Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Mathematical Modelling of the Human Vocal Tract

By XiaoBo Lu

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Auckland

The Auckland Bioengineering Institute
University of Auckland
New Zealand
December 2015
Abstract

An anatomical and customised ‘Talking Head’ model for the human vocal system has been developed. The model is created in three stages: the creation of a static model; the simulation of articulatory motion and the segmentation and generation of a 3D mesh representing the vocal tract.

To simulate the sound, we have developed a new Finite Element formulation for the Lighthill aeroacoustic analogy, coupled with incompressible flows in our customised FORTRAN code. The finite element model comprises a three-step procedure where the incompressible Navier-Stokes equations are numerically resolved by the Finite Element method in order to provide solutions for incompressible flow. The results from the incompressible flow are used to evaluate the acoustic source field assembled in different forms which then are used as the stress terms in the inhomogeneous wave equations according to the Lighthill equation.

Our modified aeroacoustic model has also been applied in a pilot study for simulating fricative sounds /s/ and /ʃ/ in the 2D domain. The study has adapted Shadle’s mechanical models as the basis to investigate the physical mechanisms of sound production with respects to three geometric factors including the length of the front cavity, the presence of the obstacle and the size of the constriction. The simulated acoustic far-fields are normalised for 3D space and compared with the data from Shadle’s experiment.
Acknowledgement

It has been a long journey since the beginning of the project. When looking back the track, there are lot of people and organisations which have offered valuable support towards the realisation of the project. I must first thank the University of Auckland and the Bioengineering Institute for providing the opportunity and funding for this project.

A big thank-you to all my supervisors Peter Hunter, William Thorpe, John Cater and Gordon Mallinson for their constant support and guidance. Without their guidance, it would be impossible to complete the project.

Throughout the project, I have taken too many favours from our institute IT teams and would like to send my deep appreciation to Tiong Lim who has provided an important service in the maintenance of the HPC and always been very competent and creative in solving a computer crisis. Many of the graphics included in this thesis are created with the aid of the Cmgui team in the Bioengineering Institute, whose skill is excellent and highly regarded.

Special thanks goes to David Nordsletten for lending me the CFD source code without which the numerical work in the project could not have been developed.

I am grateful to Peter Bier from the Department of Engineering Science, Kumar Mithraratne from the Bioengineering Institute, Catherine Watson from the Department of Electrical Engineering, Stuart Norris from the Department of Mechanical Engineering, Anna Lydon in the Advance Imaging unit in the School of Medicine, Kylie Foster from the Massey University, Callum Spencer from the Canterbury University for their assistance and contributions at various stages of the project.

I would like to thank my fellow colleagues Nina Van Essen, David Ladd, Harvey Ho, YiKun Wang, Shanon Li, Vijayaraghavan Rajapopal, Alan Wu, to name a few among many other current or past members in the Bioengineering Institute, as I have greatly benefited from the creative discussions with them from time to time.

A big thank you to my summer student Aswin Narayanan who has helped to compile many of the results of the ‘Talking Head’ model.

At last, I own too much from my family and family friends who gave enormous support both financially and emotionally to me during the course. God bless you all no matter how far you are away.
Nomenclature

\( \mathcal{P} \) Function of the voice

\( S \) Function of the source

\( \mathcal{H} \) Function of the vocal tract transmission property

\( \mathcal{R} \) Function of the Radiation

\( c \) Speed of sound

\( L \) Acoustic inertance

\( C \) Acoustic capacitance

\( R \) Acoustic resistance

\( G \) Acoustic conductance

\( Z \) Acoustic impedance

\( i, j, k, m, n, h, d \) Integers

\( \Omega \) Function space

\( \mathbb{R}^d \) Real coordinate space of \( n \) dimensions

\( L^2 \) \( \ell^2 \)-norm of a function

\( H^h \) Sobolev space containing \( L^2 \) functions whose derivatives are also \( L^2 \) up to the \( h^{th} \) order

\( f \) A mathematical function

\( \Gamma_i \) Boundary of the domain \( \Omega_i \)

\( \xi_i, \eta_i \) Local coordinates of the FE element

\( \phi, \Psi, \varphi \) FE basis functions

\( x_i \) Geometric field variable

\( \alpha, \beta \) Coefficient matrix for basis functions
s  Arc length
zi  A discrete data set
Fe  Error function for the data projection
e_d  Error projection
w  Weighting function
wi, γi  Coefficients of the Sobolev smoothing function
Fs  Sobolev smoothing function
Θ  EMA signal
T_{affine}  Affine transformation matrix
T_{3\times3}  Inner 3 by times 3 square matrix of affine transformation matrix
R_{3\times3}  Orthogonal matrix
U_{3\times3}  Upper triangular matrix
Ω_{host}  Function space of the host mesh
Ω_{slave}  Function space of the slave mesh
x_{master}  Geometric field variable in the master mesh
x_{slave}  Geometric field variable in the slave mesh
T_x, T_y, T_z  Translations in x, y and z directions
R_x, R_y, R_z  Rotations about x, y and z directions
/i:/  Phonetic symbol for a front-high vowel
/a:/  Phonetic symbol for a central-mid vowel
/ɔ:/  Phonetic symbol for a back-low vowel
/s/  Phonetic symbol for an aveolar fricative consonant
/ʃ/  Phonetic symbol for an aveo-palatal fricative consonant
ρ  Density
p  Pressure
u_i  Velocity of the fluid
\( t \)  Time
\( \omega_i \)  Vorticity/Angular velocity
\( E \)  Energy density
\( \sigma_{ij} \)  Viscous stress
\( q_i \)  Heat flux
\( H \)  Total enthalpy
\( \gamma \)  Adiabatic index
\( l \)  Length scale
\( M \)  Mach number
\( \bar{\rho}, \bar{p}, \bar{u}, \bar{\omega}, \bar{E}, \bar{h}, \bar{M} \)  Mean quantities
\( \rho', p', u', \omega', E', h', M' \)  Perturbation quantities
\( S_i \)  Lilley's acoustic source
\( L_i \)  Lamb vector
\( c_0 \)  Speed of sound in static medium
\( T_{ij} \)  Lighthill stress tensor
\( H(S) \)  Heavy side function
\( \delta(S) \)  Impulse function
\( \hat{u}_i \)  Velocity of the solid wall
\( F_i \)  Acoustic dipole function
\( Q_i \)  Acoustic monopole function
\( A_{ij} \)  Coefficient matrix
\( B_j \)  Variable vector
\( C_i \)  Source vector
\( P \)  Lagrangian multiplier
\( \lambda \)  Dynamic bulk viscosity
\( \nu \)  Kinematic shear viscosity
\( \mu \) Dynamic shear viscosity

\( \delta \) Numerical step size

\( n_i \) Surface normal

\( Re \) Reynolds number

\( D \) Diameter

\( \rho_0 \) Density at the reference state

\( \delta_{ij} \) Identity matrix

\( \hat{T}_{ij} \) Filtered Lighthill stress tensor

\( \hat{\rho} \) Normalised density

\( \Gamma \) Circulation

\( \Phi \) Velocity potential function

\( z, b \) Complex functions

\( r \) Radius

\( \theta \) Angular phase

\( T \) Rotational period

\( C_L \) Lift coefficient

\( C_d \) Drag coefficient

\( C'_L \) Fluctuating lift coefficient

\( \bar{C}_D \) Mean drag coefficient

\( u_\infty \) Far-filed velocity

\( \kappa \) Adiabatic heat capacity ratio

\( \Delta P \) Pressure drop

\( r_{\text{listening}} \) Distance between the fricative model and the listening point

\( Lb_1 \) Extended length to the back cavity in the fricative models

\( Lb_2 \) Length of the back cavity in the fricative models

\( Lc \) Length of the constriction in the fricative models
$Lf_1$ Distance between the constriction and the obstacle in the fricative models

$Lf_2$ Distance between the obstacle and the end of the front cavity in the fricative models

$LI$ Distance between the obstacle and the listening point in the fricative models

$Lo_1$ Length of the open space in the fricative models

$Lo_2$ Extended length of the open space in the fricative models

$Hb$ Height of the back and front cavities in the fricative models

$Hc$ Height of the constriction in the fricative models

$Ho_1$ Height of the lower edge of the open space in the fricative models

$Ho_2$ Height of the upper edge of the open space in the fricative models

$u_0$ Inlet velocity

$P_c$ Pressure drop across the constriction in the fricative models
## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALE</td>
<td>Arbitrary Lagrangian Eulerian</td>
</tr>
<tr>
<td>APE</td>
<td>Acoustic perturbation equations</td>
</tr>
<tr>
<td>BVI</td>
<td>Body vortex interaction</td>
</tr>
<tr>
<td>CAA</td>
<td>Computational Aeroacoustics</td>
</tr>
<tr>
<td>CFD</td>
<td>Computational Fluid Dynamics</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>DNS</td>
<td>Direct Numerical Simulation</td>
</tr>
<tr>
<td>DOF</td>
<td>Degrees of Freedom</td>
</tr>
<tr>
<td>EMA</td>
<td>Electromagnetic articulography</td>
</tr>
<tr>
<td>EMG</td>
<td>Electromyography</td>
</tr>
<tr>
<td>EPG</td>
<td>Electropalatography</td>
</tr>
<tr>
<td>FE</td>
<td>Finite Element</td>
</tr>
<tr>
<td>FEM</td>
<td>Finite Element Method</td>
</tr>
<tr>
<td>FID</td>
<td>Free Induction Decay</td>
</tr>
<tr>
<td>FOV</td>
<td>Field of View</td>
</tr>
<tr>
<td>HPC</td>
<td>High Performance Computer</td>
</tr>
<tr>
<td>LBB</td>
<td>Ladyzhenskaya-Babuska-Brezzi</td>
</tr>
<tr>
<td>LEE</td>
<td>Linearised Euler equations</td>
</tr>
<tr>
<td>LES</td>
<td>Large eddy simulation</td>
</tr>
<tr>
<td>LMS</td>
<td>Least median squares</td>
</tr>
<tr>
<td>LPCE</td>
<td>Linearised perturbed compressible equations</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>MAP</td>
<td>Muscle Activation Pattern</td>
</tr>
<tr>
<td>MC</td>
<td>Marching cubes algorithm</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic Resonance</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>PCE</td>
<td>Perturbed compressible equations</td>
</tr>
<tr>
<td>PIV</td>
<td>Particle Image Velocimetry</td>
</tr>
<tr>
<td>RANS</td>
<td>Reynolds Averaged Naiver-Stokes</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency (a MRI excitation signal)</td>
</tr>
<tr>
<td>RMS</td>
<td>Root mean square</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>TE</td>
<td>Echo Time (a MRI scanning parameter)</td>
</tr>
<tr>
<td>TR</td>
<td>Repetition Time (a MRI scanning parameter)</td>
</tr>
<tr>
<td>VT</td>
<td>Vocal Tract</td>
</tr>
</tbody>
</table>
Contents

1 Introduction 1
  1.1 Physiological View of the Human Vocal System . . . . . . . . . . . . . . . . . . . 1
  1.2 Functional View of the Human Vocal System . . . . . . . . . . . . . . . . . . . . 6
  1.3 Motivation and Goals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2 Anatomical Models of the Vocal Articulators 16
  2.1 Magnetic Resonance Imaging . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16
    2.1.1 MRI Scanning and Image Acquisition . . . . . . . . . . . . . . . . . . . . . 18
    2.1.2 Image Analysis and Segmentation . . . . . . . . . . . . . . . . . . . . . . 19
  2.2 Finite Element Discretisation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
    2.2.1 Sobolev Space . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
    2.2.2 Subdivision of the Function Space . . . . . . . . . . . . . . . . . . . . . . 21
    2.2.3 Assembly of the Finite Element Space . . . . . . . . . . . . . . . . . . . . 22
  2.3 Least-Squares Fitting . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
    2.3.1 Initial Linear Meshes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
    2.3.2 Data Projection to Parametrised Surfaces . . . . . . . . . . . . . . . . . . 25
    2.3.3 Optimisation of Nodal Parameters . . . . . . . . . . . . . . . . . . . . . . 27
    2.3.4 ‘Talking Head’ Assembly . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
    2.3.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
  2.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

3 Simulation of Articulatory Movements in Speech Sounds 39
  3.1 Data Collection and Processing . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
    3.1.1 Electromagnetic Articulography . . . . . . . . . . . . . . . . . . . . . . . . 39
    3.1.2 The EMA-Video Experiment . . . . . . . . . . . . . . . . . . . . . . . . . . 40
    3.1.3 Data Processing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
  3.2 Rigid Body Transformation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
  3.3 Modelling Deformable Bodies-Host Mesh Fitting . . . . . . . . . . . . . . . . . . . 47
    3.3.1 Host and Slave Meshes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47
    3.3.2 Matching Reference to Target Points . . . . . . . . . . . . . . . . . . . . . 47
    3.3.3 Updating the Slave Mesh . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
3.4 Contact between Two Rigid Bodies ................................................. 51
3.5 Results ......................................................................................... 51
  3.5.1 Segmentation of Phonemes ......................................................... 52
  3.5.2 Jaw Motion ............................................................................ 52
  3.5.3 Deformation of the Tongue ...................................................... 53
  3.5.4 Deformation of the Cheek-Lips ............................................. 55
  3.5.5 Contact between the Tongue and Epiglottis ............................... 56
3.6 Discussion .................................................................................... 57
3.7 Summary ...................................................................................... 58

4 Geometric Modelling of the Vocal Tract ............................ 59
  4.1 Introduction .............................................................................. 59
  4.2 Vocal Tract Segmentation .......................................................... 62
    4.2.1 Digitising the Articulator Surfaces ........................................ 62
    4.2.2 Overlapped Tree Partitioning Strategy .................................. 65
    4.2.3 Flood-Fill Method .............................................................. 67
    4.2.4 Segmentation for Vocal Tract Configurations .......................... 68
  4.3 Vocal Tract Meshing ................................................................. 68
    4.3.1 Surface Extraction ............................................................. 69
    4.3.2 Volume Reconstruction ....................................................... 73
  4.4 Validation of the Results ........................................................... 76
    4.4.1 Validation of the Rest Breathing State .................................. 79
    4.4.2 Comparison between MRI and Modelled VT ....................... 81
  4.5 Summary .................................................................................... 85

5 Aeroacoustic Modelling Framework .............................. 86
  5.1 Aeroacoustic Modelling of Human Speech Sounds .......... 86
    5.1.1 Empirical Modelling .......................................................... 86
    5.1.2 Direct Numerical Simulation .............................................. 87
    5.1.3 Hybrid Methods based on the Perturbation Equations .......... 89
    5.1.4 Hybrid Methods based on the Aeroacoustic Analogies .......... 91
  5.2 Finite Element Method for Unsteady Viscous Incompressible Navier-Stokes Equations ............................................. 93
    5.2.1 Weak Form .................................................................... 94
    5.2.2 Numerical Discretisation of the FE Fields ......................... 95
    5.2.3 Newton-Raphson Technique .......................................... 97
    5.2.4 Test Problem 1: The Backward-Facing Step ................... 98
  5.3 Finite Element Method for the Lighthill Aeroacoustic Analogy . 101
5.3.1 Approximation of the Lighthill Analogy for Compressible and Incompressible Flows ........................................... 102
5.3.2 Alternative View of the Lighthill Aeroacoustic Analogy ...................................................... 104
5.3.3 Weak Form ............................................................................................................................. 106
5.3.4 FE Discretisation of the Acoustic Fields .............................................................................. 106
5.3.5 Test Problem 2: Spinning Vortex Pair .............................................................................. 109
5.4 Test Problem 3: Aeroacoustic Simulation of 2D Flow Around a Square Cylinder 121
5.4.1 Incompressible Flow Field ................................................................................................. 122
5.4.2 Acoustic Near-Field ............................................................................................................ 126
5.4.3 Acoustic Far-Field ............................................................................................................... 131
5.4.4 Effect of Source Truncation associated with the Vortex Train ........................................ 133
5.4.5 Mach Number Effect .......................................................................................................... 136
5.5 Discussion and Conclusions ................................................................................................. 138
5.6 Summary .................................................................................................................................. 139

6 Aeroacoustic Modelling of Fricatives /s/ and /ʃ/ 140
6.1 Physical Properties of Fricatives /s/ and /ʃ/ ........................................................................ 140
6.1.1 Articulation .......................................................................................................................... 141
6.1.2 Aerodynamics ...................................................................................................................... 142
6.1.3 Acoustics .............................................................................................................................. 143
6.2 Aeroacoustic Simulations on Simplified 2D Geometries for Fricatives /s/ and /ʃ/ 147
6.2.1 Problem Description ........................................................................................................... 151
6.2.2 Case 1: The Effect of the Length of the Front Cavity ......................................................... 155
6.2.3 Case 2: 2D /s/ and /ʃ/ Models with Obstacles ..................................................................... 165
6.2.4 Case 3: The Modified 2D /s/ and /ʃ/ Models with Obstacles ............................................. 173
6.3 Discussion ................................................................................................................................ 180
6.4 Summary .................................................................................................................................. 184

7 Conclusions .............................................................................................................................. 185

8 Future Work .................................................................................................................................. 187

Appendices ...................................................................................................................................... 190

A List of Words and Sentences Used in the EMA Experiment ....................................................... 191

B Error Analysis on Articulatory Models in Speech Configurations ............................................ 194
B.1 Error Analysis of the Tongue Models .................................................................................... 194
B.2 Error Analysis of the Cheek-Lips Models ............................................................................ 196

C Segmented Vocal Tract Surface for Six Configurations ................................................................. 200
D User Guide for AcousticParallel

D.1 Compile the Program .................................................. 204
D.2 Running the Program .................................................. 204
D.3 IO Module .............................................................. 204
D.4 FE Module .............................................................. 206
D.5 Boundary Module ...................................................... 207
D.6 Matrix Module .......................................................... 208
D.7 Solver Module ........................................................... 209

E ALE Form of The Lighthill Equation .................................. 211

F New Formulation for Aeroacoustic Equations .................... 213

F.1 Two-Equation Model for Modelling Flow-Acoustic Interactions ............... 213
F.2 Evaluation of the Third-order Directives ............................ 214
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A mid-sagittal view of the human vocal tract.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>The larynx and glottis structures.</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>The pharynx cavity.</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>The nasal and oral cavities.</td>
<td>4</td>
</tr>
<tr>
<td>1.5</td>
<td>Articulation of four English vowels in mid-sagittal MRI scans.</td>
<td>5</td>
</tr>
<tr>
<td>1.6</td>
<td>Articulation of five English consonants in mid-sagittal MRI scans.</td>
<td>6</td>
</tr>
<tr>
<td>1.7</td>
<td>A schematic diagram of the functional units involved in speech production.</td>
<td>7</td>
</tr>
<tr>
<td>1.8</td>
<td>A schematic diagram of the source-filter model.</td>
<td>8</td>
</tr>
<tr>
<td>1.9</td>
<td>A schematic diagram of the concatenated tubes model of the VT.</td>
<td>9</td>
</tr>
<tr>
<td>2.1</td>
<td>Digitisation of the tongue in the Zinc digitiser.</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Tessellation of the 2D space.</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Steps of the least-squares fitting algorithm on the two articulator models.</td>
<td>26</td>
</tr>
<tr>
<td>2.4</td>
<td>The maxilla model.</td>
<td>29</td>
</tr>
<tr>
<td>2.5</td>
<td>The mandible model.</td>
<td>29</td>
</tr>
<tr>
<td>2.6</td>
<td>The tongue model.</td>
<td>30</td>
</tr>
<tr>
<td>2.7</td>
<td>The cheek-lips model.</td>
<td>31</td>
</tr>
<tr>
<td>2.8</td>
<td>The upper face model.</td>
<td>31</td>
</tr>
<tr>
<td>2.9</td>
<td>The soft palate model.</td>
<td>32</td>
</tr>
<tr>
<td>2.10</td>
<td>The backwall model.</td>
<td>32</td>
</tr>
<tr>
<td>2.11</td>
<td>The epiglottis model.</td>
<td>33</td>
</tr>
<tr>
<td>2.12</td>
<td>The larynx model.</td>
<td>34</td>
</tr>
<tr>
<td>2.13</td>
<td>Assembly of the ‘Talking Head’.</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>The EMA experiment.</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>Locations of EMA sensors and visual markers.</td>
<td>41</td>
</tr>
<tr>
<td>3.3</td>
<td>Data processing procedure.</td>
<td>42</td>
</tr>
<tr>
<td>3.4</td>
<td>The modelling frame and the reference state of the EMA sensors.</td>
<td>43</td>
</tr>
<tr>
<td>3.5</td>
<td>The resampling of the EMA signal of the tongue tip sensor in the word ‘hoard’.</td>
<td>44</td>
</tr>
<tr>
<td>3.6</td>
<td>The filtering of the acoustic data.</td>
<td>44</td>
</tr>
<tr>
<td>3.7</td>
<td>The reference state of the cheek-lips mesh and corresponding video frames.</td>
<td>45</td>
</tr>
<tr>
<td>3.8</td>
<td>The jaw movements during the word ‘hoard’.</td>
<td>46</td>
</tr>
</tbody>
</table>
3.9 The host mesh fitting on the tongue.

3.10 The host mesh fitting on the cheek-lips mesh.

3.11 Contact simulation between the tongue and epiglottis.

3.12 The segmentation of phonemes from five short words.

3.13 Positions of the tongue sensors for the three vowels and two fricatives.

3.14 Mid-sagittal contours of the simulated articulations versus MRI scans of five phonemes.

3.15 The deformation of the cheeks and lips by host mesh fitting.

3.16 Contact simulations between the tongue and epiglottis in vowels and fricatives.

4.1 The selected faces of the articulator models.

4.2 The digitised surfaces for six VT configurations.

4.3 A schematic of the overlapping cells used in a 2D implementation of the algorithm.

4.4 Segmentation of a 2D quadratic triangle.

4.5 The pseudo-code for the flood-fill algorithm.

4.6 Segmentation of the vocal tract for the rest breathing state.

4.7 The extraction of a parametric surface.

4.8 Improvements on the initial parametric surface.

4.9 Subdivision of the vocal tract mesh.

4.10 Reconstructed VT surfaces after fitting.

4.11 Reconstructed bilinear vocal tract surfaces after fitting.

4.12 Estimation of a centerline and normal planes for the rest breathing state.

4.13 The linear VT meshes for the reference state.

4.14 The fitted cubic Hermite VT mesh for the resting breathing state.

4.15 The fitted cubic Hermite mesh for the five speech configurations.

4.16 Comparison between the MRI scans and the 3D model for the vocal tract in the reference state.

4.17 Comparison between the MRI scans and the model for the sampled cross-sections in the rest breathing state.

4.18 Comparison of the area functions measured by MRI and mathematical models.

4.19 Comparison of the area functions measured by MRI and mathematical models.

5.1 The pseudo-code for the CFD solver.

5.2 The backward-facing step problem.

5.3 The mesh convergence analysis at Re=2,000.

5.4 A comparison of the steady-state solutions at Re=100-600.

5.5 A schematic diagram of the procedures for the aeroacoustic solver.

5.6 The pseudo-code for the CAA solver.

5.7 The CFD domain of the spinning vortices.

5.8 The calculated velocity gradients at r = 1.02 and θ = π/3.
6.2 The computational domain for Shadle’s fricative models. .......................... 151
6.3 The coarse mesh of Model 3. ................................................................. 154
6.4 The streamlines of the artificial steady-state for Models 1 and 2. .............. 156
6.5 The streamlines of the averaged transient flow for Models 1 and 2. ........ 157
6.6 The pressure drop across the front cavity of Models 1 and 2. .................. 158
6.7 The Lighthill source evaluated at the end of the tract in Models 1 and 2. .. 159
6.8 The time-averaged acoustic source fields of Models 1 and 2. ................... 161
6.9 The acoustic source fields evaluated in each sections of Models 1 and 2. .... 162
6.10 The acoustic source fields evaluated in the front cavities of Models 1 and 2. 162
6.11 The acoustic far-fields of Models 1 and 2. .............................................. 163
6.12 The spectra of the acoustic far-fields for Models 1 and 2. ....................... 164
6.13 The streamlines of the artificial steady-state for Models 3 and 4. .......... 165
6.14 The streamlines of the averaged transient flow for Models 3 and 4. ......... 166
6.15 The pressure drop across the front cavity of Models 3 and 4. .................. 167
6.16 The Lighthill source evaluated at the end of the tract in Models 3 and 4. .. 168
6.17 The time-averaged acoustic source fields of Models 3 and 4. .................. 169
6.18 The acoustic source fields evaluated in various sections of Models 3 and 4. 170
6.19 The acoustic source fields evaluated in the front cavities of Models 3 and 4. 171
6.20 The acoustic far-fields of Models 3 and 4. ............................................ 171
6.21 The spectra of acoustic far-fields for Models 3 and 4. ............................. 172
6.22 The streamlines of the artificial steady-state for Models 5 and 6. .......... 173
6.23 The streamlines of the averaged transient flow for Models 5 and 6. .......... 174
6.24 The pressure drop across the front cavity of Models 5 and 6. ................. 175
6.25 The Lighthill source evaluated at the end of the tract in Models 5 and 6. .. 176
6.26 The time-averaged acoustic source fields for Models 5 and 6. ............... 178
6.27 The acoustic source fields evaluated in each section of Models 5 and 6. .. 179
6.28 The acoustic source fields evaluated on the solid walls of the front cavities in Models 5 and 6. ......................................................... 179
6.29 The acoustic far-fields for Models 5 and 6. .......................................... 180
6.30 The spectra of acoustic far-fields of Models 5 and 6. ............................ 181

B.1 The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /s/. . . 194
B.2 The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /ʃ/. . . 195
B.3 The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /iː/. . . 195
B.4 The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /aː/. . . 196
B.5 The comparison between the imaged (sustained articulation in MRI) and the
simulated tongue shape using the data derived from EMA in phoneme /ɔ:/ . . . . 196
B.6 The comparison between the image and the simulated lips shape using the data
derived from video cameras in phoneme /s/. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 197
B.7 The comparison between the image and the simulated lips shape using the data
derived from video cameras in phoneme /ʃ/. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 197
B.8 The comparison between the image and the simulated lips shape using the data
derived from video cameras in phoneme /iː/. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 198
B.9 The comparison between the image and the simulated lips shape using the data
derived from video cameras in phoneme /aː/. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 198
B.10 The comparison between the image and the simulated lips shape using the data
derived from video cameras in phoneme /ɔː/. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 199

C.1 Segmentation of the vocal tract surface in the rest breathing state. . . . . . . 200
C.2 Segmentation of the vocal tract surface in the fricative /s/. . . . . . . . . . . . . 201
C.3 Segmentation of the vocal tract surface in the fricative /ʃ/. . . . . . . . . . . . . 201
C.4 Segmentation of the vocal tract surface in the vowel /iː/. . . . . . . . . . . . . 202
C.5 Segmentation of the vocal tract surface in the vowel /aː/. . . . . . . . . . . . . 202
C.6 Segmentation of the vocal tract surface in the vowel /ɔː/. . . . . . . . . . . . . 203
List of Tables

2.1 MRI Scanning Parameters .............................................. 19
2.2 Nine Articulator Meshes Created for the ‘Talking Head’ model. ............ 34
2.3 Sobolev Smoothing Coefficients used in the Least-Squares Fitting .......... 37
3.1 Sobolev Smoothing Coefficients used in Host Mesh Fitting ................. 50
3.2 Jaw Movements for 5 Phonemes ........................................ 52
3.3 Comparison between Simulated Tongue Shapes & MRI Scans ................. 55
3.4 Comparison Between Simulated Lips Shapes & MRI Scans .................... 56
4.1 Segmentation of Six Vocal Tract Configurations .......................... 69
4.2 Vocal Tract Meshes for Six Configurations ................................ 73
4.3 Cubic Vocal Tract Meshes for Six Configurations ........................... 78
5.1 Comparison between Different Flow Models ................................ 125
6.1 Parameters for Simulating Unvoiced Fricatives ............................ 143
6.2 Dimensions of the Simulated Domain ..................................... 153
6.3 Grid Parameters for the Simulated Domains ................................ 153
6.4 Mesh Analysis of the Pressure Drop across the Constriction ............... 156
6.5 Mesh Analysis of the Pressure Drop across the Constriction ............... 166
6.6 Mesh Analysis of the Pressure Drop across the Constriction ............... 174
A.1 A List of Speech Sentences Used in the EMA Experiment ................... 191
A.2 A List of Speech Words Used in the EMA Experiment (Part 1) .............. 192
A.3 A List of Speech Words Used in the EMA Experiment (Part 2) .............. 193
Chapter 1

Introduction

Speech is the acoustic end product of voluntary, formalised motions of the respiratory and masticatory apparatus [1]. In Section 1.1, the components of human vocal system and their physiological roles in speech production are examined. For most of speech sounds, subglottal systems (e.g. the lungs) provides the energy while structures above the glottis manipulating airflow into audible sounds [2]. Therefore human vocal system may be viewed as a circuit and the functions of individual vocal organs are explained in Section 1.2). After a brief review of the current speech models, the motivations and goals of the thesis are outlined in Section 1.3.

1.1 Physiological View of the Human Vocal System

The human vocal tract (VT) (highlighted in blue in Figure 1.1) may be conceptualised as a tubular air passage approximately 17cm long for an adult male, with varying cross-sectional area (0 – 20cm²) [3], surrounded by various types of soft tissues.

At the bottom end, the lung generates the force to drive the air in and out of the airway. On leaving the lung, air passes through trachea and enters the VT where speech sounds are produced. The human VT may be divided into four cavities starting from the most inferior end to the anterior: the laryngeal, pharyngeal, oral and nasal cavities. Before entering the larynx, the air has to pass through a narrow opening called the glottis which is an orifice covered by two membrane-like soft tissues known as vocal folds. The vocal folds are anchored to the cartilaginous structures of the larynx which surrounds and protects the glottis (Figure 1.2(a)) [5]. It begins at the level of vertebra C₄ or C₅ and ends at the level of vertebra C₇ and is a cartilaginous cylinder whose walls are stabilised by ligaments (i.e. vocal cords) or skeletal muscles or both (Figure 1.2(b)) [6]. Vocal folds act like gates to the VT as they open the airway to the incoming air when they part or stop the flow the moment they fold back [4]. Vocal folds are capable of sustained vibration under the pressure of the air.

The middle portion of the VT, known as the pharynx which connects the nose, mouth and throat, can be subdivided into three regions [6]: the nasopharynx which forms the superior portion of the pharynx and connects to the posterior portion of the nasal cavity; the oropharynx
Figure 1.1: A mid-sagittal view of the human Vocal Tract where the Vocal Tract is highlighted in blue. Reproduced from [4].
which extends between the soft palate, also known as velum, and the base of the tongue at the level of hyoid bone; and the larynxgopharynx which includes the area between the hyoid bone and the entrance of the esophagus, as shown in Figure 1.3.

The air pathway to the nasal cavity is controlled by the velum (i.e. soft palate), a piece of cartilage extending from the posterior end of the hard palate. The nasal cavity, shown in Figure 1.4(a), can be decoupled from the VT by voluntarily raising the velum. The hard palate, mainly made of bone, separates the oral cavity from the nasal cavity and forms the roof of the oral cavity (see Figure 1.4(b)).

The tongue is a highly mobile muscular structure, owing to its complex intrinsic and extrinsic muscle groups. Its surface extends from the hyoid bone at the back of the mouth upward and forward to the lips, which makes it form almost two-thirds of the bottom surface of the VT [6]. Its upper surface and the forward part of the lower surface are free to move; elsewhere it is attached to the adjacent components of the mouth.

Surrounding the tongue, there are the teeth, embedded in the gums of the mandible (jaw) and the maxilla. The two sides of the oral cavity are covered by the cheeks. At the end of the tract, the lips are the only exit for the air in the absence of a nasal passage. When talking, lips may be retreated or protruded, by which the length of the VT is effectively changed by 1-2cm [2].

The basic units of speech sounds are named ‘Phonemes’ by linguists and are generally classified into vowels and consonants. Vowels are always voiced sounds, meaning that they are made with glottal vibration. There are two types of vowels: monophthongs, in which the VT shape is time-invariant, and dipthongs, in which the VT dynamically changes its configuration.
Figure 1.3: The pharynx cavity, reproduced from [7]. The opened posterior view of the pharynx cavity.

Figure 1.4: The nasal and oral cavities. (a) The sagittal view of the nasal cavity, reproduced from [8]. (b) The anterior view of the oral cavity, reproduced from [9].
during the course. The articulation of English monophthongs are best described by their tongue positions. As shown in Figure 1.5, there are mid-sagittal scans of a high-front vowel /i:/, a mid-front vowel /e/, a low-central vowel /a:/ and a low-back vowel /ɔ/. On the other hand, the diptongs change noticeably between the positions which are assumed for pure vowels [10].

Figure 1.5: Articulation of four New Zealand English vowels in mid-sagittal MRI scans. (a) /i:/; (b) /e/; (c) /a:/; (d) /ɔ/.

English consonants can be voiced or unvoiced and are described by the place-of-articulation as labials (lips), dental (teeth), alveolars (gums), palatals (hard palate), velars (soft palate), glottal (glottis) and libiodentals [10]. There are two main groups of consonants, called stops and fricatives. Stops are produced by building pressure behind an blocking point and quickly releasing of the pressure. Figures 1.6(d) and (e) show the pre-release articulations of the two stops /k/ and /t/, in which the airway is completely blocked at the velum (velar) and teeth (alveolar) respectively. The airway can be also blocked at lips (labial), gums (dental), hard palate (palatal), glottis (glottal) and so on. Fricatives are made by forcing air through a narrow constriction and can be classified by the place-of-constriction such as the three fricatives shown in Figures 1.6(a) alveolar, (b) palatal and (c) dental.
1.2 Functional View of the Human Vocal System

The human VT, viewed as a voice-making apparatus, may be explained as a circuit made of various functional units demonstrated in the schematic diagram in Figure 1.7 [1].

At the beginning of the circuit, the lung is often considered as the generator as it initiates the flow with the assistance of respiratory muscles. All speech sounds are flow-induced, however in the trachea the air flow is mostly silent. The flow can induce vibrations on the vocal folds when passing through the glottis, causing a cyclic motion of opening and closure of the glottal orifice, hence effectively chopping the continuous air stream into periodic puffs. This physiological response becomes the primary sound generation mechanism in human speech. The glottal sound is still not quite what is actually heard in normal speech. The initial sound waves generated by the glottal vibration may be further modified by the acoustic properties of the VT. After the larynx, the pharynx and oral cavities have been viewed by many phoneticians as acoustic resonators, whose resonant frequencies are defined by their physical shape [11, 1, 3]. More importantly, these resonators can change their shape by moving surrounding articulators like the tongue. For instance, as the tongue moves to the front and back, it changes the length and volumes of the two cavities, altering the resonance response of the VT. The nasal cavity may add additional resonant (anti-resonant) frequencies to the system if it is connected to the main passage via the velum.

Adjusting the VT shape to produce different sounds is called ‘Articulation’ [12]. Further-
more, more sound may be generated in the supra-glottal space by forming constrictions or obstructions at various points. The nose and lips are the two radiation ports for the sounds and can be used in pairs or individually in human speech production. Generally, acoustic energy is lost in the nasal passage due to its large surface area lined with soft tissue (i.e. sinuses) [2]. Consequently, only a little is emitted through the nostrilitis compared to the mouth opening for those non-nasal sounds.

A view of the speech production is based on the acoustic theory developed by Fant [11], which may be formulated as a mathematical relation

\[ P = S \cdot H \cdot R, \]

which interpreters the speech signal \( P \) as a product of the function of sound source \( S \), the VT transmission function \( H \) and the radiation property at the exit of the VT \( R \). This equation implies that the sound production (the source) and the sound propagation (the filter) are two independent processes that can be modelled independently. For this reason, it has been given the name the ‘source-filter’ model, as illustrated in Figure 1.8 where the source function representing the glottal vibration is convoluted with the VT filter function \( H \cdot R \) to produce the voice.

It is generally recognised that there are three sound generation mechanisms involved in
human speech [3], including

1. Glottal vibration;

2. Rapid release of the pressure behind an obstruction;

3. Turbulent air streams due to flow passing through narrow spaces.

The first source mechanism is the primary sound source among vowels [11]. The vibration of the vocal folds converts the air stream into periodic volume flows which are coupled with pressure waves in the supra-glottal space. The frequency at which the vocal folds vibrate, is controlled by the air pressure in the lungs, the stiffness and mass of the vocal folds, and the area of the glottal opening in the rest position [3], however at some frequency, source-tract interaction can also occur [1]. Among source-filter models, the glottal vibration is normally modelled via a glottal impedance which converts the given pressure gradients across the glottis into volume flows. The calculation of the glottal impedance may include the effect of area changes of the orifice, the viscous loss (which could be significant in small orifice areas), the wall compliance and other physiologically derived parameters [1, 3].

Stop consonants are made by dynamic articulatory motion of a closure followed by a rapid release somewhere along the tract. For this reason, this group of sounds are often modelled by introducing a step function of volume flow at the point of release, and then gradually transient to a fricative sound source as the closure opens up [2].

Compared to the glottal vibration, the third group of sources is aperiodic, even being described as noise [11]. Fricative consonants are produced by turbulent air formed by either one or more of the following ways: forcing air streams through narrow channels, passing around obstacles (e.g. teeth) or impinging jets on the VT walls [1]. There are many places that such sound sources can be initiated and while their exact locations are difficult to pin-point, they are generally considered to be at the most constricted area of the tract or just anterior to the constriction [1]. Less is known about the sound associated with the turbulence than with the glottal vibration, therefore its spectral character is often derived from empirical models based on experimental results [1]. The strength of the sound source is believed to be related to the
volume of the air flow [3]. For voiced consonants, both the glottal vibration and turbulence source are present [3].

For modelling the propagation behaviour, the VT has been viewed as a series of concatenated uniform tubes of finite length [3], as illustrated in Figure 1.9. It can be shown that under some approximations of the compressible Navier-Stokes equations, the pressure propagation inside these tubes can be solved by a 1D wave equation whose solutions produce both forward travelling (towards the lips) and backward travelling (towards the glottis) waves at the speed of sound \( c \) at each intersection of the tubes and can be calculated based on area functions measured as the cross-sectional areas versus the lengths of the tubes [3]. As a result, some source frequencies are selectively amplified according to the VT shapes and such frequencies are called ‘formants’. Each vowel sound is characterised by a unique VT configuration, which arises to a unique formant structures in their individual spectra.

In a source-filter model, the resonator property can be mathematically represented by a combination of acoustic inertance \( L \) derived from Newton’s second law and the acoustic capacitance \( C \) based on the constitutive relation between the pressure and density [1]. For a more advanced VT transmission model, more acoustic elements might be added, like the acoustic resistance \( R \), representing a power loss due to viscous dissipation; the acoustic conductance \( G \), providing a power loss from the heat conduction through the VT wall, and the acoustic impedance \( Z \) which is related to the compliance of the tract wall and other parameters derived from various physiological responses [1, 2]. Monophthongs and some fricative consonants can be modelled by static VT configurations while the modelling of dipthongs and stops require estimates of the time-varying area functions in order to represent the dynamic VT shapes.

![Figure 1.9: A schematic diagram of the concatenated tubes model of the VT.](image)

### 1.3 Motivation and Goals

In previous sections, a general picture of the human vocal system and how it produces speech sounds has been given, based on the knowledge gained in past studies. However, there are still many limitations to our understanding of speech production process, which hinders the developments of speech-related technologies.
Over the years, articulatory models have shown to be useful in many applications including clinical and surgical studies [13, 14], speech synthesis [11, 15, 16, 17, 18, 19, 20] as well as speech perception and recognition [21, 22, 23], to name just a few. Articulatory models have the advantage of following the physiological process of speech production when compared to other methods such as concatenative models, therefore they have the most potential for advancing speech science as well as for synthesising high quality speech sounds [24].

Earlier articulatory models [15, 17] were constrained in 1.5D space, based on mid-sagittal contours and area functions of the VT. Given advances in both the available computer power and experimental technologies in recent years, 3D articulatory models have gained more popularity. According to Badin [25], 3D models are better than 1.5D not only for their ability to model transverse acoustic modes, which can be important for frequencies above 4-5KHz, but also for simulating the acoustic source mechanisms for some of the fricative consonants (e.g. jets impinging on the teeth). Furthermore, 3D models are better suited for visual speech synthesis and identifying the roles of individual articulators in the overall VT dynamics than 1D or 2D models.

In literature, 3D articulatory models can be either classified as data-driven parametric models [26, 25, 27] or muscle-activated biomechanical models [28, 29, 30]. The former group derives a finite number of parameters for controlling the deformation of each articulator from a group of reference articulations (often artificially sustained) while the latter aims to reproduce the physiological process responsible for the voluntary movements of speech organs.

At first glance, muscle-activated biomechanical models represent the most straightforward modelling strategy for simulating human speech articulation, however it is a non-trivial task to determine the so-called muscle activation pattern (MAP) needed for the musculature controls of the vocal articulator in practice. For control of the tongue in humans, lingual deformations are determined by the synergistic contraction of the intrinsic and extrinsic muscle fibres [31]. Early attempts using electromyographic (EMG) techniques for investigating muscle fibre activity have yielded little success for articulatory modelling largely due to mixed electrical signals from different groups of interlaced muscle fibres (11 groups in total) at the same electrode, which makes it difficult to deduce activation patterns for a particular muscle or muscle group [32, 33]. Hence EMG measurements are still limited to modelling only the activation pattern of large extrinsic muscles [29, 30]. To further complicate the model, finding appropriate constitutive relations for the biological tissue is an extremely challenging task [33] and different articulators may interact with each other (e.g. jaw-tongue contacts) during speaking [34]. Inverse methods have been adopted to derive MAPs from tracked kinematics of the tongue [29, 35, 36, 33]. Despite using such a strategy in simulating the speech articulation presented in [35], many simplifications have to be made in order to reduce the complexity of the model, including the reduction of dimensionality from 3D to 2.5D (i.e. the geometric model is constructed by extruding from mid-sagittal contours), replacing the continuum with a semi-continuum (i.e. a mass-spring approximation for the tongue tissue), no contact modelling of the tongue with the
surrounding hard and soft tissues (e.g. hard and soft palates) and so on. Moreover, muscle-activated biomechanical models often demand significant amount of computational time which may prevent its interactive usage in many of the applications [14].

On the other hand, data-driven parametric models bypass the dynamics and therefore have none of the complexity associated with the simulation of biomechanics. They parametise articulator shapes based on statistical analysis (e.g. principle component analysis) from experimental measurements. The control parameters are usually derived from measured sustained articulation by imaging techniques such as X-ray [11, 15], MRI [37, 19], ultrasound [38], video cameras [37] and others. However, these imaging techniques all fail to produce sufficient spatial or temporal resolution for tracking the movements of major vocal articulators (e.g. the tongue) in physiological tasks [33]. In contrast, point tracking technologies like electromagnetic articulography (EMA) and electropalatography (EPG), can provide a sufficient temporal sampling rate (100-500Hz) to study speech articulation but are handicapped by the intrusive manner of attaching sensors, and are subsequently restricted to operate on only a small area of articulator surfaces. More detail and a discussion of the current experimental tools will be presented in Chapters 2 and 3. Some researchers have sought to combine the measurements of different modalities in their articulatory models (e.g. MRI and video cameras in [25], EMA and motion tracking in [39]; MRI, EMA and EPG in [40] and etc.) and showed their usefulness in improving the accuracy for articulatory speech synthesis.

Overall, data-driven parametric models can only reflect the statistical behaviour but not the physiological and biomechanical constraints and therefore questions need to be raised as to how faithfully the simulated movement by a parametric model can be compatible with the physiology of speech organs [35]. Although various studies have tried to link control parameters with the physiological response of the articulators, [41] suggests that such results will remain ambiguous as long as the intrinsic structures of articulators are not modelled and the mechanical aspects not included. This view is supported by [40, 27], stating that in the long run a more physiology-oriented model is needed, in which the data-driven approach needs to be carefully combined with the biomechanical likelihood constraints.

Regarding the acoustic stages of speech production, source-filter models have been widely used in speech synthesis studies [42, 43, 44, 45]. Although a model of such a type is easy to implement and has achieved a high level of success in synthesising speech sounds, especially for vowels [11], it is based on simplified physics of aeroacoustics and faces a number of challenges for synthesising unvoiced consonants.

First of all, the mathematical properties of the source for fricative sounds are often modelled empirically based on a combination of theoretical relations and experiment measurements. Over the past decades, the theoretical framework for synthesising unvoiced sounds (e.g. fricative and plosive consonants) have been gradually moved towards more and more explicit modelling of the aeroacoustic and aerodynamic fields. Early models simulated the acoustic source for fricative sounds with random noise in an electrical analogy of the VT [43]. Later, mathematical
relations, developed to link acoustic production with aerodynamics and aerodynamic parameters (pressure drop across the constriction [46], the constriction size in [47], the jet vorticity and the mean velocity potential [48], etc.) are incorporated into the noise source model, often derived from experimental measurements on simplified VT models of fricatives. Shadle studied a number of geometric factors with respect to the impact on acoustic production in her mechanical models of fricatives [49]. In the study, inverse filtering of the far-field spectra with the known acoustic propagation property of the mechanical model was applied in order to extract the spectra for the noise source. The method is largely successful but it suffers two major setbacks, namely the pre-assumption of source-filter independence and an inability to be used on generalised geometries. The former restricts the application of the model to be only in the obstacle-configurations while the later limits the model to be used on only fixed and localised source models [49]. Given advances in computing speed and parallel programing, direct simulation of the nonlinear equations of aeroacoustics for VT flow becomes a viable option and numerical models based on aeroacoustic analogies have become an increasingly popular choice for fricative synthesis. Numerical models based on aeroacoustic analogies, such as the finite difference model in [50] and the finite element model in [51], have demonstrated the feasibility of direct aeroacoustic simulation of fricative sounds in realistic VT geometries. Compared to the theoretical method, the biggest advantage of the numerical models is the ability to use generalised VT geometries, thus allowing a comprehensive investigation of the source-tract interactions. As Shadle has pointed out [52], it is necessary to take account of the full 3D geometry of the human VT for the modelling of the source mechanism in fricative consonants.

Secondly, the filter property should represent the independent acoustic transmission property of the VT according to the acoustic theory [11]. In [49], the author admitted that the models without the presence of an obstacle have a stronger source-tract interaction than those models with an obstacle included and a realistic fricative model is likely to have a more spatially distributed source field due to the less distinct features of constriction and obstacle. Exactly how spatially distributed the source field must be for the fricatives is still debated in the literature [49, 47, 52, 53, 54] and more detail of this discussion will be given in Chapter 6. Such ambiguity between the two assumed independent physical properties raises the question about exactly what properties should be included in the ‘source’ and the ‘filter’ respectively, which reveals yet another weakness of the source-filter method. The problem is more cumbersome in the modelling of unvoiced sounds than vowels. Different source-filter models may define the ‘source’ and ‘filter’ properties based on the selected level of modelling of aerodynamics in idealised VT geometries, which makes the comparison of their performance very difficult.

Lastly, in-vivo measurements of aerodynamics are still limited to some basic quantities such as subglottal/intraoral pressure and the volume flow rate from the lung, much is owed to the invasive nature of the devices, while other methods like hot-wires and stereoscopic particle image velocimetry (PIV) are better for use on mechanical models in controlled environments [55, 24, 56]. Most of the experiments have been conducted on static and simplified geometries
of the VT. Amongst them, PIV together with 3D printing technology seems to have the best potential for moving towards realistic geometric models of VT (i.e. phantoms) despite the current limitation of their spatial and temporal resolutions, as discussed in [56]. Albeit experimentation on dynamic VT models is still a distant goal.

In summary, it can be seen that physiologically-constrained data-driven models as a way to balance the complexity of the model and physical realism, have the potential to become a good choice for studying speech kinematics. An understanding of the source mechanism is essential for modelling unvoiced sounds while an accurate description of the VT geometry is crucial in simulating the aeroacoustic source mechanism. Due to the limitations of the current experimental technologies, numerical modelling based on aeroacoustics theory and physiology-driven articulatory models therefore offer the best opportunity for synthesising high quality fricative sounds. The main objectives of the present work are summarised in the following two goals:

- Create physiologically-based 3D models of the vocal tract for speech articulation;

- Develop an aeroacoustic model and apply it to simulate flow-induced speech sound with an emphasis on fricatives.

The chapters in this thesis are organised as follows:

- **Chapter 1: Introduction.** This chapter provides a brief overview of the physiological process which leads to human speech production. Both the anatomy of various human vocal organs and their functional roles in creating speech sounds are reviewed.

- **Chapter 2: Anatomical Models of the Vocal Articulators.** In this chapter, 3D mathematical models for a system of speech organs based on the anatomy of a human subject are introduced. These models, assembled into a ‘Talking Head’, later used for simulating speech articulations for a few selected vowels and consonants, are created as finite element (FE) meshes based on the information extracted from MRI scans.

- **Chapter 3: Simulation of Articulatory Movements for Speech Sounds.** In this chapter, numerical methods are introduced for simulating realistic speech articulation based on combined measurements from MRI, EMA and video experiments, where some of the customised articulator models created in the previous chapter are dynamically wrapped by fast EMA and video recordings.
• **Chapter 4: Geometric Modelling of the Vocal Tract.** The chapter presents a method which segments the VT domain from its surrounding surfaces of the ‘Talking Head’ model, and then creates a parametric approximating its shape.

• **Chapter 5: Aeroacoustic Modelling Framework.** In this chapter, a method for simulating flow-induced sounds is introduced. In the model a two-stage FE formulation is built on the Lighthill aeroacoustic analogy for solving acoustic problems in incompressible flows for the human vocal tract.

• **Chapter 6: Aeroacoustic Modelling of Fricatives /s/ and /ʃ/.** The chapter introduces 2D aeroacoustic simulations, using the modified Lighthill equations and based on Shadle’s mechanical models of fricatives /s/ and /ʃ/ [49]. In the simulations, several geometric factors have been investigated with respect to their acoustic impacts and the resulting acoustic far-fields are compared with Shadle’s experiment measurements.

• **Chapter 7: Conclusions.** The chapter summarises the two major findings and contributions of the thesis, including the creation of a customised FE model of the human VT and the development of a new formulation for the Lighthill aeroacoustic analogy and its application for simulating the fricative sounds.

• **Chapter 8: Future Works.** In this chapter, some of the limitations of the current model is discussed and potential improvements are proposed for future studies.
The outcome of this thesis has been used in the following list of publications.

- **A three-dimensional computational model of human vocal tract systems**, in preparation for submission.

- **A flood-filling algorithm for segmentation and reconstruction of geometric models from unorganised data points**, in preparation for submission.

- **Finite element approximations of the Lighthill aeroacoustic analogy for incompressible vortex shedding from a square cylinder**, in preparation for submission.

- **From experiments to articulatory motion-a three dimensional talking head model**, presented in the 10th Annual Conference of the International Speech Communication Association 2009.

- **Aeroacoustic modelling of fricatives /s/ and /sh/**, presented in the 18th International Congress on Sound and Vibration 2011.

- **Different ways of viewing New Zealand English**, published in *Australian Acoustical Society*.

- **Stereoscopic PIV measurement of airflow in human speech during pronunciation of fricatives**, published in *Experiments in Fluids*.
Chapter 2

Anatomical Models of the Vocal Articulators

The ‘Talking Head’ assembly consists of nine articulator models (the jaw, maxilla, tongue, soft palate, pharynx wall, cheek and lips, epiglottis, larynx and upper face), all of which are constructed as finite element meshes and generated from high resolution Magnetic Resonance Imaging scans. Section 2.1 gives the introduction on how the anatomical information of the vocal articulators is obtained via means of MRI. In Section 2.2, the mathematical base for constructing parametised models is summarised and Section 2.3 presents the method for creating such mathematical models according to the information extracted in Section 2.1.

2.1 Magnetic Resonance Imaging

In 1960, Fant first used X-ray imaging to study vocal tract shapes in Russian phonemes and made considerable progress in understanding the speech production mechanism, leading to the birth of the Acoustic Theory of Speech Production [11]. Since then, X-ray based imaging techniques such as CT (Computed Tomography) have been largely abandoned in studying speech articulations after the discovery of the adverse health effects of ionising radiation. This largely halted any further large-scale VT imaging work until the introduction of Magnetic Resonance Imaging.

MRI is based on the principles of Nuclear Magnetic Resonance (NMR) which allows us the detection of internal biological structures without invasive procedures [57]. The biological basis of MRI is that the human body is primarily water and fat. Both water and fat are rich in hydrogen atoms and hydrogen nucleus are able to generate NMR signals once magnetised [57]. In the past, MRI has been shown to be a reliable imaging technique for studying speech production [58] and numerous studies of human speech articulations have been conducted with MRI applied to human speakers [59, 60, 61, 62, 63, 64, 65, 66, 67, 68] among many others. A validation study of the MRI measurement on a tubular phantom filled with air and mounted in a water tank was performed in [65]. The results showed that the errors of estimated cross-
sectional areas were in the range of 1.5% and 10%, where large errors arose where the airway was narrow and the proportion of the interface voxels in the total area was large [65].

Compared to X-ray based imaging techniques, MRI has two major advantages. First, it does not use ionising radiation and has no known harmful health effects [57], allowing researchers to obtain large amount of scans to gain insights into the relevant anatomy. Moreover, MRI can produce tomographic images in planes at any given angles without repositioning the subject, which has been shown useful in obtaining unskewed cross-sectional scans following curved VT pathways [62].

On the other hand, there are a number of downsides of MRI for scanning the VT. Among them, the biggest issue preventing real-time MRI for speech articulations is the insufficient acquisition rate due to the physical principle upon which MRI operates. In MRI systems, a strong static external magnetic field (e.g. typically at 1.5T) is applied to align objective nuclei. The overall scanning time depends on the length of each scanning sequence (the repetition time, TR), the number of pixels per image, total number of scans and other factors. A typical MRI scanning sequence starts with a detecting signal which causes resonance of the target nuclei and is generated by a much weaker and short-lived oscillating magnetic field, normally operating at radio frequencies (RF) [69]. Each affected nucleus rotates around its own axis and generates a magnetic field around the nucleus due to the fact each nucleus carries a certain amount of electrical charge [69]. After the removal of the RF pulses, the spin system gradually returns to its thermal equilibrium state under the static magnetic field, a process known as free induction decay (FID). This period is when the data is acquired and two magnetisation vector fields, the longitudinal ($T_1$) and transverse ($T_2$) magnetisation, are obtained. By convention, the external magnetic fields are applied to be vertical to the longitudinal axis. $T_1$ is defined as the time taken for 63% of the longitudinal magnetisation to recover after the initial RF signal, known as longitudinal relaxation caused by an energy exchange process known as spin lattice energy transfer. $T_2$ is the time taken for the transverse magnetisation to drop to 37% of its original magnitude, caused by the process of transferring energy between nuclei, also known as transverse relaxation. Both $T_1$ and $T_2$ are dependant on the intrinsic properties of the detected nuclei [70], and therefore pose a fundamental limit on the scanning time.

Most studies to date were performed on artificially sustained vowels and consonants due to the relatively low image acquisition rate. In [60], it took 3.4 minutes to complete thirty-seven $T_1$ weighted images by a spin-echo scanning sequence and the subjects had to perform repetitions of the same vowels during the session. In 1996, Demolin and his team were able to obtain 14 slices in a single breath in 13.8 seconds [64] with the help of a proton-contrast scanning sequence; however, there were problems with the lack of contrast at air-tissue interfaces. An attempt at dynamic imaging for continuous speech was made in [71] where the author achieved an acquisition rate of 8-9 images per second for continuous speech sentences on 2D mid-sagittal scans via a tailored spiral imaging sequence. The fast sampling rate is partially due to the multiple-pixel data acquisition during a single excitation and temporal averaging of successive
frames (whereas the true data acquisition rate is 110ms per image) [71]. Further improvement was made in [72], using a radial FLASH MRI scanning sequence [73], allowing acquisition rates of 30 images per second but at the expense of pronounced under-sampled spatial resolution (e.g. ∼ 7% sampling rate [74]). In practise, high temporal resolutions are obtained often at the cost of low spatial resolutions due to reduced SNR (signal-to-noise ratio) [64].

Another major challenge related to MRI of the VT is the lack of contrast agents in hard tissues such as the hard palate and teeth in oral cavities. As a result, MRI often produces blurred interfaces between the air space and nearby hard tissues [60, 67]. Such image defects may cause failure of some automated algorithms designed to segment the airway due to leaks into tissue regions. In [67], a teeth phantom generated from CT images, was manually registered into MRI slices.

### 2.1.1 MRI Scanning and Image Acquisition

In the current study, MRI scans were obtained on a 1.5T Siemens Magnetom Avanto model MRI scanner located at the Centre for Advanced MRI of the University of Auckland. The subject is an adult male native speaker of New Zealand English. The experiment was divided into two sessions where two different sets of scanning parameters (Table 2.1) were used. At all time, the subject was lying still in a supine position inside the magnet.

A full volumetric scan was completed first while the subject was quietly breathing with the mouth closed and tongue resting on the teeth. The aim of the scanning is to acquire accurate anatomical information of the subject’s vocal articulators, therefore high spatial resolution and high contrast among different types of tissues are preferred in the finishing images. $T_2$ weighted images have been shown to provide more differentiation amongst some types of biological and pathological tissue and are less susceptible to inhomogeneities in the magnetic field than $T_1$ weighted images [75]. Therefore, a scanning protocol with long TR and TE, was selected for the 3D volumetric imaging. In addition, the subject was instructed to breath quietly through the nose, front teeth placed together, and the tongue tip held against the lower teeth, as this practice has been shown to reduce motion artifacts in the images [76]. The soft-palate was in a lowered and relaxed state, thus allowing the nasal passage to be open.

After that, nine artificially sustained phonemes were scanned using a fast sequence so that the scanning could be completed in one breath, amongst which were four vowels (Figure 1.5), three fricatives and two stops (Figure 1.6) frozen at their pre-releasing states (i.e. no air flow). Every speech sound was acquired using the same fast scanning sequence and collected in one breath. The main benefit of such a fast scanning sequence is the avoidance of repetition of the utterance, therefore eliminating potential human errors and motion artifacts [65], however it comes at the cost of lower image resolution due to much shorter scanning time than the sequence adopted for the quiet breathing. To provide the best resolution for identifying the articulator shapes, sagittal scans were obtained, but with a wider spacing between planes. These images
Table 2.1: MRI Scanning Parameters

<table>
<thead>
<tr>
<th>View</th>
<th>Quiet Breathing</th>
<th>Sustained Articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissue contrast</td>
<td>$T_2$ weighted</td>
<td>$T_1$ weighted</td>
</tr>
<tr>
<td>TR(s)</td>
<td>3200</td>
<td>1660</td>
</tr>
<tr>
<td>TE (s)</td>
<td>585</td>
<td>9.4</td>
</tr>
<tr>
<td>FOV(mm)</td>
<td>256×256</td>
<td>200×250</td>
</tr>
<tr>
<td>In-plane resolution(mm)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>In-depth resolution(mm)</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Total No.Slices</td>
<td>160</td>
<td>20</td>
</tr>
<tr>
<td>Total scanning time(s)</td>
<td>450</td>
<td>21</td>
</tr>
</tbody>
</table>

can provide clear outlines for most of the articulators of interest, especially in regions centered around the mid-sagittal planes. Compared to the first scanning sequence, the new one has a smaller FOV, a shorter TR and TE which leads to $T_1$ weighted images, and a lower in-depth resolution (i.e slice thickness), as summarised in Table 2.1. Similar designs of the scanning sequence for VT imaging can be found in [77, 64, 78].

During each scan, the subject held the positions of all speech organs (e.g. tongue, lips, jaw and etc.) as steady as he could while maintaining a constant air stream coming out of the mouth. Compared to normal speech, the loudness of the sustained voice was significantly lowered, but it was still at an intelligible level.

2.1.2 Image Analysis and Segmentation

The acquired images were stored in DICOM format and later converted into BITMAP files for digitisation whereby the air-tissue interface is obtained by identifying points on the images in a 3D coordinate system. Each grey-scale BITMAP image has a resolution of 1012×916 pixels. The purpose of digitisation is to obtain the geometric information of various articulators at the reference state (i.e. resting breathing) for later creating parametric models. In the model the VT domain covers the airway starting from the glottis and terminating at the lips, therefore all structures surrounding this part of the airway were modelled. However, the nasal cavity was not included in the current model. The segmentation and digitisation were carried out manually in a digitiser where articulator shapes were marked directly on the MRI images. The choice of manual segmentation is mainly due to the lack of satisfactory automated algorithms. Automated segmentation algorithms may fail on the MRI images which do not provide sufficient contrast among some types of the biological tissues (e.g. hard tissues) [79], while manual segmentation can take advantage of pre-knowledge of human anatomy [80, 81] and therefore corrects possible defects in the original images. In addition, some of the articulator models are artificially truncated from their surrounding structures, such as the bottom of the tongue model which is separated from the muscular structures in the jaw floor, hence manual definition of an artificial boundary is required. However, manual entry is time consuming and demands
knowledge of physiology in the scanned area from users.

The Zinc digitiser [82] is software built from the computer graphics package Cmgui [83] developed at the Bioengineering institute in the University of Auckland. Each stack of parallel sagittal images was loaded into a volume texture in a box form, designed to fit the physical size of the imaged object. For the volumetric scans taken in the quiet breathing state, the size of the box is 256×160×256mm and the sagittal slices are aligned in the X-Z planes (as shown in Figure 2.1(b)). In digitising some organs, it is better to view them in planes other than the original sagittal ones (e.g. axial views for the larynx structure). Therefore, axial and coronal slices were calculated based on the original parallel sagittal images at 1mm intervals. Since each original image has 1mm in-plane resolution and 1mm thickness with no gaps between neighbouring slices, the reproduced axial and coronal images also have 1×1mm resolution.

The tissue boundaries were manually marked on either the axial, coronal, or sagittal images as appropriate, using spheres of 1mm diameter (same as image resolution), as shown in Figure 2.1(a) for the tongue. Most of the structures (the tongue, cheek and lips, larynx, epiglottis, jaw, maxilla, pharyngeal wall and upper face) were digitised on the axial images with assistance from sagittal views. To exclude the nasal passage, the soft palate was required to be raised, but it was in a lowered-down position in the static scans. Therefore, the shape of the soft palate was extracted from the coronal slices collected for the vowel /i:/ where it was raised. The reason for choosing the vowel /i:/ is that it provides the maximum distance between the tongue and the velum, and therefore gives the best separation of the two organs amongst all the test phonemes.

Figure 2.1: Digitisation of the tongue in the Zinc digitiser. (a) Digitising on an axial plane. (b) The volume texture (marked in lines) and the axial (reconstructed) and sagittal views of the MRI images for the rest breathing state.
2.2 Finite Element Discretisation

This idea of approximating some unknown function from a set of functions with known behavior is the fundamental concept behind the finite element (FE) method. In a geometric sense, the mathematical space can be divided into piecewise ‘elements’ each represented by a mathematical basis function. In practise, the FE basis is normally chosen to be a collection of linearly independent polynomials [84].

2.2.1 Sobolev Space

In the Galerkin finite element method, the solution is obtained in Sobolev space. Consider a mathematical function \( f(x) \) spanning a Sobolev space \( (\Omega \in \mathbb{R}^d, d = 1, 2, 3) \), defined as

\[
H^h(\Omega) = \{ f(x) \in L^2(\Omega) | \partial^m f/\partial x^m \in L^2(\Omega), m \in [0, h] \},
\]

(2.2.1)

where \( L^2(\Omega) \) is defined as all the square integrable functions over the space \( \Omega \). Integer \( m \) denotes the order of the partial derivative of functions \( f \), which also has to be square integrable over \( \Omega \). Notice that \( L^2(\Omega) = H^0(\Omega) \) [85].

Function \( f(x) \) is now approximated with \( n \) subfunctions (basis) \( f(x)_i \), which all fall into the subspace of \( \Omega \). The new mathematical space becomes

\[
H^h(n, \Omega) = \{ f(x) | f(x)_i \in H^h(\Omega), i \in [1, n] \}.
\]

(2.2.2)

For most engineering applications, it is interested in finding a solution in \( H^1 \) space [84].

2.2.2 Subdivision of the Function Space

For approximating complex functions, the function space is subdivided into piecewise elements so that instead of using a long high order polynomial, the space can be approximated using a group of lower order ones. One advantage of such an approximation is that each subdivision \( \Omega_i \) enclosed by \( \Gamma_i \) only consists of a small group of local variables, which is especially useful when defining boundary conditions. There are many ways of dividing a continuous space into piecewise elements. In a geometric sense, elements might be described as lines \( (d = 1) \), triangles and squares \( (d = 2) \) or tetrahedrons and cubes \( (d = 3) \). Each element has a finite number of local nodes where function variables are located. The following conditions may be imposed on the subdivisions \( \Omega(n) \) according to [86]:

(i) There must be a finite number of subregions;
(ii) Any two subregions \( \Omega_i \) and \( \Omega_j \) are either:

\[
\Omega_i = \Omega_j, \text{ or } \quad \Omega_i \cap \Omega_j = 0, \text{ or }
\]
\( \Omega_i \cap \Omega_j = \Gamma_i \cap \Gamma_j; \)

(iii) \( \Omega(n) = \Omega \cup \Gamma. \)

These rules ensure that there are no overlapping elements or gaps between neighbouring elements and the approximated domain fully fills the original function space. Two 2D examples of tessellations satisfying such rules are illustrated in Figure 2.2(a) using triangular elements and in Figure 2.2(b) with the quadrilateral type. Once the subdivision is done, the new domain will consist of \( n \) elements, each mathematically represented by a local polynomial defined by both the variables located at nodes and the properties of the chosen basis function.

\[ \text{Figure 2.2: Tessellation of the 2D space. (a) Subdivision of a triangle with triangular elements. (b) Subdivision of a square with quadrilateral elements.} \]

Figures 2.2(a) and (b) also represent two types of tessellation strategies: unstructured and structured meshing. Structured meshes are the ones having fixed topological relations among all the elements. In other words, each interior node shares the same number of neighbouring elements (e.g. 4 elements per interior node, as shown in Figure 2.2(b)). Unstructured meshes do not have such valency constraints and can be made of triangles in 2D (Figure 2.2(a)) or tetrahedrons in 3D. The choice of using a structured or unstructured mesh depends upon the domain. While unstructured meshes provide more flexibility in subdividing complex geometries, structured meshes may offer some benefits by employing high order polynomial functions (e.g. cubic Hermite).

### 2.2.3 Assembly of the Finite Element Space

For each subdivision of the global domain \( (\Omega_n) \), the local variable field may be constructed as a Lagrange polynomial series of order \( n - 1 \) where the coefficients are derived based on the
specified point values of the functions \( \{x| x_i, i \in [1,n]\} \) and the variables are defined as the local scaling factors \( (\xi \in [0,1]) \) based on either of the length, area or volume of the element. To calculate the coefficients, \( \{\phi(\xi)|\phi_j, j \in [1,n]\} \) need to be independent interpolation functions of \( x_i \) so that a 1D polynomial function in subspace \( \Omega_m \) can be approximated as

\[
x(\Omega_m) = \sum_{i,j=1}^{n} \phi_j x_i, \quad (2.2.3)
\]

where \( \phi \) is defined as

\[
\phi = \begin{bmatrix}
\alpha_{11} & \cdots & \alpha_{1n} \\
\vdots & \ddots & \vdots \\
\alpha_{n1} & \cdots & \alpha_{nn}
\end{bmatrix}
\begin{bmatrix}
1 \\
\vdots \\
\xi_{n-1}
\end{bmatrix}.
\quad (2.2.4)
\]

Here, if \( \phi_i \) is only non-zero at \( x_i \), this type of vector function is referred to as a Cardinal basis function and, given the values of \( \xi \) for each node (also referred to as nodal parameters/values), all coefficients \( \alpha_{ij} \) can be resolved by solving a systems of equations

\[
\begin{bmatrix}
\xi^0 \\
\xi^1 \\
\vdots \\
\xi^n
\end{bmatrix}
\begin{bmatrix}
\alpha_{11} & \cdots & \alpha_{1n} \\
\vdots & \ddots & \vdots \\
\alpha_{n1} & \cdots & \alpha_{nn}
\end{bmatrix}
\begin{bmatrix}
1 \\
\vdots \\
0
\end{bmatrix} = \begin{bmatrix}
1 \\
\vdots \\
0
\end{bmatrix},
\quad (2.2.5)
\]

As long as the left hand side matrix is invertible, the resulting basis function shall always be valid for the polynomial. In a multidimensional space, for length-based basis functions (e.g. quadrilateral and hexagonal elements), the tensor product can be used to assemble the new basis from a lower-dimensional space via

\[
\phi(\xi_1,\xi_2) = \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_i(\xi_{1,i})\phi_j(\xi_{2,j}).
\quad (2.2.6)
\]

However this is not true for triangular or tetrahedral type elements, in which case basis functions are area/volume based and can only be calculated according to Equation (2.2.4) [87].

It can be seen immediately that such polynomials of order \( n - 1 \) fall into \( H^{n-1} \) space as the function is differentiable up to the \( (n - 1) \)th order. On the other hand, across a neighbouring element, such differentiation may yield different derivative values at the shared borders [88]. For this reason, Lagrange type functions are \( C^0 \) continuous basis. To achieve higher order continuities, conditions need to be imposed on the derivatives of the function space, which naturally leads us to ‘Hermite’ polynomials. Equation (2.2.3) is rearranged into the form

\[
x(\Omega_m) = \sum_{i,j=1}^{n} \psi_j^0(\xi) x_i + \sum_{i,j=1}^{n} \psi_j^1(\xi) \frac{\partial x_i}{\partial \xi},
\quad (2.2.7)
\]
where the $\psi^0_j$ and $\psi^1_j$ represent the basis functions for the $0^{th}$ and $1^{st}$ order derivatives of the function at the $j^{th}$ node of the element, respectively. The relations defined in Equations (2.2.4) and (2.2.5) may be applied for resolving the coefficients. Only this time, a new matrix $\partial \psi / \partial \xi$ needs to be added to complete the system. In cubic Hermite elements, $C^1$ continuity is achieved for the domain $\Omega$, while obtaining a cubic polynomial space within each element $\Omega_m$.

In the case of two adjacent cubic Hermite elements significantly different in size, the curvature of the smaller element could be severely biased by the derivatives of the neighbouring larger ones through the shared derivative values at common nodes [89]. To address this problem, arc length-based scaling factors are introduced and the nodal derivative of a 1D cubic Hermite element is rewritten using the chain rule

$$\frac{\partial x}{\partial \xi} = \frac{\partial x}{\partial s} \frac{\partial s}{\partial \xi}, \quad (2.2.8)$$

where $\partial s / \partial \xi$ are stored on local elements and the arc length $s$ can be calculated as

$$s = \int_0^1 \sqrt{\left(\frac{\partial x}{\partial \xi}\right)^2} \, d\xi, \quad (2.2.9)$$

while normalised derivatives $\partial x / \partial s$ are preserved across neighbouring elements [90].

2.3 Least-Squares Fitting

Following the digitisation process, geometric information is extracted from the MRI scans, in the form of discrete data sets, as described in Section 2.1. For creating FE meshes of the vocal articulators, a least-squares fitting algorithm is employed to minimise the projection distances between the segmented data points and predefined FE surfaces. It involves three basic steps: creation of an initial mesh, calculation of data projections to the surface and optimisation of nodal parameters, which will be demonstrated below in the fitting of the tongue and cheek-lips models while more detail of the method and its applications on more biological structures can be found in [91, 78, 92, 89].

2.3.1 Initial Linear Meshes

Like any optimisation problem, the technique requires an initial mesh outlining the shape of the object, acting as a starting point. This can often be done by manually connecting selected data points (i.e. nodes) into linear elements. There are many ways that an initial mesh can be constructed, however certain conditions need to be met in order to reach a good quality finished product.

The number of nodes and elements in the mesh needs to be decided by the number of DOFs (degrees of freedom) provided by the data cloud so that the optimisation problem will not be
seriously under-defined or over-constrained. Although subsequent procedures, like converting elements into higher order basis (i.e. linear Lagrangian to cubic Hermite) or refining the existing element, can bring more DOFs onto the mesh, care needs to be taken in order to make sure that the data projections are more or less uniformly distributed across different surface patches. In the case of the tongue, as shown in Figure 2.3(b), for each surface facet (4-noded cubic Hermite element), there are 48 DOFs on the mesh and between 12-20 data projections, giving 36-60 specified DOFs. As the data digitisation is reasonably evenly-spaced across the tongue surface, the tongue model is made of elements almost homogeneous in size. While in the cheek-lips case, there are initially over 40 data projections per surface patch for some elements (Figure 2.3(c)), so refinements have been introduced to achieve a better fitting quality in the finish product (Figure 2.3(d)).

All articulator meshes are constructed using cubic Hermite type basis functions, therefore the meshes have to be made in structured forms (e.g. quadrilateral elements in 2D and hexagonal elements in 3D, as discussed in Section 2.2.2). The initial mesh for the tongue is made of 10 trilinear elements in arrangements of two columns of five elements each, aligned in the sagittal direction, as shown in Figure 2.3(a). Such arrangements produce local axes approximately aligned with some of the major muscle groups (e.g. styloglossus, transverse and vertical muscles) hence likely in-line with the directions of principle strains during the tongue deformation [93]. Forty trilinear FE elements are grouped in a $4 \times 10 \times 1$ arc form which brings symmetrical topology about the frontal and sagittal axes. In this way, the local coordinates $\xi_1$ and $\xi_2$ form a plane covering the lower half of the face while $\xi_3$ is pointing to inside the tissues. One advantage of this topology for the cheek-lips is that all the data projections are landed on the planes spanned by $\xi_1$ and $\xi_2$ as only external surfaces of the tissue are digitised, but not any interior space. Meshes of similar design have been successfully applied in biomechanical models of facial tissues in [94, 95].

As the optimisation process does not change the topology of the mesh but merely deforms it, those topological features in the target geometry should be reflected in the initial mesh. As shown in Figure 2.3(c) for the cheek-lips mesh, a mouth opening is created by breaking up affected elements and by specifying multiple nodal derivatives in the transverse direction at the corners nodes (highlighted in Figure 2.7(a)) during the optimisation process.

### 2.3.2 Data Projection to Parametrised Surfaces

Given a discrete data set $\{z_i, i \in [1, n]\}$ defined in $\mathbb{R}^3$ and a parametric surface $(f(\Omega), \Omega \in \mathbb{R}^3)$ constructed with cardinal basis functions defined in Equation (2.2.3), a mapping between the two spaces can be established through the projection method and the mathematical function can be calculated as

$$F_e = \sum_{i=1}^{n} ||(z_i - x(\xi))||^2,$$  \hspace{1cm} (2.3.1)
Figure 2.3: Steps of the least-squares fitting algorithm on the two articulator models, where the data projections are represented by yellow arrows. (a) The initial trilinear tongue mesh. (b) The refined and fitted tricubic tongue mesh. (c) The initial linear cheek-lips mesh. (d) The fitted tricubic cheek-lips mesh.
where errors are defined as the $L_2$ norm of the data points and their corresponding projection points $x(\xi)$ on $\Omega$.

The projection point $x(\xi)$ should have the least $L_2$ norm over the domain $\Omega$, therefore it may be evaluated by differentiating the above function (i.e. errors) with respect to the element local coordinate $\xi$ and equating to 0,

$$\frac{\partial (F_e)}{\partial \xi} = \sum_{i=1}^{n} 2(x(\xi) - z_i) \frac{\partial x(\xi)}{\partial \xi} = 0.$$  \hspace{1cm} (2.3.2)

As can be seen, with the exception of 1st order elements, this operation will lead to a nonlinear function of $\xi$, which requires iterative solvers like the Newton-Raphson method to find local optima [91]. In Figures 2.3(a) and (c), the data points segmented from the previous section were projected onto the external surface of the linear elements, in the form of arrows whose lengths represent the error.

In the case of fitting volume elements where there are both interior and exterior elemental faces, it is beneficial to single out external facets together with associated DOFs from the original mesh and leave interior DOFs out of the fitting equations [89]. Volumetric meshes are needed for both soft tissues (e.g. tongue, cheek-lips, etc.) and hard tissues (e.g. jaw, epiglottis, etc.) in order for implementing physiological constraints (e.g. volume conservation, contacts modelling, etc.) during the animation stage. Therefore, it is wise to intentionally minimise the number of interior nodes on the initial meshes. As a result, there is no interior nodes created in the initial mesh for the newly created articulator models. However, for the cubic Hermite type basis functions, there are always nodal derivatives related to the interior space, hence these values are left untouched during the least-squares fitting process.

### 2.3.3 Optimisation of Nodal Parameters

Once all projection points have been evaluated, objective functions can be established by differentiating function $F_e$ with respect to each nodal parameters $x_j$ and equating to zero, as

$$\frac{\partial F_e}{\partial x_j} = \sum_{i=1}^{n} 2w_i(\phi_i)(x_i - z_i) = 0,$$  \hspace{1cm} (2.3.3)

where there shall have a linear system of equations for nodal variables $x_i$ and $w_i$ is a weighting function for each data point. However, when using cubic Hermite elements with arc-length based derivatives, the objective function (2.3.3) becomes nonlinear due to the evaluation of the arc-length $s$ which involves the use of the values of $x_i$. In such cases, $s$ may be treated as a constant and multiple fittings may be used, in which each uses a renewed value of $s$ based on the updated solution field. This simplifies the formulation to a linear system of equations which is relatively easy to solve [89].

The projections, shown as arrows in Figures 2.3(b) and (d), were minimised in this way
while all elements were transformed into the cubic Hermite type in order to provide more DOFs to the geometry. The cheek-lips mesh was further refined in the span-wise direction and the final product contains 80 tricubic elements with dual nodal versions specified for the four corner nodes of the lips. In addition, the nodal coordinates are fixed on the bottom face of the tongue and the top and bottom planes of the cheek-lips mesh, where both models are artificially separated from the related tissues.
2.3.4 ‘Talking Head’ Assembly

In total, the ‘Talking Head’ assembly consists of nine individual FE articulator meshes based on the information acquired from the statics MRI scans, covering the VT domain from the glottis end to the mouth opening. These models are constructed mainly based on their tissue types as hard tissues such as bones (maxilla and mandible) and cartilages (epiglottis, soft palate and larynx) and soft tissues including structures made of both muscular and connective tissues (tongue, pharynx wall, upper and lower faces). In addition, arrangements are made so as to facilitate the kinematic simulations which will be introduced in the next chapter, where hard tissues are treated as rigid bodies while soft tissues are simulated as deformable materials for speech articulation. Independent active or passive articulators have been identified based on both past literature [4] and the observations made on those MRI scans of sustained phonations.

![Figure 2.4: The maxilla model. (a) The FE mesh of the mandible. (b) Comparison between the model and the MRI scans during the rest breathing state.](image)

Figure 2.4: The maxilla model. (a) The FE mesh of the mandible. (b) Comparison between the model and the MRI scans during the rest breathing state.

![Figure 2.5: The mandible model. (a) The FE mesh of the maxilla. (b) Comparison between the model and the MRI scans during the rest breathing state.](image)

Figure 2.5: The mandible model. (a) The FE mesh of the maxilla. (b) Comparison between the model and the MRI scans during the rest breathing state.

The FE meshes of the mandible (jaw) and maxilla, shown in Figures 2.4(a) and 2.5(a) respectively, were originally created in a project of modelling the human skull in [92]. Both the
mandible and maxilla meshes were customised to fit the MRI scans of the test subject using the host mesh fitting technique. The technique will be given more introduction in the next chapter. Additional modifications, including removing the wisdom teeth and manually adjusting the size of the teeth, have been made in the teeth area to match subject’s dental profile, as shown in Figures 2.4(b) and 2.5(b) respectively.

The tongue model occupies a large part of the oral cavity and extends into the pharynx. In the current implementation, as shown in Figure 2.6(a) and compared to the MRI scans in Figure 2.6(b), only the main body of the tongue is included in the mesh while detail of its intrinsic and extrinsic muscles are ignored as they are not directly involved in defining the VT shapes.

![Figure 2.6: The tongue model. (a) The FE mesh of the tongue. (b) Comparison between the tongue model and the MRI scans during the rest breathing.](image)

The cheeks and lips are merged into one volume mesh (Figure 2.7(a)), surrounding the exterior surface of the mandible and the maxilla, as shown in Figure 2.7(b). The mouth opening is created with the help of the multiple nodal versions specified at the nodes located at the corners of the lips (highlighted in Figure 2.7(a)). There are two sets of values of the transverse derivatives specified at corner nodes, which enables different curvatures on the upper and lower lips while maintaining $C^0$ continuity at the same locations.

Since the nasal cavity is not included in this study, the upper half of the facial structure is irrelevant to the VT modelling. For this reason, a bicubic upper face mesh (Figures 2.8(a) and (b)) is created to complete the facial model. The bicubic mesh of the upper face can be viewed as a topological extension of the frontal face of the cheek-lips mesh, where the curvatures on the shared border with the cheek-lips mesh are conserved by replacing the related nodal values (i.e. coordinates and associated derivatives) with the ones from the upper edge of the cheek-lips mesh after the fitting.

The soft palate is in an elevated position in all the sustained articulation scans (Figures 1.5 and 1.6), decoupling the nasal cavity from the VT domain. Therefore a tricubic mesh (Figure
Figure 2.7: The cheek-lips model. (a) The FE mesh of the cheek-lips where the nodes with multiple nodal derivatives are identified in red. (b) Comparison between the cheek-lips model and the MRI scans during the rest breathing state.

Figure 2.8: The upper face model. (a) The FE mesh of the upper face. (b) Comparison between the upper face model and the MRI scans during the rest breathing state.

2.9(a)) is made according to the data digitised on the reconstructed coronal planes from the images of the vowel /i:/, as shown in Figure 2.9(b). The structured mesh is created in a way so as to align the elements on the sagittal planes, similar to the 2D biomechanical model presented in [96]. Notice that the soft palate created in this way, is in a deformed state due to primarily the contraction of an extrinsic muscle group (i.e. veli palatini) [97].

A large portion of the posterior part of the VT is constituted by layers of pharyngeal muscles. As the current MRI images can not provide enough contrast in order to identify different muscle groups (e.g. pharyngeal constrictors and others), only the air-tissue interface can be clearly identified on the images. As a result, the decision was made to model this part of the VT as a bicubic surface mesh (Figure 2.10(a)), starting from the back end of the soft palate, finishing on the top of the larynx and extending to the lateral side of the pharynx, as shown in Figure 2.10(b) along with the MRI scans. Such a model may not be sufficient for a biomechanical simulation driven by the muscle activation, but it is still capable of modelling the narrowing
effect of the VT domain due to lateral constrictions by the pharyngeal constrictor muscles; this phenomenon has been observed in [98] and simulated on a mathematical model developed for speech synthesis [99].

In [100, 101], it was discovered that the epiglottis is an active, independent and acoustically significant articulator. This view is supported by the observations made on some of the images of sustained sounds in the MRI study, such as the cases presented in Figures 1.5(c) and (d), where it can be seen that the epiglottis moved backward and flipped downward by the contact with the tongue during the two vowels /a:/ and /aː/. Therefore the epiglottis (Figures 2.11(a) and (b)) is created as an independent volume mesh in the ‘Talking head’ model. The mesh is sitting in an upright posture on top of the larynx and between the backwall and the tongue in
the rest breathing state.

Figure 2.11: The epiglottis model. (a) The FE mesh of the epiglottis. (b) Comparison between the epiglottis model and the MRI scans during the rest breathing state.

Constructing a detailed physiological larynx model is beyond the scope of this project, so a simplified geometric model, combining the hyoid bone, the thyroid, arytenoid and cricoid cartilages, and associated ligaments and muscles, is represented by a cylindrical volume mesh, stretching from the bottom of the tongue to the region of glottis opening, as shown in Figure 2.12(a) and compared to the MRI scans in Figure 2.12(b). As can be seen in Figures 1.5 and 1.6, the positions of glottal opening can be roughly identified on the images, however those substructures like vocal folds and vocal cords can not be accurately segmented from the images due to insufficient contrast and resolution. In the current implementation, the glottis is fixed at an opening posture in the larynx model. Although complex physiological dynamics such as vocal fold vibrations can not be explicitly simulated, the current model may still be considered as adequate for representing the VT shape in this part of the airway during human speech production for the following reasons:

1. The dynamics of the muscular structure mainly affect the property of the vocal fold vibration [102, 2], but not much the overall shape of the airway [103, 104];

2. The main movement of the larynx tube can be described by a translation in the longitudinal and/or horizontal directions [105, 106, 26];

3. The phonation of unvoiced sounds (e.g. voiceless consonants) does not involve glottal vibrations [11].

A summary of the statistics of these articulator models is provided in Table 2.2 and various views of the assembly are shown in Figure 2.13. The RMS errors on most of the articulator models are kept close to 1mm (i.e. the MRI resolution) so that no additional errors are introduced to the geometric modelling.
Figure 2.12: The larynx model. (a) The FE mesh of the larynx. (b) Comparison between the larynx model and the MRI scans during the rest breathing state.

Table 2.2: Nine Articulator Meshes Created for the ‘Talking Head’ model.

<table>
<thead>
<tr>
<th>Models</th>
<th>Basis Type</th>
<th>No. Nodes</th>
<th>No. Elements</th>
<th>RMS error (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandible</td>
<td>Tricubic</td>
<td>603</td>
<td>292</td>
<td>1.01</td>
</tr>
<tr>
<td>Maxilla</td>
<td>Tricubic</td>
<td>580</td>
<td>266</td>
<td>1.47</td>
</tr>
<tr>
<td>Cheek-lips</td>
<td>Tricubic</td>
<td>210</td>
<td>80</td>
<td>1.16</td>
</tr>
<tr>
<td>Upper face</td>
<td>Bicubic</td>
<td>68</td>
<td>48</td>
<td>0.55</td>
</tr>
<tr>
<td>Tongue</td>
<td>Tricubic</td>
<td>36</td>
<td>10</td>
<td>1.19</td>
</tr>
<tr>
<td>Soft palate</td>
<td>Tricubic</td>
<td>90</td>
<td>32</td>
<td>1.15</td>
</tr>
<tr>
<td>Pharynx wall</td>
<td>Bicubic</td>
<td>117</td>
<td>96</td>
<td>0.96</td>
</tr>
<tr>
<td>Larynx</td>
<td>Tricubic</td>
<td>240</td>
<td>108</td>
<td>0.67</td>
</tr>
<tr>
<td>Epiglottis</td>
<td>Tricubic</td>
<td>30</td>
<td>8</td>
<td>0.65</td>
</tr>
</tbody>
</table>
Figure 2.13: Assembly of the ‘Talking Head’. (a) The right side view, (b) the left side view, (c) the front view, (d) the back view, (e) the top view, (f) the bottom view.
2.3.5 Discussion

As discussed in Section 2.3.1, it is preferable for all the DOFs in the mesh to be fully constrained by the data to extract as much information from the data set as possible. Therefore it is desirable to have a fully matched system of equations for the variables in Equation (2.3.1). However, the problem is better solved as an optimisation process due to imperfections in both the data set and the initial mesh.

When calculating projections for unevenly sampled data clouds, some elements may have more projection points than local DOFs while others could have no projections from any data points at all. In the first case, one of solutions is to refine those overly-specified elements or convert them to use a higher order basis function (e.g. transform from a Linear Lagrange to a cubic Hermite basis), such as the procedures done for both the tongue and cheek-lips meshes. At the same time, some additional constraints need to be provided in order to prevent excessive deformation and other unrealistic physiological responses (e.g. volume changes in incompressible tissues) due to insufficient sampling [78]. One way to do this is adding the Sobolev smoothing function \( F_s \) to the objective functions, Equation (2.3.1) is modified into

\[
F_e = \sum_{i=1}^{n} |(z_i - x(\xi))|^2 + F_s,
\]

(2.3.4)

where the smoothing function is defined as a series of weighted Sobolev norms designed for minimising the arc-length, arc-curvature, area and volume of the element. Discussions of the individual components of the Sobolev smoothing function can be found in [89, 91]. Increasing the dimensionless weights of each norm against the error projections tends to constrain the amount of deformation during the fitting at the expense of larger projection errors, while unrealistic deformations may occur due to the defects in the data clouds if no smoothing function is specified. In practice it is often up to the user to find a balance point between accuracy and consistency.

It has been found that the Sobolev smoothing functions are useful in all the least-squares fitting examples created in this project. In Table 2.3, a summary of the Sobolev coefficients, corresponding to the norms \( (||\partial x_i/\partial \xi_1||^2, ||\partial x_i/\partial \xi_2||^2, ||\partial^2 x_i/\partial \xi_1^2||^2, ||\partial^2 x_i/\partial \xi_2^2||^2, ||\partial^2 x_i/\partial \xi_1 \partial \xi_2||^2, \) \( i \in [1,3] \)) respectively, is listed for the seven articulator meshes. On the cheek-lips mesh, larger weights have been specified for the elements in the lips region than those in the cheeks region, primarily due to the close proximity between the upper and lower lips.

The least-squares fitting method for approximating parametric surfaces from discrete data clouds is very efficient, but the computational cost for the method rises dramatically owing to the matrix inversion operation during the optimisation procedure and is further complicated by the nonlinearities arising at the stages of calculating data projections and optimising nodal parameters. Additionally, the creation of the initial mesh also demands considerable human intervention. In practice, the least-squares fitting technique is rarely used for large and complex...
Table 2.3: Sobolev Smoothing Coefficients used in the Least-Squares Fitting

<table>
<thead>
<tr>
<th>Model</th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
<th>$w_4$</th>
<th>$w_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongue</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.01</td>
</tr>
<tr>
<td>Cheek-Lips (Cheeks region)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Cheek-Lips (Lips region)</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Upper Face</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>Soft Palate</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Backwall</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Larynx</td>
<td>0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Epiglottis</td>
<td>0.005</td>
<td>0.005</td>
<td>0.01</td>
<td>0.01</td>
<td>0.001</td>
</tr>
</tbody>
</table>

geometries and is usually constrained to models containing less than a few hundred elements. A close examination of the algorithm reveals that the global optimisation process can be in fact divided into individual elemental variables. As for calculating the projection errors, instead of Equation (2.3.3) for each elemental space, a preliminary selection algorithm based on a contouring method can be used for initial pairing the data points to the nearest facets. The global space is then divided up into a number of patches, each for the independent optimisations of nodal parameters, hence reducing the problem size. After the optimisation, the shared nodal parameters are averaged across different patches while the global space is re-assembled. This concept will be demonstrated on 2D and 3D examples in Chapter 4.

Compared to other 3D articulatory models in the literature, the ‘Talking Head’ has two distinctive advantages, being patient-specific and employing a high-order FE basis. Other models [26, 18, 107] collect the anatomical information from multiple sources, and as a result they can not be used in applications needing customised anatomy, such as some of the surgical applications [14] as well as being useful in the validation of mathematical models. Moreover, the base models created for the ‘Talking Head’ could also be easily customised to suit anatomical data collected from other individuals. The approach using the host-mesh fitting technique has yielded many good results in the customisation of biological structures [89]. The high-order interpolation function (cubic Hermite) used in constructing each of the articulator models, has been shown to produce a better approximation to the biological structures than linearly interpolated ones [79], despite requiring less elements [89].

In the current state of the ‘Talking Head’, each articulator model contains accurate geometric descriptions for the air-tissue interfaces which define the physical space of the VT, while neglecting details such as intrinsic and extrinsic muscle groups. As a result, these models were not implemented as muscle-actived biomechanical types in the articulatory simulations presented in the next Chapter, but rather their speech kinematics were calculated via methods based on the experimentally measured data. In the future, the articulator models may become more biomechanically oriented by adding the material properties of the muscle fibres, such as the ones presented in [92] for the jaw, in [93, 108] for the tongue and in [94, 95] for facial modelling, to supplement physiological constrains.
Apart from the absence of the underlying internal structures of the articulators, there are several other major limitations of the ‘Talking Head’ model with respect to speech modelling. The nasal passage has not been included, which reduces the model’s ability to simulate nasal sounds in which the resonance and anti-resonance frequencies are important in the resulting spectra. This is partially due to the insufficient resolution of the MRI scans (1x1x1mm) for differentiating the narrow airways in the nasal cavity. Other experimental tools like CT are preferable to MRI for scanning nasal passages because they have a better spatial resolution and higher definition of the air-tissue interface [109]. Similar issues exist with regards to imaging the vocal folds, where the MRI scanning sequence is unable to produce sufficient contrast and resolution for segmenting the soft tissues from their surrounding cartilaginous structures. For the geometric modelling of the vocal folds, other imaging tools such as CT [110] or MR microimaging [111] may be employed.

Another limitation of the current method is with the time consuming manual segmentation of the MRI scans, which demand high levels of knowledge of the human head and neck anatomy. If accurate and reliable automatic segmentation methods can be developed, these will enhance the ability of the ‘Talking Head’ to be used in more practical applications as well as becoming a more useful research tool. Automatic delimitation of articulator surfaces is a highly challenging task for the VT domain, owing to numerous factors including the inter-connected tissues, insufficient image resolution/contrast, and noise in the signals amongst many others. Despite the difficulties, atlas-based supervised automatic/semi-automatic algorithms seem to have become standard tools for exploiting prior knowledge in the segmentation of medical images in recent years [112, 113, 114], where a reference geometry (atlas) is first obtained, then a mapping algorithm is established for registering the image data to the reference. It may also be possible to use the current articulatory models as the reference frame in an atlas-based segmentation algorithm.

2.4 Summary

Nine customised 2D and 3D FE models of the vocal articulators were created according to the digitised shape extracted from the MRI images. High-resolution MRI scans were taken during the rest breathing state. Manual segmentation was performed and the results are a group of discrete point clouds created on the surface of each articulator. A least-squares surface fitting algorithm and the host-mesh fitting technique were applied to deform the existing models to match the digitised articulator shapes. The resulting ‘Talking Head’ covers the vocal tract from the glottal end to the opening of the mouth, excluding the nasal branches.
Chapter 3

Simulation of Articulatory Movements in Speech Sounds

Electromagnetic articulography (EMA) and video data of speech was collected and applied to animate four of the articulator models (Section 3.1). Three types of articulatory motion (rigid body in Section 3.2; deformable in Section 3.3 and contact in Section 3.4) have been identified and four articulatory models (the jaw in Section 3.5.2; the tongue in Section 3.5.3, the cheek-lips in Section 3.5.4 and the epiglottis in Section 3.5.5) are simulated separately. In total three vowels and 2 consonants have been simulated using the ‘Talking Head’ model and the simulated articulator shapes are compared with their corresponding configurations in sustained phonation in the MRI scans/video images in Section 3.5.

3.1 Data Collection and Processing

3.1.1 Electromagnetic Articulography

Like MRI, EMA is a technique widely adopted in studies of human speech articulation, involving the use of electrical signals induced by magnetic fields [115, 116]. However, unlike MRI it relies on miniature sensors rather than biological materials to generate signals. The physical basis of EMA is that alternating magnetic fields can induce currents on affected sensors while the strength of the current is proportional to the intensity of the flux whose size is inversely proportional to the distance between the sensor and the transmitter. Therefore the spatial location of each sensor can be calculated as a function of its signal strength measured by receivers in different locations [115].

The first EMA study was performed by Sonoda in 1974 [115], in which a permanent magnetic rod was fixed on the tongue and two sensors set outside the head and measuring the horizontal and vertical movements of the rod. This was followed by Giet in 1977, who built a system which allows the tracking of 2D movements of 4 sensors simultaneously [117]. In 1987, Schönle and his co-workers demonstrated the use of magnetically induced current to track sensor positions and
orientations, which forms the basis of many modern day EMA devices [118]. The measuring concepts of EMA can be found in detail in [119]. All early EMA devices were designed to take measurements in 2D space and often required subjects to wear electrical helmets (magnetic transmitters). A 3D EMA device (AG500) was commercially launched by Carstens in Germany. It is capable of measuring the sensor movements in 3D (e.g. x,y,z coordinates) as well as the orientations (e.g. two phase angles) [120].

Over the years, EMA has been shown to produce reliable and accurate articulatory measurements [121] and has been employed in many linguistic [122, 123, 124, 125, 126] and clinical studies [127]. That being said, there are still some noticeable drawbacks, especially the intrusive manner of placing sensors on the subject’s vocal organs, which may lead to unnatural speech behaviour [123]. A detailed comparison among different measuring techniques for studying speech articulation can be found in [128]. Among those using the results obtained by EMA to reconstruct articulator geometries, [129] demonstrated a method for determining mid-sagittal contours of the tongue from the EMA sensors placed on the surface. The model has an estimated error within 1mm. Following that, Engwall presented a 3D tongue model driven by arbitrarily defined parameters derived from EMA measurements [40] in 2003.

### 3.1.2 The EMA-Video Experiment

The experiment was conducted in the laboratory of the Department of Food Science at the Massey University in Auckland. The EMA machine is the Carstens model AG500, which is equipped with 12 HS220 EMA sensors, recording signals at 200Hz, and an analog audio channel, recording at 16kHz. The machine has a measuring area of a 300mm diameter sphere and a measuring accuracy of approximately 0.5mm [130]. To gain more information on the kinematics of multiple articulators, video recordings were also made to trace the facial movements in parallel to the EMA which was mainly applied in the subject’s oral cavity. The video files were acquired by two Sony Digital Handycams at 30Hz and have a image resolution of 720x576 pixels. Pieces of 5mm diameter circular silver tapes were placed at various facial locations as visual markers. The arrangement of the experiment is shown in Figure 3.1.

All 12 EMA sensors were used in the experiment. The sensors were glued onto a number of designated locations on the articulators as illustrated in Figure 3.2(a). Visual markers were placed on the lips, chin and cheeks (Figure 3.2(b)). The one attached to the middle of the upper lip was co-registered with an EMA channel. The placements of the tongue and lips sensors/markers have taken considerations of the experiment performed in [39], while adding two more EMA sensors to the tongue blades for measuring the tongue width.

The same subject as in the MRI scans participated this EMA-Video experiment. The data consists of a list of English words and sentences focusing on a range of vowels, stops, fricatives, nasals, liquids and consonant clusters. They were chosen so that

1. All phonemes are in the initial, medial and final position;
2. Combinations of initial, medial and final consonants are paired with each of the vowels;

3. Consonant clusters are mixed with a wide range of vowels.
Each word and sentence was recorded in an individual file and repeated twice consecutively. In total, 208 English words and 13 sentences were recorded. A complete list is given in Appendix A.

3.1.3 Data Processing

Before applying the measured data from the above experiment to produce articulatory motion on the ‘Talking Head’ model, the collected raw data has been further processed. The procedures are illustrated in Figure 3.3.

![Data processing procedure diagram](image)

Figure 3.3: Data processing procedure.

Data Optimisation

The EMA signals were generated by 6 transmitters placed on the measurement cube. To determine the coordinates of the sensor in a 5D space (e.g. three coordinates and two orientations), a non-linear formulation needed to be solved, which can be described as

\[ f(x) - \Theta = 0, \]  

(3.1.1)

where \( x \) is the position of the transmitter, \( f(x) \) is a non-linear function of \( x \) (e.g. a magnetic field model) and \( \Theta \) is the measured sensor signal [131]. An iterative algorithm like the Newton-Raphson technique is needed to solve this nonlinear system. The calculation was performed by the software CalcPos that came with the EMA machine from Carstens [132].

Frame Transformation

After the optimisation, the calculated spatial data set needs to be normalised with respect to the head movements. Three EMA sensors (the left ear, right ear and upper incisor) are chosen to represent the head coordinate system and their movements are zeroed. The normalisation process was carried out by the Carstens software (NormPos) [131].
The data acquired from the EMA is presented in the coordinate system of the device and needs to be transformed into the coordinate system of the computer model. Both coordinate systems are of rectangular Cartesian type, so the task was to match their origins and orientations. This was done by exporting the reference EMA frame at the rest breathing state into a visualisation package (Cmgui) together with the model. As the ‘Talking Head’ model is also created at the rest breathing state from the same subject, those anatomical locations to which the EMA sensors were attached are used as guide to align the two coordinate systems, as shown in Figure 3.4. The data was rotated and translated to match their locations and in this way the transformation matrix was acquired, which was later applied to the rest of EMA data set.

![Figure 3.4: The modelling frame and the reference state of the EMA sensors (1,2,3,5,6), marked as spheres.](image)

**Resampling**

The original EMA signal is recorded at 200Hz. It is believed that for supra-glottal organs (e.g. tongue, lips, jaw, etc) involved in speech production, the minimum time taken to produce one period of a rapid repetitive movement is approximately 150-200ms [2]. Therefore, those high frequency signals were most likely contributed by the noise in the measurements. Resampling was applied to the EMA data set with a cutoff frequency at 30Hz which is the same rate used by the two video cameras, as shown in Figure 3.5 for the longitudinal movements of the tongue tip sensor during the word ‘hoard’.

**Data Synchronisation**

Two sets of video images (front and right side views) were used to provide information on the lips. The video frames share the same sampling frequencies with the resampled EMA data.
Figure 3.5: The resampling of the EMA signal of the tongue tip sensor in the word ‘hoard’. (a) The original measurements in the x-direction are marked in circles while the resampled signals are presented as a red line, (b) the audio signal recorded by the EMA machine.

The task was to synchronise the images with the EMA data. This was done with the aid of the audio files from both measuring devices. Spectra of the two audio files have been used to align the onset of the speech signals. As shown in Figures 3.6(a) and (b), two spectra for the word ‘hoard’ measured by both the EMA audio equipment and the Camera B, where it can be seen that video frames lag approximately 1 second behind the EMA signals. After the offset, the video frames were exported as images in bitmap format.

Figure 3.6: The sound spectra of the word ‘hoard’ from the EMA and Video Camera B, filtered by a Hanning window of 256 samples and 50 percent overlapping. (a) EMA. (b) Camera B.
Image Segmentation

The shape of the lips was measured on the images. The images were loaded into Cmgui along with the ‘Talking Head’ model and scaled to match the physical size of the model as shown in Figures 3.7(a) and (b). Relative positions between each visual marker with respect to the reference marker (i.e. Marker 3 in the middle of the upper lip co-registered with an EMA sensor) were measured. The y and z coordinates were taken from the front views while the side view provided the x coordinates. Symmetrical movement of the lips is assumed in the medial direction. With this spatial information plus the EMA measurements of the reference marker, the locations of these labial markers could be fully recovered in the ‘Talking head’ frame.

Figure 3.7: The reference state of the cheek-lips mesh and corresponding video frames scaled to match the physical size of the model. (Images are reproduced with permission from the participant.) (a) The frontal view. (b) The side view.

3.2 Rigid Body Transformation

In the current model, the jaw is set to perform rigid body motion (i.e., rotations and translations) during all the speech utterances. The Carsten model AG500 system provides both positional and rotational information of each sensor. As a result, 4 sets of 3D coordinates can be extracted from the two jaw sensors at each sampling instance, which allows us to derive Euclidean transformation tensors from these coordinates. First the affine transformation tensor can be calculated as

\[
T_{affine} = \begin{bmatrix}
t_{11} & t_{12} & t_{13} & t_{14} \\
t_{21} & t_{22} & t_{23} & t_{24} \\
t_{31} & t_{32} & t_{33} & t_{34} \\
0 & 0 & 0 & 1
\end{bmatrix},
\]

(3.2.1)
Figure 3.8: The jaw movements during the word ‘hoard’. The position of the jaw is shown at (a) 0.2s, (b) 0.5s and (c) 0.8s.

which can be calculated from the four pairs of deformed \((x, y, z)\) and undeformed coordinates \((X, Y, Z)\), via relations

\[
\begin{bmatrix}
X_1 & Y_1 & Z_1 & 1 \\
X_2 & Y_2 & Z_2 & 1 \\
X_3 & Y_3 & Z_3 & 1 \\
X_4 & Y_4 & Z_4 & 1
\end{bmatrix}
\begin{bmatrix}
t_{i1} \\
t_{i2} \\
t_{i3} \\
t_{i4}
\end{bmatrix}
= \begin{bmatrix}
x_{1,i} \\
x_{2,i} \\
x_{3,i} \\
x_{4,i}
\end{bmatrix} \quad i = 1, 3. \tag{3.2.2}
\]

The resulting transformation includes the rotation, translation, shearing and scaling components. A negligible amount of deformation of the jaw is assumed during any speech utterance, so any such non-zero components in the transformation matrix are most likely due to the noise in the measurements. To eliminate the shearing and scaling components, the orthogonality property of the rotation matrix is used and the inner \(3 \times 3\) matrix which contains the rotation, shearing and scaling components, is singled out. The resulting matrix can be written as

\[
T_{3\times3} = R_{3\times3} U_{3\times3} \quad \text{and} \quad T_{3\times3}^t T_{3\times3} = U_{3\times3}^t R_{3\times3} U_{3\times3} R_{3\times3} U_{3\times3} = U_{3\times3}^2, \tag{3.2.3}
\]

where \(R_{3\times3}\) is the rotation tensor and an orthogonal matrix, and \(U_{3\times3}\) is a upper triangular matrix containing the shearing and scaling components. Then the scaling and shearing parts of the transformation can be removed via the formula

\[
R_{3\times3} = T_{3\times3} (T_{3\times3}^t T_{3\times3})^{-\frac{1}{2}}. \tag{3.2.4}
\]

Finally, the inner \(3 \times 3\) matrix of \(T_{affine}\) is substituted with the rotation tensor \(R_{3\times3}\) to get the Euclidean transformation for the jaw. The resulting jaw positions during the recording of the ‘hoard’, were sampled at three instances and are presented in Figure 3.8 where the maxilla model represents the static head.
3.3 Modelling Deformable Bodies-Host Mesh Fitting

The soft tissue models, including the tongue, cheeks and lips, are treated as deformable bodies simulated using the host mesh fitting technique. The host mesh fitting technique is designed for under-constrained problems where the information available is insufficient to provide a unique solution. Detailed mathematical derivations and more examples of the host mesh fitting can be found in [89]. A brief overview of the technique and its applications on the tongue and cheek-lips models in the ‘Talking Head’ are provided in the following sections.

3.3.1 Host and Slave Meshes

In host mesh fitting, the geometric model of interest, also called the slave mesh (e.g. the tongue and cheek-lips mesh) is embedded inside a host mesh which is often chosen to be a simpler geometry with fewer DOFs (e.g. 864 DOFs in the tongue mesh vs 288 DOFs in its host). As the slave mesh is fully embedded in the host (i.e. $\Omega_{\text{master}} \cap \Omega_{\text{slave}} = \Omega_{\text{slave}}$), the slave domain can be expressed in terms of the parameters of the host. For any point on the slave mesh, assuming both domains are interpolated using cardinal polynomial type basis defined in Equation (2.2.4), its value can be calculated by,

$$x_{\text{slave}} = \alpha_{ij} \xi_j x_{i,\text{master}} + \beta_{mn} \eta_n x_{m,\text{slave}},$$  \hspace{1cm} (3.3.1)

where $\xi$ and $\eta$ are the basis vectors for the master and slave domains respectively. Given the values of $x_{\text{slave}}$ and its mapping in $\Omega_{\text{slave}}$, a unique basis vector $\xi$ can be found for it in $\Omega_{\text{master}}$. The rationale behind the operation is that since the slave mesh is now a part of the internal structure of the master, as long as a good quality host mesh is managed, so shall its embedded structure.

The topology of the host is chosen to fit the shape of the slave mesh and should have sufficient DOFs to match the prescribed movements of material points. In Figure 3.9(a), a subdivided tricubic element is used as the host mesh for the tongue model while a more complex one is made for the cheek-lips model, where the host mesh is refined in a way so that the lips are isolated in a few central elements, as shown in Figure 3.10(a).

3.3.2 Matching Reference to Target Points

Every spoken word sampled in EMA, was initiated from the reference state, that is with the mouth closed and the tongue resting on the teeth. During the host mesh fitting, the measured reference material points are also embedded in the host. At their reference positions, they are referred as ‘landmark points’ and their local coordinates in the host domain ($\Omega_{\text{host}}$) may be calculated in the same way via Equation (3.3.1). After the movement, their new locations are called ‘target points’ and can be either inside or outside of the host mesh.
Figure 3.9: The host mesh fitting on the tongue. (a) The reference state where the slave mesh (the tongue in red) is embedded in a subdivided cubic Hermite element (shown by white lines) and the ‘landmark points’ are represented by the green spheres. (b) The deformed state of the host and slave meshes and the updated ‘target points’ for the vowel /ɔ:/.

In Figure 3.9(a), there are six landmark points on the top surface of the tongue, and their trajectories during the speech production are measured by the EMA. Four of them (sensors 7,8,9,12) lie approximately 10-20mm apart along the tongue groove and two more (sensors 10,11) are placed halfway along each side of the tongue blades. In addition, another eight virtual landmark points are created around the waist of the tongue and coupled their movement to the jaw kinematics. The purpose of this is to bring a level of coupling between the tongue and jaw as mechanically the tongue is sitting inside the jaw and connected to the jaw via various muscle groups and connective tissues.

The distance between a landmark and its targeting position is expressed in terms of the Euclidean norm $\|e_d\|^2$. There are more DOFs on the host mesh than the ones provided by the data points, therefore Sobolev smoothing functions are built into the object functions in order to apply further constraints on the deformation of the host mesh. Overall, the operation can be formulated as an optimisation problem where the objective function can be formed by differentiating the distance function by the nodal parameters of the host and equating to zero

$$\frac{\partial E}{\partial \mathbf{x}_f} = 0,$$

where $E = \sum_{d=1}^{D} e_d^2 + F_s$ and $D$ is the total number of landmark or target points. The 3D
Figure 3.10: The host mesh fitting on the cheek-lips mesh. (a) The host mesh (marked in lines), the slave mesh (the cheeks and lips) and the markers (marked in spheres around the lips) in the reference state. (b) The cheek-lips mesh after the surface fitting to the jaw movements in /ɔ:/ (c) The cheek-lips mesh after the host mesh fitting for the /ɔ:/ where the updated marker locations are marked as spheres.

Sobolev smoothing function

\[
F_s = \gamma_1 \| \partial x / \partial \xi_1 \|^2 + \gamma_2 \| \partial x / \partial \xi_2 \|^2 + \gamma_3 \| \partial x / \partial \xi_3 \|^2 + \gamma_4 \| \partial^2 x / \partial \xi_1^2 \|^2 + \gamma_5 \| \partial^2 x / \partial \xi_2^2 \|^2 + \gamma_6 \| \partial^2 x / \partial \xi_3^2 \|^2 + \gamma_7 \| \partial^2 x / \partial \xi_1 \partial \xi_2 \|^2 + \gamma_8 \| \partial^2 x / \partial \xi_1 \partial \xi_3 \|^2 + \gamma_9 \| \partial^2 x / \partial \xi_2 \partial \xi_3 \|^2 + \gamma_{10} \| \partial^2 x / \partial \xi_1 \partial \xi_2 \partial \xi_3 \|^2,
\]

(3.3.3)

is added to Equation (3.3.2), where coefficients \( \gamma \) are called Sobolev smoothing factors that constrain the geometry of host mesh by penalising the arc-length \( \gamma_i(i = 1..3) \), the curvatures \( \gamma_i(i = 4..6) \), the surface areas \( \gamma_i(i = 7..9) \) and the volume \( \gamma_i(i = 10) \) [89].

Constraints are applied on the element volume, enforcing incompressibility, and on the area, arc-curvature and arc-length of each facet, implying a stiffness constraint. In a way, the method might be viewed as an analogy to the muscle-activation model such as the ones presented in [30, 133], where the deformation is determined by both the Sobolev constraints (e.g. the passive tissue property) and the Euclidean norms due to the movements of material points (e.g. active muscle force) in an optimisation process. Similar to what has been done in the previous examples of the least-squares fitting, different sets of Sobolev coefficients were trialled on the
tongue and cheek-lips models before the final version was taken based on both the accuracy (e.g. RMS errors on the material points) and the consideration of the physiological property (e.g. volume conservation). Large penalty weights have been chosen, as illustrated in Table 3.1 and similar values have been used for the customisation of the skeletal muscle and lung in [89]. The deformed states of the tongue for the word ‘hoard’ are illustrated in Figure 3.9(b).

<table>
<thead>
<tr>
<th>Model</th>
<th>γ_1</th>
<th>γ_2</th>
<th>γ_3</th>
<th>γ_4</th>
<th>γ_5</th>
<th>γ_6</th>
<th>γ_7</th>
<th>γ_8</th>
<th>γ_9</th>
<th>γ_{10}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongue</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.01</td>
</tr>
<tr>
<td>Cheek-Lips</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.01</td>
</tr>
</tbody>
</table>

For the cheek-lips model shown in Figure 3.10(a), there is the positional information of the eight material points around the lips provided by the video frames. The simulation for the cheeks and lips have also been carried out by the host mesh fitting method in conjunction with the least-squares face fitting technique. At first, the lower half of the cheek-lips mesh is coupled to the jaw movements by refitting the affected elements (Figure 3.10(b) for the /ɑ:/ sound). Next the segmented video data is applied, in the form of eight material points around the lips, into the host mesh fitting (Figure 3.10(c) for the /ɑ:/). For the host mesh fitting for the cheek-lips model, all the coordinates (e.g. x, y and z) values are fixed on all the nodes of the host mesh and only the derivatives (i.e. curvatures) are optimised.

### 3.3.3 Updating the Slave Mesh

After the deformation, new nodal values on the host mesh ($x_{i,\text{host}}$) are obtained. As the mapping between the slave and host domains remains the same as established in Equation (3.3.1), the nodal values on the slave mesh may be updated with the pre-calculated values of $\xi_j$. For a slave mesh element with $n$ number of DOFs, their new nodal values can be reached by solving a matrix system,

$$
\begin{bmatrix}
\psi_1(\eta^1) & \psi_2(\eta^1) & \ldots & \psi_n(\eta^1) \\
\psi_1(\eta^2) & \psi_2(\eta^2) & \ldots & \psi_n(\eta^2) \\
\vdots & \vdots & \vdots & \vdots \\
\psi_1(\eta^n) & \psi_2(\eta^n) & \ldots & \psi_n(\eta^n)
\end{bmatrix}
\begin{bmatrix}
x_{1,\text{slave}} \\
x_{2,\text{slave}} \\
\vdots \\
x_{n,\text{slave}}
\end{bmatrix}
= 
\begin{bmatrix}
x(\xi^1) \\
x(\xi^2) \\
\vdots \\
x(\xi^n)
\end{bmatrix},
$$

(3.3.4)

where $\xi^n$ and $\eta^n$ denote the pre-calculated local coordinates of the $n^{th}$ value of $x$ on the host and slave domains respectively. The deformed states of the tongue and cheek-lips model for the vowel /ɑ:/ are shown in Figures 3.9(b) and 3.10(b), respectively, and more examples for the speech sounds can be found in Figures 3.14, 3.15 and 3.16.
3.4 Contact between Two Rigid Bodies

The epiglottis can be described as a piece of cartilage sitting obliquely upward behind the root of the tongue and the hyoid bone [4]. From the experiment there are no direct EMA or visual measurements from the region of the pharynx, but it is known that the epiglottis can tilt down to attain a transverse position, covering the larynx passage during deglutition (swallowing), also known as the first epiglottis movement [134]. According to [134], the first movement of the epiglottis is mainly passive and is induced by the muscle groups around the tongue root. It is therefore assumed that the potential motion of the epiglottis is only a passive response to the contact by the tongue during the speech. Furthermore, the epiglottis has been treated as a rigid body and set to rotate about the point where it is attached to the tongue root.

An algorithm was developed based on the above assumptions for modeling the epiglottis movements. It can be summarised in two stages: collision detection and collision correction. The former is achieved by projecting 11 material points defined on the back surface of the epiglottis onto its front surface and to the surface of tongue. A collision occurs when any of the projections to the tongue is less than its distance to the front surface of the epiglottis. Once a collision is detected, the epiglottis is rotated about its root until the collision detection routine returns a false. A demonstration of the algorithm is presented in Figures 3.11(a)-(c) where the tongue is translated along the x-axis towards the backwall.

![Figure 3.11: Contact simulation between the tongue and epiglottis. (a) Before contact. (b) Contact phase 1. (c) Contact phase 2.](image)

3.5 Results

Three vowels and two consonants were simulated on the ‘Talking Head’ model, using the methods outlined above. These articulations are compared with their sustained versions in the MRI scans and video images of the same subject in the next few sections.
3.5.1 Segmentation of Phonemes

The kinematic data was segmented from five short words, where three vowels are separated from CVC (consonant-vowel-consonant) structured words of ‘heed’, ‘hard’ and ‘hoard’, respectively, whereas two fricatives are taken as the preceding consonants in ‘seem’ and ‘sheet’, respectively. The phonemes are identified by their spectra (Figure 3.12) and the sampling points were decided with consideration of the articulatory measurements.

![Segmentation of phonemes from five short words](image)

Figure 3.12: The segmentation of phonemes from five short words, where the selected EMA frames of each target vowel and fricative are marked by arrows.

3.5.2 Jaw Motion

The rotations and translations for the five selected vowels and fricatives, calculated using the method introduced in Section 1.2, are summarised in Table 3.2.

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Tx (mm)</th>
<th>Ty (mm)</th>
<th>Tz (mm)</th>
<th>Rx (rad)</th>
<th>Ry (rad)</th>
<th>Rz (rad)</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i:/</td>
<td>5.57</td>
<td>-1.27</td>
<td>-3.99</td>
<td>0.027</td>
<td>0.034</td>
<td>-0.0024</td>
</tr>
<tr>
<td>/a:/</td>
<td>19.47</td>
<td>1.87</td>
<td>-18.16</td>
<td>0.0040</td>
<td>0.12</td>
<td>-0.0095</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>17.34</td>
<td>4.86</td>
<td>-15.27</td>
<td>0.021</td>
<td>0.11</td>
<td>-0.010</td>
</tr>
<tr>
<td>/s/</td>
<td>4.92</td>
<td>-2.00</td>
<td>-5.94</td>
<td>0.0069</td>
<td>0.035</td>
<td>0.0052</td>
</tr>
<tr>
<td>/ʃ/</td>
<td>6.36</td>
<td>-0.45</td>
<td>-2.82</td>
<td>0.00095</td>
<td>0.029</td>
<td>-0.00012</td>
</tr>
</tbody>
</table>

Table 3.2: Jaw Movements for 5 Phonemes
Amongst all five simulated phonemes, the three primary components in the jaw movements are the horizontal and vertical translations (Tx and Tz) and yaw (Ry), similar to the finding by Edwards in [135]. The mid-sagittal contours of the jaw are also compared to the corresponding MRI scans during sustained articulation of the same phonemes in Figure 3.14, where the contours of the jaw model are outlined in white.

### 3.5.3 Deformation of the Tongue

The positions of the EMA sensors attached to the tongue surface for five selected phonemes, with respect to their reference states, are presented in Figures 3.13. Figure 3.13(a) shows characteristics of the frontal vowel /iː/, the central vowel /aː/ and back vowel /ɔː/ in the x-z plane (i.e. in the mid-sagittal view), reconstructed by connecting the 4 sensor positions along the tongue groove. Compared to the lateral (x axis) and vertical (z axis) translations, there is very little transversely movement of the sensors, as shown in Figure 3.13(b). The data on the right side of the tongue blade for the fricative /ʃ/ seems to be in error, possibly due to a loose sensor. It was therefore removed from the simulation.

![Figure 3.13](image_url)

Figure 3.13: Positions of the tongue sensors for the three vowels and two fricatives. (a) Sensors 12, 9, 7 and 8 on the X-Z plane. (b) Sensors 10, 7 and 11 on the X-Y plane.

The five deformed tongue shapes are shown in Figure 3.16 and their mid-sagittal contours
are extracted and outlined in red lines in Figure 3.14 against the MRI scans of sustained articulations. The maximum, minimum, mean and standard deviation between the simulated tongue shapes and their digitised forms extracted from MRI scans are listed in Table 3.3 and illustrated in Appendix B.1. The bottom and side surfaces of the tongue have not been included in the comparison due to the insufficient image resolution (7mm depth in the axial planes) to allow clear separation of the tongue from its surrounding tissues. For the back, front and top surfaces of the tongue, the digitisation was conducted on six sagittal planes at 7mm intervals.

![Figure 3.14: Mid-sagittal contours of the simulated articulations versus MRI scans of five phonemes. (a) /s/; (b) /ʃ/; (c) /iː/; (d) /aː/; (e) /oː/.](image)

54
Table 3.3: Comparison between Simulated Tongue Shapes & MRI Scans

<table>
<thead>
<tr>
<th>Model</th>
<th>Frontal Face (mm)</th>
<th>Top Face (mm)</th>
<th>Back Face (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Min</td>
<td>Mean</td>
</tr>
<tr>
<td>/s/</td>
<td>8.2</td>
<td>0.0</td>
<td>3.1</td>
</tr>
<tr>
<td>/f/</td>
<td>6.5</td>
<td>0.1</td>
<td>2.7</td>
</tr>
<tr>
<td>/i:/</td>
<td>8.8</td>
<td>1.5</td>
<td>4.8</td>
</tr>
<tr>
<td>/a:/</td>
<td>7.7</td>
<td>0.0</td>
<td>3.4</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>7.6</td>
<td>0.1</td>
<td>2.2</td>
</tr>
</tbody>
</table>

3.5.4 Deformation of the Cheek-Lips

In Figures 3.14(a)-(e), the customised cheek-lips models are compared with the corresponding video frames for each segmented speech sound and with the shapes digitised from corresponding MRI scans in Appendix B.2. In most views, the shapes of the simulated lips have matched their respective visual contours well, however the RMS fitting error is generally larger than for the tongue model. The shape of the lips are also presented in Figure 3.15 on their respective MRI scans during sustained phonation and the difference between the simulated and imaged labial shapes are quantified in Table 3.4, via the projection methods.

Figure 3.15: The deformation of the cheeks and lips by host mesh fitting. (Images are produced with permission from the participant.) (a) /s/ RMS = 1.94mm; (b) /f/ RMS = 1.59mm; (c) /i:/ RMS = 1.87mm; (d) /a:/ RMS = 3.00mm; (e) /ɔ:/ RMS = 1.18mm.
Table 3.4: Comparison Between Simulated Lips Shapes & MRI Scans

<table>
<thead>
<tr>
<th>Model</th>
<th>Upper Lip(mm)</th>
<th>Lower Lip(mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Min</td>
</tr>
<tr>
<td>/s/</td>
<td>4.0</td>
<td>0.0</td>
</tr>
<tr>
<td>/f/</td>
<td>5.4</td>
<td>0.1</td>
</tr>
<tr>
<td>/i:/</td>
<td>7.1</td>
<td>0.0</td>
</tr>
<tr>
<td>/a:/</td>
<td>6.1</td>
<td>0.1</td>
</tr>
<tr>
<td>/ə:/</td>
<td>10.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

3.5.5 Contact between the Tongue and Epiglottis

As shown in Figure 3.16, the relationship between the simulated tongue and epiglottis is illustrated for the three vowels and two fricatives. There is no contact between the tongue and the epiglottis in the vowel /i:/, which adds a branch of the airway into the VT domain. This finding is supported by the MRI scans of the same sustained sound, as shown in Figure 3.14(c). The branch disappears in the two other vowels /a:/ and /ə:/ and there is some degree of contact between the tongue and epiglottis with the epiglottis being displaced more in the latter case, which is also supported by the MRI scans in Figures 3.14(d) and (e) respectively. The match between the simulated and measured epiglottis shapes lessens in the two fricative cases. In Figure 3.14, it can be seen that in the MRI scans, the location of the epiglottis were elevated coupled with a forward tongue articulation in /s/, /f/ and /i:/, while it was in a lowered position in /a:/ and /ə:/ where the tongue articulated towards the back of the pharynx cavity.

Figure 3.16: Contact simulations between the tongue and epiglottis in vowels and fricatives.
(a) /s/; (b) /f/; (c) /i:/; (d) /a:/; (e) /ə:/.
3.6 Discussion

From the comparisons between the simulated deformed state of the tongue and the MRI scans, it seems that errors (both the maximum and mean values) are generally the largest on the back side of the tongue for four of the test sounds with the exception of the back vowel /ɔ:/.

This is most likely due to lack of direct measurements coming from that region. Given the incompressibility constraints introduced in the host mesh fitting algorithm, the errors on the back side will also negatively impact on the definition of tongue shapes elsewhere. As each articulator was simulated separately and contact constraints were loosely defined by additional material points in the host mesh fitting, it can be observed in the mid-sagittal contours (Figure 3.14) that there are some interceptions between the surfaces of different articulator models, such as in the case of the bottom of the tongue and the jaw model, but it does not seem to have adversely affected the vocal tract shapes.

On the other hand, the modelled lip shape appears to match better on the exterior surfaces with the images than the tongue model does. The possible explanation for this could be the better spread of the material points while the only direct measurements on the tongue come from the top face.

These errors may also be caused by the different recording conditions between the MRI and EMA data. Engwall in [136] provided a comparison study for these ‘unnatural’ factors. His conclusions suggested:

1. The artificial sustained sound could cause the articulations to be hyper-articulated compared to normally sustained articulations.

2. A supine position could draw the tongue towards the pharynx by gravity and cause a narrower pharynx passage.

Similar findings on the gravitational effects on the tongue are confirmed in [30]. As the test subject was lying supine in the MRI while sitting upright during the EMA recordings, it is possible that his tongue was retracted more in the MRI scans than in the EMA study. Also notice that each sagittal image is in fact averaged over a depth of 7mm, which could contribute to the discrepancies between the model and the MRI.

In addition, coarticulation effects [21] may also play a role in bringing the tongue into a more forward position since in the two fricatives segmented from the words ‘seem’ and ‘sheet’, respectively, and both followed by a front vowel /i:/.

The ‘Talking Head’ model directly applies the experiment measurements in simulation of articulatory motion without the need for arbitrarily defined control parameters, compared to other similar data-driven articulatory models [26, 25, 27]. This feature gives the ‘Talking Head’ model an advantage in terms of flexibility and compatibility with different types of experimental measurements. Without the need for a finite number of statistically-derived controlled parameters, the articulatory models in the ‘Talking Head’ have more DOFs than the other
models, which can be important in simulating deformations in different speech environments (e.g. artificially sustained vs normal speech). At the same time, the host mesh deformation technique converts the input data into material points of the interested geometry, meaning that different types of measurements can be collected anywhere inside or on the surface of the body. The use of Sobolev smoothing functions as part of the optimisation problem allows the introduction of physiological constraints on the deformation of the articulators (e.g. volume conservation), thus bringing the mathematical model closer to physical realism.

The main weakness of the model is the limited accuracy. When compared to similar models, the ‘Talking Head’ has significantly higher RMS errors for the tongue (1.2-1.6mm in [25], 1.2-1.3mm in [26], 1.4-6.4mm in the current model) and the lips (1mm in [25], 1.3-3mm in the current model) in simulated speech articulation. This is partially due to the lack of measurements in the pharynx and larynx regions and partially caused by the fact that the kinematics of the ‘Talking Head’ are not statistically derived from artificially sustained articulation but driven by the measurements taken in a different speech environment, as discussed above.

It is difficult to compare the ‘Talking Head’ with muscle-activated biomechanical-type models as there is so far no published results of simulated speech articulation by a 3D biomechanical model to the author’s knowledge. On the other hand, less computational resource is generally required by data-driven models than mechanical types driven by the muscle activation ([41, 29, 107, 14]). For instance, 40min is required for 100ms simulation by the muscle-activation model of the tongue in [30] on a single processor while less than 4s is needed for the same period of simulation using the ‘Talking Head’ model.

3.7 Summary

Experiments are performed to collect the kinematic measurements of the vocal articulators by the EMA and video cameras during a range of speech tasks (words and sentences). Numerical methods are developed to translate the measurements into simulated articulatory movements with added physiological constraints, on the ‘Talking Head’ model introduced in the last chapter. By directly representing the underlying imaging and motion data, the model combines the high spatial resolution MRI measurements with the high temporal resolution measurements of EMA and video cameras. Five simulated articulator shapes of speech utterance (3 vowels and 2 consonants) are compared with their MRI images during sustained articulation.
Chapter 4

Geometric Modelling of the Vocal Tract

Given each set of simulated articulation, the segmentation was performed to isolate the airway from the surrounding surfaces and both surface and volume reconstructions were conducted for the segmented VT domain. A brief review of the modelling methods for the VT is provided in Section 4.1. For the segmentation method described in Section 4.2, a flood-fill algorithm has been developed to extract the airspace in a 3D space partitioned by an overlapping tree scheme. Parametric surfaces can be extracted from the resulting cell structures via a fast algorithm and a high order volume mesh can be built with a semi-automatic algorithm, introduced in Section 4.3. Six VT configurations, produced by the newly developed methods, have been compared to their sustained versions in MRI both on images and through their area functions in Section 4.4.

4.1 Introduction

To segment the VT domain, there exist two schools of thoughts in the literature. The first approach is centered upon the idea that a starting configuration may be defined for the VT, which can be customised in coordination with the movements of the surrounding articulators. Engwall [26] presented a parametised model for simulating dynamics of a 3D VT model which is controlled by 10 parameters to represent effects of the motions of the larynx, jaw, lips, velum and tongue on the final VT shape, where each parameter maps certain vertices of the VT mesh to the corresponding articulator surfaces and deforms them symmetrically to the mid-sagittal plane. Fels and his co-workers also adopt a similar strategy for extracting 3D VT meshes from their mechanical models of various articulators (jaw, tongue, face and lips) by setting the airway vertices to corresponding surface vertices or a weighted sum of such vertices near gaps or junctions [107].

The second approach aims to directly segment the airway without any assumptions about its configuration. In [78], the author manually segmented five vowel sounds from the corresponding
MRI scans during sustained speech articulations, where the tract was digitised on calculated normal planes along the estimated center lines. A more automated segmentation method was trialled by Story and his team who applied a line of image processing techniques such as thresholding and region growing amid some manual modifications to the original images [65]. Those processed frames were then used to reconstruct the VT volume and subsequently reprocessed for evaluation of area functions for vowels and consonants. Takemoto extended this approach to study the temporal behavior of area functions for vowels sounds, segmented from a group of scans acquired by the 3D-cine MRI [137].

For the ‘Talking Head’ model, a second methodology is selected for segmenting the VT domain, mainly for its flexibility in handling complex geometries. In a speech environment, surfaces coming from different organs are constantly moving against each other, resulting in either addition or deletion of articulator surfaces from the VT domain (e.g. the sub-cavity between the epiglottis and the back of the tongue in /a:/ in Figure1.5(c) and /i:/ in Figure 1.5(d)). In some extreme cases such as the situations in stop sounds, as shown in Figures 1.6(d) and (e), the tract can be broken into two pieces, making defining a starting configuration a very difficult task. It is also desirable to minimise the amount of human intervention in the segmentation process as the method is aiming at reconstructing the VT for a large group of configurations collected from the EMA-video experiments introduced in Chapter 3. Therefore, a 3D seeded-region growing method similar to the ones presented in [138] was implemented with the new flood-fill algorithm. However, there are some major issues related to this type of method, and care needs to be taken to isolate the VT domain from the external (i.e. open space) and internal (e.g. lung and trachea) airspace and prevent leakage through gaps between different articulators [67]. Techniques are developed to address these two issues.

After segmentation, a parametised mesh needs to be created for the VT domain so that it can be used in visualisation, morphological analysis and numerical simulations (e.g. CFD, CAA). There exists two classes of methods for reconstructing VT surfaces/volumes among past studies. The first group is via contouring methods of which the Marching Cubes (MC) algorithm is the most common choice in VT modelling such as the examples in [139] for the two Czech vowels /a/ and /i/; in [140] for a liquid sound /r/ and in [51] for the fricative /s/. The MC algorithm operates on a grid structure (e.g. voxels). The algorithm visits each grid point in a group of eight, arranged in a cubical form, and decides the shape of linear facets in the neighbourhood according to a scalar value (e.g. ‘1’ for a surface and ‘0’ elsewhere), using a look-up table made of 256 cases [141]. The end result is a triangulated mesh suitable for arbitrary shapes. The MC algorithm is particularly popular in image processing as the voxel structure in the images provides a natural contour for the algorithm to operate upon, however the MC method can generate undesirable features on the reconstructed surface due to defects in the input data (e.g. sampling noise, etc), which is one of the major drawbacks when applied to experimentally collected data [142] such as the case reported in [140] for a VT model.

Deformable surface methods such as the bicubic Hermite meshes introduced in [78] for 5
vowel sounds, are designed to reach a compromise between the approximation accuracy and the quality of the mesh. Such a method explores the concepts of free-form surfaces by building an initial seed mesh, then deforms it according to both the ‘internal force’ and ‘external force’. Here the ‘internal force’ refers to the constraints which the user specifies on the finished surface (e.g. no tears, no holes or no serious distortions etc.), while the ‘external forces’ are related to the original geometry. The implementation of this type of method often involves an energy minimization procedure where the two forces are combined in a single system. This method has been shown to be robust to noise and produces desired features such as continuity of the finished surface. For this reason, this type of method has been particularly welcome among mathematical modellers who are always dealing with noisy inputs [89]. However, one serious downside of the method is the requirement of a pre-defined seed mesh as the starting point for the optimization process, which could be a non-trivial task for complex and dynamic geometry of the vocal tract. In many cases [91, 78, 92], a stage of time-consuming manual creation of an initial mesh was performed by experienced users, otherwise pre-knowledge of the geometry has to be incorporated into the algorithm itself in order to automate the process [143].

After noticing the strengths and weaknesses of the two aforementioned reconstruction methods, some researchers have attempted to combine the two methods into one robust algorithm for modelling arbitrary shapes. Sinha proposed a two-stage process, whereby a local optimization technique LMS (least median of squares) was adapted to obtain the initial mesh from a natural grid and the linear mesh was converted to a smoother version with cubic spline functions during the second stage of the process [144].

In [145], the author approximated the surface first by a group of tangent planes based on local optimisation of $k$ nearest neighbours of point $x_i$ for the entire data set. The resulting tangent planes were converted into simplicial (i.e. triangle) meshes via the MC algorithm. In the author’s later work [146], further improvements were made by fitting the mesh and converting it into a higher order type. A similar line of work has been investigated by Stoddart who converted the triangular elemental mesh extracted by MC into the quadrilateral type, then fitted them to the original data cloud in order to produce $G-1$ continuous smooth surface [147]. In [148], it was realised that contour-based mesh extraction methods may not only apply to those naturally gridded data like the voxel structures in the images but can also apply to more general vector data sets where the 3D space can be subdivided into cubical elements each assigned a binary value (e.g. ‘1’ for containing a data point, ‘0’ for empty space).

For the reconstruction stage in the VT model, two methods have been developed for creating a single continuous mathematical domain segmented by the flood-fill algorithm. The first method takes advantage of the contouring structure used by the segmentation and uses it for extracting a parametrised surface made of linear basis functions. The extracted surface is smoothed and fitted to the original data in order to achieve a better quality and accuracy. The process is automated and only requires a minimal amount of human input. In the second method, a structured tricubic mesh is created via the least-squares fitting method which has
been used for constructing all the articulator models. The procedure for creating the initial mesh is largely automated by applying the pre-knowledge of the unstructured surface mesh.

4.2 Vocal Tract Segmentation

From each simulated frame of the ‘Talking Head’ model, a single continuous VT mesh can be extracted by filling the airway with regular volume cells. All related articulator surfaces are decimated into data clouds for the selection by a flood-fill algorithm, combined with an overlapping space partitioning scheme.

4.2.1 Digitising the Articulator Surfaces

The following surfaces extracted from the ‘Talking Head’ model, were considered to be relevant to the VT boundary and are decimated into data clouds after the deformation.

1. The inner surfaces of the cheeks and lips (Figure 4.1(a)).

2. The surface of the teeth and gums on both the mandible and maxilla (Figures 4.1(c) and (d)).

3. The bottom face of the hard palate (Figures 4.1(c) and (d)).

4. The entire tongue exterior surface except for the bottom part (Figure 4.1(e)).

5. The pharyngeal wall and bottom face of the soft palate (Figure 4.1(f)).

6. The inner epiglottis and larynx surfaces (Figure 4.1(g)).

In addition to these surfaces, the two ends of the tract (lips in Figure 4.1(b) and glottis in Figure 4.1(i)) had to be sealed in order to isolate the region of interest from the upper trachea and the external airspace, respectively. Another two surface patches were added to provide boundaries for the bottom of the jaw as shown in Figure 4.1(d) and in front of the epiglottis, as shown in Figure 4.1(h). The density of the data cloud is approximately 125 points per $1cm^2$. Figure 4.2 give the resulting data clouds for the six VT configurations.
Figure 4.1: (a) The selected faces of the cheek-lips model. (b) The surface added to the opening of the lips. (c) The selected faces of the maxilla and jaw models. (d) The added jaw floor surface. (e) The selected surfaces of the tongue model. (f) The selected surfaces of the backwall and soft palate models. (g) The selected faces of the larynx and epiglottis models. (h) The added larynx surface in front of the epiglottis. (i) The added bottom surface to the glottis opening.
Figure 4.2: The digitised surfaces for six VT configurations. (a) The data cloud of the rest breathing state (282,345 points). (b) The data cloud of /s/ (287,099 points). (c) The data cloud of /ʃ/ (270,860 points). (d) The data cloud of /i:/ (295,369 points). (d) The data cloud of /a:/ (288,095 points). (e) The data cloud of /ɔ:/ (297,871 points).
4.2.2 Overlapped Tree Partitioning Strategy

In the field of medical imaging, the image space is usually divided into cubes called voxels, each of which occupies a unique subspace. In the new scheme, the centers of the cells are placed on the borders of their neighbouring cells, giving partially overlapped cubes, as illustrated in Figure 4.3 for a 2D example where the cell in the center shares its space with 9 other cells. In 3D, this means that each interior cell shares the space with 26 neighbours of the same size. When applying this partition scheme to a data set, those cells which contain one or more data points are flagged as boundary cells, by which a vector space is effectively reduced to a scalar field (i.e. ‘1’ for boundary cell and ‘0’ for interior ones).

Figure 4.3: A schematic of the overlapping cells used in a 2D implementation of the algorithm. Cells are numbered from 1 to 9 at their center locations. Cell 1 (solid lines) shares some of its space (shading diagonal) with cell 3 (dashed lines); Cells 6 (dotted lines) and 8 (dashed and dotted lines) share part of the area of cell 1 (shaded in vertical and horizontal lines respectively) and each other (shaded in both horizontal and vertical lines).

Compared to the conventional non-overlapping grid, the new scheme produces more compact boundary descriptions using the same resolution. The algorithm is demonstrated on a 2D triangle which is a quadratic simplex-type element, shown in Figure 4.4(a), with its edges decimated into discrete data points using an evenly spaced sampling size of 500 points per side. The space is partitioned by both the non-overlapping (Figure 4.4(c)) and the overlapping grid (Figure 4.4(b)) centered at point (2.6,1.7) with a uniform cell size of 0.0169 normalised by the length of the longest side of the triangle. On the triangle surface, the boundary ranks are intercepted by the original data points.

In Figures 4.4(b) and (c), it is observed that the boundary cells are more closely packed in the overlapping structure than those based on the non-overlapping scheme. Moreover there is a
more serious problem in the boundary representation by the non-overlapping scheme identified in Figure 4.4(d) where cell 516 is not part of the boundary rank by not containing any data points, although it should be if the input boundary description is continuous. In this case, if the compact growth strategy (i.e. 8 growing directions per interior node) is applied to this non-overlapping grid, leakage will occur from interior cell 547. This shows that the non-overlapping structure is prone to discrete representation of curved surfaces even though the cell size is much larger than the average size of the gaps in the discrete boundary representations. In contrast, the partition formed by the overlapping scheme, shown in Figure 4.4(b), successfully completes the segmentation of the boundary by including all the data points in its boundary ranks.
4.2.3 Flood-Fill Method

After partitioning the space, an advancing front is initiated at a specified interior cell and grows into its peripheral space in the same way as in a region growing method. The algorithm has been written as pseudo-code in Figure 4.5.

```plaintext
Pick a seeding point;
Partition the space;
queue(1) = seeding point;
queue counter = 1;
stack counter = 0;
DO WHILE queue counter != stack counter
    stack counter++;
    stack(stack counter) = queue(stack counter);
    DO 'Total number of neighbours
        IF The cell is not visited and not at the boundary
            queue counter++;
            queue(queue counter) = neighbour;
        ELSE IF The cell is not visited and at the boundary
            Flag it as a boundary cell;
        END IF
    END DO
END DO
```

Figure 4.5: The pseudo-code for the flood-fill algorithm, where the variable ‘queue’ and ‘queue counter’ represent the advancing front, the ‘stack’ and ‘stack counter’ represent the segmented cells.

The algorithm starts with a user-defined seeding point inside the target domain and adds it into the advancing front (the queue). Before the initiation of the flood-fill, the entire domain is partitioned with uniformly sized overlapping cells (squares/cubes). Within the loop, neighbouring cells (e.g. 8 in $\mathbb{R}^2$ and 26 in $\mathbb{R}^3$) are visited. If the neighbouring cell is not already visited, a new cell is created and added to the queue. Skip the cells flagged as boundary cells and keep processing the queue until all the cells in the queue are added to the segmented domain (the stack).

Upon a successful segmentation, there is a space partition made up of overlapping cubes and a selected data cloud made of data points intercepted by the boundary cells. The curved surfaces presented in most of the articulators make its digitised form more prone to leakages from the non-overlapping scheme, as shown in Figure 4.6(a) for the resting breathing state. Once leakage occurs, the flood-fill segmentation can not be successfully completed. On the other hand, the overlapping scheme (Figure 4.6(b)) produces the correct boundary definitions with the same cell size.
4.2.4 Segmentation for Vocal Tract Configurations

Six VT configurations have been segmented from the ‘Talking Head’ model, introduced in Chapter 3, by the flood-fill algorithm with an overlapping grid. With the exception of the vowel /æ:/ where the seeding point is in the oral cavity, the flood-fill algorithm all starts at a pharynx location for all five other VT configurations. The cell size has been adjusted so that the algorithm can complete the segmentation for the whole tract and the growth axes are aligned with the x, y and z axis in the ‘Talking Head’ model. The segmentation results are summarised in Table 4.1 and the resulting cell structure and segmented digitised surface (data points intercepted by the boundary cells) are shown in Figures in Appendix C.

4.3 Vocal Tract Meshing

In the previous section, methods are introduced for the segmentation of the simulated VT shapes in the form of digitised surfaces intercepted by an overlapping grid. For such representations, there is a single resolution of cells on all boundary ranks, from which a parametric surface may be extracted in a consistent manner. From the segmented structure, surface nodes can
be initially created as the centers of those boundary cells, and further connected into a surface mesh made of quadrilateral elements. From that point, two different ways of improving and constructing VT meshes are tested. In Section 4.3.1, improvements are applied to the bilinear mesh via the Laplacian smoothing and local optimisation techniques. In the second approach, introduced in Section 4.3.2, pre-knowledge of the VT shape gained by the first method, is incorporated into the an automated design of an initial structured volumetric mesh which is later converted into the cubic Hermite type via the least-squares fitting method.

### 4.3.1 Surface Extraction

Given a single resolution among all the boundary cells, the surface may be extracted from the overlapping tree structure in a regular manner. Every interior cell should be able to subdivided into eight individual cubes each made of eight center nodes as shown in Figure 4.7(a). The centres of each boundary cells have to be on the border of at least one internal cell, so all center nodes are included as a part of the new subdivisions. In this way, the external facets can be identified by the fact they are singular and only owned by a single cube. The end surface, consisting of nodes only from those boundary cells, carries a ‘Lego’-like pattern and is made of quadrilateral elements (4 nodes per element) as shown in Figure 4.7(b) for the VT model for the rest breathing state.

#### Boundary Accommodation

By definition, each surface node represents a cell which contains \( n \) (\( n \geq 1 \)) number of original surface data points, therefore their nodal values (e.g. \( f(u) = x,y \) and \( z \)) may be updated by the rule

\[
 f(u) = \frac{1}{n} \sum_{n} f(n). \tag{4.3.1}
\]

This effectively moves the node to the centroid of the enclosed data points. This movement is bounded within the existing cube. In fact, the node can only be displaced in part of its sector not overlapped with any of its internal neighbour(s). The update does not alter the topology of the surface but rather brings it closer to the original version. The resulting surface, consisting of nodes only from those boundary cells, is made of quadrilateral elements (4 nodes per element).
Figure 4.7: (a) The internal arrangements of the partitioned space. In 3D, every internal cube, such as cell 1 (marked in black) shares space with 26 neighbouring cells whose centers are marked in green circles. The space is divided with 8 non-overlapping cubes each formed with the centers of the internal cell 1 and 7 others, such as the one marked in red. (b) The extracted bilinear surface of the vocal tract for the rest breathing state.

per element) as shown in Figures 4.8(a) for the rest breathing model.

**Surface Smoothing**

The surface reconstruction method described above produces a continuous linear surface as long as the segmentation process completes successfully. The VT surface created in this way (Figure 4.8(a)), looks irregular due to the noise, non-uniform sampling in the original data cloud (e.g. some boundary cells may contain more than one data points) or defects in the samplings (e.g. data points collected from internal surfaces of the object). Laplacian filtering is a common practice for smoothing a parametric surface by placing a node in the center of its neighbours [149]. Given the surface connectivity information created in the previous step, a ‘stiffness’ factor can be further introduced as

$$f(x) = \frac{1}{n + m} \left[ \sum_n f(x_n) + m.f(x) \right],$$

where $n$ is the number of connected surface nodes and $m$ is called the ‘stiffness factor’, increasing its value tends to keep the node where it is. In practice, the zero stiffness setting produces the most homogeneous elements all over the surface, such as the case shown in Figure 4.8(b), but it could lead to a over-smoothed version where sharp features in the original geometry have been lost. For the VT configuration shown in in Figure 4.8(c), a stiffness factor of 10 is selected and
Least-Squares Fitting of the Surface-A Subdivided Strategy

Bounded displacement of boundary nodes can help to build a surface close to the original data cloud, however such movement of nodes carries no direct bearing on the fitting quality of the surface. As observed in the example in Figure 4.8(a), the unevenly sampled data cloud caused some level of ‘roughness’ on the updated parametric surface. While Laplacian smoothing can produce a smoother surface, as a reductive process, it reduces the fitting accuracy in return. A better way of approximating a surface is by the mean least-squares fitting method which aims to minimise the projection distances between each original data point to the nearest surface patch, similar to the fitting of articulator models. However the parametric surface of the VT contains many more DOFs than any of the articulator models and special treatment is needed to address the increased computational cost.

Compared to contour-based surface reconstruction methods, the least-squares optimisation technique requires much more computational resource as it demands solving a $n \times n$ matrix system and the computational cost can be shown to be proportional to $n^3$ [91]. One way to reduce the cost is by dividing the entire mesh into patches, each made of certain number of elements and a subgroup of data points. As each boundary node on the mesh represents a boundary cell intercepted by a number of data points on the overlapping grid, a connection is established between the data points and their associated group of elements.

In Figure 4.9, the VT model for the resting breathing state has been divided into 8 patches.
of equal numbers of elements. In this way, the fitting problem can be drastically sped up (3000s per iteration without subdivision vs 25s per iteration with subdivision of 8 patches) and yet there is no major loss in fitting accuracy for the final mesh after smoothing (0.32mm RMS without subdivision vs 0.33mm RMS with subdivision of 8 patches). As each fitting of the subdivided surface patch is independant, the program can be ran on multiple computers simultaneously.

There is still one issue associated with the method, which is caused by the break-up of the closed surface. The ‘edge effect’ creates irregularities on the interface nodes shared among different patches, as shown in Figure 4.10(a). In the current implementation, procedures of the Laplacian smoothing, same as the operation demonstrated in Figure 4.8(c), are applied to the nodal values of those interface nodes and the end results are shown in Figure 4.10(b). In the future, improvements may be made by fixing all the interface nodes during the fitting stage and updating their values locally by smooth interpolation of their neighbours across different surface patches.

In Figure 4.11, the other five VT configurations have been illustrated and Table 4.2 summarises the mesh statistics for all the tested VT configurations.
Figure 4.10: Reconstructed VT surfaces after fitting. (a) Before smoothing. (b) After smoothing.

Table 4.2: Vocal Tract Meshes for Six Configurations

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>No. Nodes</th>
<th>No. Elements</th>
<th>RMS Fitting Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rest Breathing</td>
<td>16,131</td>
<td>16,130</td>
<td>0.33mm</td>
</tr>
<tr>
<td>/s/</td>
<td>26,243</td>
<td>26,246</td>
<td>0.26mm</td>
</tr>
<tr>
<td>/f/</td>
<td>30,846</td>
<td>30,846</td>
<td>0.24mm</td>
</tr>
<tr>
<td>/i:/</td>
<td>20,052</td>
<td>20,054</td>
<td>0.38mm</td>
</tr>
<tr>
<td>/a:/</td>
<td>23,380</td>
<td>23,394</td>
<td>0.34mm</td>
</tr>
<tr>
<td>/a:/</td>
<td>22,419</td>
<td>22,430</td>
<td>0.38mm</td>
</tr>
</tbody>
</table>

4.3.2 Volume Reconstruction

The method introduced above can generate an unstructured bilinear surface mesh for the segmented VT shape, however its usefulness is limited by its topology and inability to apply higher order basis function (e.g. cubic Hermite). In the second method, a way to generate a tricubic volume mesh is explored for the segmented domain, which is similar to the strategy presented in [78]. Major improvement has been made in the automation of the creation of an initial mesh based on previously created surface meshes.
Estimation of Centerline Position

The process of the estimation of a centerline may be explained in the form of a Fortran-style pseudo-code, as shown in Figure 4.12(a) and demonstrated for the rest breathing state in Figure 4.12(b). Since all the VT models share a fixed larynx structure at the current implementation, it was started with a pre-defined normal plane at the glottal end, parallel to the vertical axis. From that point, a 1cm step was taken along the normal direction and another normal plane which intercepts with the VT surface produced by the flood-fill algorithm was sampled. A new planar center point was evaluated as the centroid of the intercepted normal plane and the new plane normal was estimated by connecting the two neighbouring center nodes and a new normal plane could be sampled from the surface mesh. These steps were done in an iterative fashion until the movement of the center nodes changes by less than 1mm or the number of iterations exceeded 10. The movement of the centers may be constrained on the sagittal planes in order to prevent the potential zig-zag effect of the centerline in the circumferential direction. At the
end, the process returns a group of estimated center nodes and the same number of normal planes, as shown in Figure 4.12(b).

![Pseudo-code for estimating a centerline](image)

**Initial Linear Mesh**

From the estimated centerline and normal planes, a linear mesh made of hexagonal elements can be created by connecting the nodes placed on each of the normal planes. In Figure 4.13(a), six nodes and two facets were placed on each plane and connected in a way to give a mesh made of 28 volume elements aligned approximately to the centerline of the tract, as shown in Figure 4.13(b). The nodal placement is calculated by projecting nodes initially placed around a large circle onto the intercepted VT surface, as shown in Figure 4.13(a).

In the initial mesh created for the rest breathing state in Figure 4.13(b), it can be seen that there is an obvious weakness in the construction which is the lack of a representation for a side branch between the back of the tongue and the front of the epiglottis. Due to the arrangement of the normal planes shown in Figure 4.12(b), such a sub-cavity is merged into the main tract in the initial mesh. To correct such mistakes, manual adjustment was made by stretching the front row nodes on planes 6 and 7 to cover the sub-cavity, as shown in Figure 4.13(c) for the rest breathing state.

**Least-Squares Fitting of the Volume Meshes**

The final step involves transforming the linear mesh to fit to the segmented data cloud. Once again, the least-squares fitting algorithm and cubic Hermite basis functions are used for the
Figure 4.13: The linear meshes for the reference state. (a) The projection of the plane nodes, where the initial nodal positions are marked in black, the intercepted VT surface are represented by spheres and the updated nodes are shown in red. (b) The linear mesh created by connecting the nodes defined on the normal planes. (c) The linear mesh created by connecting the nodes defined on the modified normal planes.

task. The initial results for the rest breathing state is shown in Figure 4.14(a) which has many intercepted surfaces from bad projections due to the imperfections of the initial mesh. Manual modification is made to correct the intercepted surface especially in the epiglottis region where the sub-branch requires careful adjustment around the sharp curvatures at the top of the epiglottis, as compared in Figures 4.14(c) and (d). Furthermore, refinements have been made along the tract in order to provide more DOFs to the mesh. The resulting cubic Hermite mesh is shown in Figure 4.14(b). These procedures have been performed for the five speech configurations and their cubic Hermite meshes are shown in Figure 4.15 and summarised in Table 4.3.

4.4 Validation of the Results

In this section, VT models, constructed using the methods introduced in Sections 4.3.1 and 4.3.2, were compared with their artificially sustained versions by MRI for six configurations.
Figure 4.14: The fitted cubic Hermite VT mesh for the resting breathing state. (a) The fitted tricubic mesh with the initial mesh defined in Figure 4.13(a). (b) The fitted tricubic mesh with the linear mesh defined in Figure 4.13(b). (c) A close-up of the epiglottis region of the mesh given in Figure 4.14(a). (d) A close-up of the epiglottis region of the mesh given in Figure 4.14(b).

The comparisons were conducted on both the sampled normal planes along the calculated centerlines and via area functions.
Figure 4.15: The fitted cubic Hermite mesh for the VT in five speech configurations. (a) /s/; (b) /ʃ/; (c) /i:/; (d) /a:/; (e) /ɔ:/.

Table 4.3: Cubic Vocal Tract Meshes for Six Configurations

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>No. Nodes</th>
<th>No. Elements</th>
<th>RMS Fitting Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rest Breathing</td>
<td>261</td>
<td>28</td>
<td>1.47mm</td>
</tr>
<tr>
<td>/s/</td>
<td>126</td>
<td>40</td>
<td>0.836mm</td>
</tr>
<tr>
<td>/ʃ/</td>
<td>120</td>
<td>38</td>
<td>0.721mm</td>
</tr>
<tr>
<td>/i:/</td>
<td>108</td>
<td>34</td>
<td>1.415mm</td>
</tr>
<tr>
<td>/a:/</td>
<td>300</td>
<td>152</td>
<td>1.022mm</td>
</tr>
<tr>
<td>/ɔ:/</td>
<td>180</td>
<td>76</td>
<td>1.418mm</td>
</tr>
</tbody>
</table>
4.4.1 Validation of the Rest Breathing State

For measuring the accuracy of the model, 15 normal planes were sampled at an interval of approximately 1cm along an estimated centerline (Figure 4.16(a)) and manually segmented from the MRI scans during rest breathing (Figure 4.16(b)), using the method first developed in [78]. The centreline was estimated in the middle sagittal plane and created as a cubic spline made of 14 line elements and the normal planes are defined as $128 \times 128$mm square sections centered at each center nodes.

![Figure 4.16: Comparison between the MRI scans and the 3D model for the vocal tract in the reference state. (a) The estimated centerline and the placement of the normal planes. (b) The estimated centerline and the segmented cross sections (in green) on 15 normal planes for the model and MRI scans.](image)

In Figures 4.17(a)-(k), the measured cross-sections from MRI are compared with the ones extracted from the 3D models. Notice that the soft palate was in a lowered position in the MRI scans, but in the model it was fixed into a permanent raised position in order to decouple the tract from the nasal cavity, therefore MRI slices 8-11 are discarded in the comparison. As can be seen in Figure 4.17, the extracted cross-sections from the linear mesh matches the MRI scans well in most of the calculated planes with the exception of Slice 14. The cross-sections extracted from the cubic mesh also show a good match to the ones of the linear mesh, except for the Slice 5 (Figure 4.17(e)) where the physical tract is split into two branches which are merged in the cubic mesh.
Figure 4.17: Comparison between the MRI scans and the model for the VT cross-sections in the rest breathing state, where the normal planes are numbered starting from the glottis end, as shown in Figure 4.16(b) and contours extracted from the linear model are marked in red dots and the cubic model are shown as green lines. (a) Slice 1; (b) Slice 2; (c) Slice 3; (d) Slice 4; (e) Slice 5; (f) Slice 6; (g) Slice 7; (h) Slice 12; (i) Slice 13; (j) Slice 14; (k) Slice 15.
The area functions of the linear and cubic meshes, presented in Figure 4.18(a) where Slices 8-11 are discarded in the comparison since the VT is collapsed in the related region during the rest breathing, also show good agreement with the MRI measurements within 10% maximum difference. The two area functions calculated from the two mathematical models show close agreement to each other despite the cubic mesh having nearly 4 times greater RMS error than the linear mesh. A close inspection show the major contributor of the error in the cubic mesh comes from the region around the epiglottis (Figure 4.17(e)). This error does not significantly affect the calculation of the local area and shows a limitation of the area function in representing the VT geometry.

4.4.2 Comparison between MRI and Modelled VT

Five sets of area functions for the speech sounds are presented in Figures 4.18(b)-(f), where each VT model was divided into 21 slices approximately 0.8cm apart so that slices 1-5 cover part of the trachea and the entire larynx, slices 6-12 extracted from the pharynx and slices 13-21 are sampled in the oral cavity. Normal planes were calculated along the centerline and the cross-sectional areas were calculated for both the MRI and models using the method described in the previous section.

In the larynx cavity, the largest cross-sectional area amongst the measured VT shapes is located near Slice 5 (approximately 4cm away from the glottis) for the rest configuration and phonemes with a forward tongue position (/s/, /ʃ/ and /iː/) and near Slice 4 (approximately 3cm away from the glottis) in the phonemes /a:/ (with a middle and lower tongue position) and /æ:/ (with a back tongue position). This may be explained by the lowering of the larynx by the backward movement of the tongue and is consistent with the findings presented in [150]. On the other hand, there are significant reductions in the cross-sectional areas between Slices 2 and 3 for four of the five speech phonemes (/s/, /ʃ/, /iː/ and /aː/) compared to the resting configuration. The cause is unclear, but this could be potentially the result of horizontal movements of the hyoid bone which is sitting approximately anterior to the affected part of the tract, as a forward motion of the hyoid bone can cause the thyroid cartilage to rotate in a direction to stretch the vocal cords [151]. If this is true, it means there is a need for a more explicit geometric model of the larynx cavity where different mobile components need to be represented individually in order to simulate the kinematics in speech. As the larynx cavity is treated as a fixed structure in the current model of the ‘Talking Head’, the first 4 sections of the modelled VT show no variation in the cross-sectional areas in all simulated configurations.

In the pharynx cavity, the variation in that part of the VT are due to the movement of the tongue or the tongue-induced epiglottis movement, since both the backwall and the soft palate models are fixed in the ‘Talking Head’. As shown in the mid-sagittal contours in Figure 3.14, all five test phonemes are produced with a raised velum; this supports the use of a fixed soft palate model for the simulated sounds. In the pharynx section (Slices 6-12) of the area
functions for /s/, /ʃ/ and /iː/, shown in Figures 4.18(b), (c) and (d) respectively, it can be seen that the modelled functions match the envelope of the measured shapes, but there seems to be a phase shift for the modelled functions toward the mouth opening by 1-1.5cm. This may be the result of the more backward articulation in the MRI due to the gravitational effect on the tongue compared to the more natural speaking posture (upright) in the modelled data (EMA), as discussed in Section 3.6. However, this hypothesis needs to be validated in the future since some studies have showed such effect might be subject to population variation [152]. The modelled /aː/ and /ɔː/ have showed better agreement with the measured area functions in the pharynx region despite having larger cross-sectional areas (less than 20%) in most of the sampled sections.

In the oral cavity, the area functions of the two modelled fricative sounds (/s/ and /ʃ/) show similar patterns to the ones measured in the MRI, however the modelled cross-sectional areas are much larger than the measured ones whose narrowest section is approximately half the size of the modelled versions. A possible cause for this may be the different volume rates in the two speech environments (artificially sustained in the MRI versus normal speech in the EMA) and this issue will be discussed in more detail in Chapter 6. The three simulated vowels models are shown to produce cross-sectional areas within 20% variation to the corresponding measured VTs in most of the sampled sections in the oral cavity.

The five cubic models generally show a good match to their respective linear models in most of the sampled cross-sections. It can be seen in Figures 4.11 and ?? that the cubic mesh has a much smoother surface and fewer elements than those produced with linear elements, which makes it more suitable for graphics rendering and numerical modelling [91]. The cubic meshes of /s/ and /ʃ/ have been made into plastic casts for the PIV (Particle Image Velocimetry) study of VT flows in [56].

On the other hand, the C¹ continuity property of the cubic Hermite function may act as a constraint for modelling sharp geometry features such as the teeth; this is evident in the modelling of /s/ where the cubic model shows twice the cross-sectional areas in Slice 19 (around the teeth region) compared with the linear model. Local refinement may be used to bring more flexibility to the cubic mesh, however interceptions are likely to occur during least-squares fitting in regions where the two surfaces are too close to each other, such as in the case of the bicubic VT models presented in [78] where some spatial derivatives of the affected elements had to be manually edited to correct the defects. For modelling the teeth on the maxilla and jaw meshes, nodal versions were implemented and manually toned to fit the teeth contour in cubic models created in [92]. To fully represent the shape of teeth in the cubic VT model, similar arrangements may need to be implemented, or alternatively converted to use linear elements so that no nodal derivatives needs to be specified.
Figure 4.18: Comparison of the area functions measured by MRI and mathematical models.

(a) The rest breathing state (a) /s/; (b) /ʃ/; (c) /i:/; (d) /æ:/; (e) /õ:/.
4.5 Summary

The newly developed overlapping space partition scheme is implemented in a flood-fill algorithm for segmenting the VT domain from the surrounding articulator surfaces of the ‘Talking Head’ model. The entire process is largely automated to minimise human intervention. The resulting boundary representations and segmented data clouds are used by a contour-deformable surface reconstruction algorithm which has been demonstrated on the VT models for six speech and rest configurations. A method of volumetric reconstruction for the VT has also been developed with a subdivided least-squares fitting algorithm and cubic Hermite FE basis functions.
Chapter 5

Aeroacoustic Modelling Framework

To simulate the sound, a new finite element approximations of the Lighthill aeroacoustic analogy is used in the context of incompressible flows. The FE model comprises of a two-stage modelling where the incompressible Navier-Stokes equations are numerically resolved in order to provide the acoustic source field used in solving the Lighthill equation in variational forms. Three variations of the Lighthill acoustic source fields are investigated, including the Reynolds stress, the Reynolds stress plus the viscous stress and the Lagrangian multiplier. Section 5.1 provides a brief overview of the existing modelling methods with an emphasis on vocal tract modelling. Section 5.2 validates the CFD code against a benchmark test problem of a backward-facing step, Section 5.3 introduces the numerical method for the CAA simulation, which is applied to the numerical study of the 2D spinning vortices. Complete two-stage modelling is demonstrated for the vortex shedding from a 2D square cylinder in Section 5.4.

5.1 Aeroacoustic Modelling of Human Speech Sounds

In the previous chapter, anatomically realistic vocal tract models were obtained for the speech articulation. Based on the calculated area functions for the vowels and fricatives, presented in Chapter 4, and taking account of the volume flow rate for an average adult male during normal speech (200-700cm³/s [2]), flow Reynolds number has been estimated between 400 and 12,000 and the maximum Mach number is approximately 0.1, which lies within the field of low Mach and low-moderate Reynolds number aeroacoustics.

5.1.1 Empirical Modelling

In the field of speech modelling, many researchers in the past have developed empirical models to study the aeroacoustics of VT. Most such models are based on the ‘source-filter’ decomposition method first developed by Fant [11], which has been briefly touched upon in Chapter 1. In this framework, the ‘source’, refers to the production of acoustic energy, that are generally classified as voiced or unvoiced. The voiced source is produced by glottal vibration [2]. The mechanism of
glottal vibration has been modelled as a self-oscillating system such as the spring-mass-damper model (also referred to as the one mass model) [153], the two mass model [154, 155], the three mass models [156], where physiological properties (e.g. shape, tissue mass and stiffness, etc.) have been used for simulating vibrations. The vibration is driven by the pressure gradient across the glottis and is converted into a steady stream into a train of pulses which provide excitations to the VT [157].

The unvoiced sound source generally associated with the production of consonants is a kind of ‘turbulence’ noise [2], which is caused by the velocity fluctuations created by forcing air through the narrow constrictions that may be formed at various locations along the VT and between different articulators. The mathematical representation of this type of sound source is often derived via inverse-filtering of the experimentally collected data from mechanical models which are designed to emulate the VT geometries for fricative sounds and many models have been proposed in the literature [49, 53, 47, 54, 158].

The sound source for stop consonants has been modelled in two phases where an initial ‘burst’, lasting for the first 0.5ms after the release of the obstruction, is followed by transient turbulence noise [2]. In [159], the close-release motion of the VT during stop sounds was modelled as a constricted tube of a fixed length initially closed at one end and with changes to its cross-sectional area according to a linear function of time as the blockage is lifted. Under such assumptions, the ‘burst’ sound can be calculated as a combination of the acoustic impedance, resistance, dissipation of the energy to the wall, and acoustic mass of the air in the constriction site [159].

5.1.2 Direct Numerical Simulation

With the increasing computing power of modern HPC (high performance computers), attempts have been made to apply Direct Numerical Simulation (DNS) for modelling sound. At first glance, DNS seems to be a straightforward option as it is consistent with the exact physics principles and resolves the flow and acoustic fields simultaneously in one coupled system of compressible Navier-Stokes equations. The system described by a set of equations, representing the conservation relations of the mass, momentum and energy, can be written as

\[
\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_i)}{\partial x_i} = 0,
\]

\[
\frac{\partial (\rho u_j)}{\partial t} + \frac{\partial (\rho u_i u_j)}{\partial x_i} + \frac{\partial p}{\partial x_j} - \frac{\partial \sigma_{ij}}{\partial x_i} = 0,
\]

\[
\frac{\partial E}{\partial t} + \frac{\partial (u_i E + u_i p - u_j \sigma_{ij} + q_i)}{\partial x_i} = 0,
\]

where \( \rho, u_i, p, \sigma, E, q \) are the density, velocity, pressure, viscous stress, energy density and heat flux respectively. For Newtonian flows, the viscous stress can be calculated as \( \sigma_{ij} = \lambda \frac{\partial u_i}{\partial x_j} - \mu \frac{\partial u_j}{\partial x_i} \) where the constants \( \lambda \) and \( \mu \) are the coefficients for the dilatational and deviatoric
components of the viscous stress respectively. To complete the system, an equation of state such as the perfect gas law must be specified, which can be written as

$$p = \rho(\gamma - 1)(E - \frac{1}{2}u^2),$$

(5.1.4)

where $\gamma$ is called the adiabatic index.

Among attempts at DNS modelling for speech production, compressible Navier-Stokes equations have been numerically solved in an idealised rigid 2D axisymmetric model of the human larynx, using a sixth-order compact finite difference scheme [160, 161]. In the simulation, the flow Mach number was artificially increased. Based on the results, the authors found that the dipole source, due to the unsteady force acting on the glottis wall, is primarily responsible for sound generation, whose strength is regulated by the area ratio between the glottis orifice and the duct.

In 2005, Nomaura and others applied their compressible viscous fluid model to a 2D rigid glottis model [162]. Their findings suggest that, although the initial glottal flow might be described as a symmetrical jet, it quickly develops into highly complex vortical structures spread in 2D space. For a similar geometry, Larsson and his team [163] investigated the vortex evolution using a high-order finite difference scheme and the main conclusion is that the previous axis-symmetric model tends to over-estimate the vorticity production compared to a complete 2D model. 2D and 3D models of fricative configurations have been simulated with DNS in [50] for a number of numerical methods including RANS (Reynolds Averaged Navier-Stokes) [164] and LES (Large Eddy Simulation) [165]. Their results suggest that the 2D geometry seems to over-predict the acoustic strength compared to a 3D model and the RANS method is unable to predict broadband noise like LES [50].

Despite the ability to fully resolve the physics, the DNS method is still not widely used for solving aeroacoustic problems in low Mach number flows. Studies have shown major disparities between the aerodynamic and aeroacoustic fields in low Mach number flows, which may be summarised as below:

1. **Time scale disparity:** the nature of the aeroacoustic problem requires transient solutions of the flow. The human ear responds to a wide range of frequencies from 50Hz to 20kHz with the peak frequencies centered around 1-2kHz. These frequencies are generally much higher than those considered as important in aerodynamic simulations to resolve the flow [166].

2. **Length scale disparity:** the wavelength of the radiated sound is generally much larger than the physical length of the aerodynamic source ($l$) (e.g. the eddies) [167].

3. **Energy scale disparity:** the acoustic energy is generally much smaller ($O \sim M^4$) than the aerodynamic energy in the near-field [167].
Due to these disparities, DNS methods for the Computational AeroAcoustics (CAA) domain often have different numerical requirements (e.g. time and space resolution, etc) than for CFD problems in low Machs number flows [168]. Moreover, DNS incurs a very high computational cost and is not feasible for large scale aeroacoustic simulations [169]. As a result, researchers have resorted to new methodologies, known as hybrid methods where the problem is solved as separate systems consisting of independent aerodynamic and aeroacoustic domains. Depending on the way the two domains are defined, such models may be further classified as perturbation methods and analogy-based methods.

5.1.3 Hybrid Methods based on the Perturbation Equations

The first group of hybrid methods is directly derived from the fundamental principles. The methods, generally termed perturbation equations, divide the problem into the aerodynamic ‘mean’ and acoustical ‘perturbation’. The ‘mean’ is considered as large scale components of the fields and can be modelled independently to the relatively small ‘perturbation’ fields. Therefore perturbations can be solved in a set of transport equations based on the resolved ‘mean’. There are many different formulations based on the mean-perturbation decomposition, which differ by the definition of the three essential elements of the model, the mean, the perturbation and the acoustic source. One of the groups is called the linearised Euler equations (LEE) [170] which can be viewed as a linearised form of Equations (5.1.1),(5.1.2) and (5.1.3), assuming no heat conduction and neglecting the viscous forces,

\[
\frac{\partial \bar{\rho}^\prime}{\partial t} + \frac{\partial (\bar{u}_j \rho^\prime + u_j^\prime \rho)}{\partial x_j} = 0,
\]

\[
\frac{\partial \rho^\prime}{\partial t} + \frac{\partial (\bar{u}_j \rho^\prime + u_j^\prime \rho)}{\partial x_j} + (E + \bar{p} + \rho^\prime \delta_{ij}) \frac{\partial \bar{u}_j}{\partial x_j} = S_i,
\]

\[
\frac{\partial E'}{\partial t} + \bar{u}_j \frac{\partial E'}{\partial x_j} + (\bar{E} + \bar{p}) \frac{\partial u_j'}{\partial x_j} + (E' + p') \frac{\partial \bar{u}_j}{\partial x_j} + u_j' \frac{\partial (\bar{E} + \bar{p})}{\partial x_j} = 0,
\]

where the perturbations \((\rho^\prime, u_i^\prime, p^\prime)\) are assumed to be related to the acoustic field while the mean quantities \((\bar{u}_i, \bar{\rho}, \bar{p})\), have been modelled as an analytic function derived by fitting experimentally collected data [171], an incompressible LES [172], compressible LES [173], incompressible RANS [174] and others. The source terms used in the right hand side of the LEE, are derived from Lilley’s equation [175] \((S_i = -\partial \rho u_i' u_j'/\partial x_j)\) which represents the nonlinear velocity fluctuations [173], acting as a stress term in the momentum equations. Given different choices for the ‘mean’ (e.g. incompressible versus compressible flows, steady versus unsteady flows, etc), it is not always clear what relationship exists between the ‘perturbation’ field and the actual acoustic signals because the LEE describes not only the propagation of acoustic waves but also the transportation of vorticity and entropy waves [168]. Another significant short-coming of the LEE is the lack of feedback between the mean and perturbation fields [176], which may
lead to unphysical acoustic production due to hydrodynamic instabilities [177]. Therefore, other perturbation-based methods have been proposed where perturbation fields and acoustic loadings are carefully defined to represent the acoustic signals.

The acoustic/viscous splitting method proposed by Hardin and Pope in [178] is designed to adapt the low Mach condition and relies on incompressible flows to be the ‘mean’. The main difference in this approach compared to the previous LEE formulations is the introduction of a hydrodynamic density to account for the pressure fluctuations due to periodic flows, which, according to the authors, is quite large compared to the acoustic quantities, therefore they need to be removed from the solutions of the incompressible flow [179].

The approach has been further developed into the PCE (perturbed compressible equations) method [180], where the authors developed a remedy to the inconsistencies in the previous method due to the interactions between the incompressible vorticity and the perturbed velocity in the near wall region, by adding a viscous term to the perturbed momentum equations in order to dissipate the perturbed vorticity [180]. Furthermore, in [181], Seo and others converted the PCE method into the LPCE (Linearised Perturbed Compressible Equations) in which the author claimed that the numerical stability issue caused by perturbed vorticity is completely solved by dropping the related source terms.

In a separate development, a group of formulations belonging to the Acoustic perturbation equations (APE), are derived to suit different flow conditions. In [177], four derivatives of the APE systems are introduced (APE1-4). APE-1 is built on a unsteady compressible mean flow and decomposes the perturbed velocity fields into solenoidal and irrotational parts of which the former is said to be associated with the vortical mode of the flow while only the second part of the velocity field contributes to the acoustic field, therefore the source term is filtered in order to suppress the vorticity waves [177]. APE-2 deals with unsteady incompressible flows as the mean and accordingly, the pressure perturbation field defined in APE-1 is decomposed in order to remove the hydrodynamic part of the pressure fluctuation in a similar way to the PCE method. In the APE-3 [182], the acoustic perturbation field is re-defined as the perturbed total enthalpy \( H = \bar{h} + h' \) and APE-4 introduces the Lamb vector \( L' = (\omega \times u') \) as the primary form of the acoustic source in low Mach number flows [183].

In the speech modelling field, Bae and his colleagues modelled the aeroacoustic field in an axisymmetric glottis model by employing a splitting method in which the flow field is resolved by the incompressible Navier-Stokes equations and the acoustic field is resolved by a perturbed compressible equation [184]. He concluded that the quality of the voice is closely related to the vortical structures in the shear layer of the pulsating jet. Later in [185], a 2D axi-symmetric grid was applied with the LES/LPCE method to simulate a vibrating larynx at a fixed frequency and the author states that there exists some correlation among motion of the vocal folds, the dynamics of vortices and the resulting acoustic signals. Three-dimensional VT configurations for two vowels /i/ and /u/ based on the area functions measured in [65], have been used in the simulation of perturbation equations akin to PCE and APE-2 in [186] where the incompressible
flow was numerically solved only in the larynx region while the acoustic domain was extended to cover the rest of VT and part of the open space.

5.1.4 Hybrid Methods based on the Aeroacoustic Analogies

The second group of hybrid methods, which also has been widely used by researchers for modelling flow-induced sound especially for low Mach number flows, may be called ‘analogy-based’ methods. Under such schemes, the acoustic fields are decomposed into the ‘near field’ where all the acoustic-related aerodynamics are collectively grouped into a ‘source’ which provides the driving perturbation to an acoustic ‘far field’ where the plane wave propagation is assumed.

There are different aeroacoustic analogies which are all derived from the same fundamental principles but differ in either the mathematical representation of the source term or the definition of the acoustic variables. One of the most famous is called the Lighthill aeroacoustic analogy [187]. The equation can be arrived at by differentiating the equations of conservation of mass (Equation (5.1.1)) with respect to time, taking the divergence of the momentum equations (Equation (5.1.2)), combining the two scalar equations and subtracting $-c_0^2 \nabla^2 \rho'$ from both sides of the combined equation. The final form can be written as

$$\frac{\partial^2 \rho'}{\partial t^2} - c_0^2 \frac{\partial^2 \rho'}{\partial x_i \partial x_i} = \frac{\partial \left( \rho u_i u_j - \sigma_{ij} + (p - c_0^2 \rho) \delta_{ij} \right)}{\partial x_i \partial x_j} = \frac{\partial T_{ij}}{\partial x_i \partial x_j},$$

(5.1.8)

where on the left hand side, the propagation of the acoustic field is defined as the density perturbation $\rho'$ traveling at the speed of sound in a stagnanted medium $c_0$, and on the right hand side, there are three aerodynamic source terms in a group called the Lighthill stress tensor $T$. From the left to right these terms are: the Reynolds stress made of vector products of the velocity field $u_i$, the viscous stress $\sigma_{ij}$, and the nonlinear part of the constitutive relation between the pressure $p$ and density $c_0^2 \rho$.

Other commonly used derivatives of Lighthill’s analogy includes Curle’s equation [188] where a surface integral on any hard surfaces is explicitly separated from the volume source in the acoustic source region; the Ffowcs Williams-Hawkings equation [189] extended Curle’s formulation by introducing a permeable surface for the boundary integral, as

$$\frac{\partial^2 \left( (\rho')H(S) \right)}{\partial t^2} - c_0^2 \frac{\partial^2 \rho'}{\partial x_i \partial x_i} H(S) = \frac{\partial T_{ij}}{\partial x_i \partial x_j} H(S) - \frac{\partial (F_{ij} \delta(S))}{\partial x_j} + \frac{\partial (Q_i \delta(S))}{\partial t},$$

(5.1.9)

where

$$T_{ij} = \rho u_i u_j + (p \delta_{ij} - \rho c_0^2) - \sigma_{ij};$$

(5.1.10)

$$F_{ij} = p \delta_{ij} - \sigma_{ij} + \rho u_i (u_j - \hat{u}_j);$$

(5.1.11)

$$Q_i = \rho_0 u_i + \rho_0 (u_i - \hat{u}_i).$$

(5.1.12)
The terms on the right hand side represent sound sources as quadrupoles, dipoles and monopoles. \( \dot{u}_i \) is the velocity of the wall (boundary of the domain) whose surface is expressed as a function \( S \), while \( u_i \) is the velocity of the fluid. The dipoles (\( F_{ij} \)) and monopoles (\( Q_i \)) only exist on the boundary as they are multiplied by \( \delta(S) \) (the impulse function). \( T_{ij} \) represents a volume source in the flow region (\( H(S) = 1 \) in the fluid domain).

There are also the vorticity-based analogies [190, 191] which take a mathematical form of the source terms derived from the vorticity transport equations. All these analogies adapt the homogeneous wave operator to form the left hand side of the equations while all of the right hand sides contain acoustic variables, assumed to carry significant effect only in a relatively small domain coinciding with the core of the flow. Therefore, in practice, mathematical models based on such analogies normally divide the problem into two non-interacting domains where the acoustic field is resolved based on the source field evaluated in a separate flow domain, in a similar way to the aforementioned perturbation methods.

In the speech modelling field, McGowan studied the glottal flow experimentally and with a numerical model based on Powell’s aeroacoustic analogy [192]. It was hypothesised that in addition to the acoustic energy exchange via the vibrating glottis, strong vorticity-velocity interactions also produce significant amounts of acoustic energy during phonation. In [193], Suh adopted the Ffowcs Williams-Hawkings aeroacoustic analogy, coupled with compressible LES, for predicting the acoustic field resulting from a pulsatile flow through a rigid glottis. In this study, he concludes that the acoustic source on the glottal walls are the main contributors to the far-field noise. Link and his co-workers [194] have been using the Lighthill analogy combined with the ALE (Arbitrary Lagrangian Eulerian) method in modelling human speech for a 2D dynamic glottis model. In their model, the near-field Lighthill stress is represented by the lone Reynolds stress which is evaluated from the numerically resolved solutions of incompressible Navier-Stokes equations. A similar two-stage modelling strategy which combines the incompressible flow and the Lighthill aeroacoustic analogy, was applied to 3D vibrating vocal folds [195].

After comparing the perturbation methods with the acoustic analogies, it is found that:

1. Both are derived from fundamental principals, therefore they should be all exact in their complete forms;

2. The main difference between the two groups of methods lies in the treatment of acoustic propagation, where the analogy-based methods treat any non-linear wave propagation of the acoustic signals as part of the ‘source’ while those perturbation-based methods explicitly account for various non-linear transportation phenomenons (e.g. diffraction);

3. In general, due to the extra terms on the left hand side, the perturbation equations are more computationally demanding than the numerical methods based on acoustic analogies while both methods are computationally more efficient than the DNS [176].
In a speech environment, the acoustic source is generally localised in a small part of the VT [11, 49] while the listening point (i.e. the acoustic far-field) is normally placed some distance away from the mouth opening, in areas where the air can be treated as a static medium. Under such conditions, the plane wave operator adopted by the aeroacoustic analogy is more suitable for describing the acoustic far-field for the speech sounds, and thus it is better to allow the separation of the computation procedures for modelling the source and propagation fields. For this reason, an aeroacoustic analogy is used for building the framework for the numerical model introduced in the following sections.

5.2 Finite Element Method for Unsteady Viscous Incompressible Navier-Stokes Equations

A finite element code for solving the incompressible Navier-Stokes equations, called CMheart, was written in Fortran90 platform and developed by David Nordsletten at the University of Oxford [196], which uses the MPICH2 library [197] for parallel computing. The code was originally written for 3D numerical problems and has been modified to solve 2D incompressible flows. The solver may be explained via the pseudo-code in Figure 5.1.

![Figure 5.1: The pseudo-code for the CFD solver.](image)

The process starts with building the FE fields, including the geometry \( x \), velocity \( u \), Lagrangian multiplier \( P \) and their derivative fields \( \partial u / \partial x \), \( \partial P / \partial x \), \( \partial u / \partial t \) (defined in Section 5.2.2), and user-specified boundary conditions (i.e. Dirichlet and Neumann types, as defined in Section (5.2.1)). The FE fields are calculated at each Gauss point for numerical integrals and then used to build the matrix systems for the governing equations (Equations (5.2.2),(5.2.3)).
In FEM, the governing equations are numerically solved as a system of linear equations, such as

\[ A_{ij}B_j = C_i, \quad (5.2.1) \]

where \( A_{ij} \) is called the coefficient matrix which is calculated from both the linear and non-linear parts of the governing equations; \( B_j \) is a vector of nodal variables (i.e. \( u_i \)) and vector \( C_i \) contains the specified boundary conditions. Given a static grid, the linear part of the \( A_{ij} \) has constant values, therefore the subroutines for calculating the terms \( a, b, b^* \) and their temporal derivatives (defined in Section 5.2.1) are left outside the main loop and only calculated at the beginning of the code.

The non-linear part of the \( A_{ij} \) is due to the advection term \( c \) (defined in Section 5.2.1) in the Navier-Stokes equations. As the values of the \( c \) are dependant on the variable field itself, the subroutine for calculating the operator \( c \) is embedded in the iterative Newton-Raphson solver which reaches the convergence until the numerical errors (ERR) drops below specified tolerance (TOL) (defined in Section 5.2.3). This system of linear equations is resolved in the numerical package MUMPS (Multifrontal Massively Parallel Solver) [198].

### 5.2.1 Weak Form

Assuming that the fluid is incompressible, the constitutive relation required by the compressible system, can be eliminated and replaced by a divergence-free constraint on the velocity field. The incompressible system can be formulated as

\[ \frac{\partial u_i}{\partial x_i} = 0, \quad (5.2.2) \]

\[ \frac{\partial u_j}{\partial t} + \rho_0 \frac{\partial u_i u_j}{\partial x_i} = -\frac{\partial P}{\partial x_j} + \frac{\partial \sigma_{ij}}{\partial x_i} + f_i. \quad (5.2.3) \]

Here the ‘true’ pressure \( p \) has been replaced by \( P \), known as the Lagrangian multiplier, which is a redundant quantity whose only purpose is constraining the velocity field to be in the divergence-free plane, the viscous stress is reduced to consist of the only deviatoric component for a Newtonian fluid, \( \sigma_{ij} = \nu \frac{\partial u_j}{\partial x_i} \) where \( \nu \) is the kinematic viscosity, and the \( f_i \) represents the external force.

Applying the variational principle to the above equations, the following formulae for the Navier-Stokes equations can be reached

\[ (u_t, \phi) - a(u, \phi) - b(P, \phi) - c(u, u, \phi) = (f, \phi) + d(\sigma, \phi), \quad \forall u, \phi \in H^1(\Omega) \quad \text{and} \quad \forall P \in L^2(\Omega); \quad (5.2.4) \]

and for the equation of conservation of mass,

\[ b^*(u, \Psi) = 0, \forall \Psi \in L^2(\Omega), \quad (5.2.5) \]
where the above operators are defined as the following:

\[
(u_t, \phi) = \int_{\Omega} \frac{\partial u_i}{\partial t} \phi_i d\Omega; \quad (5.2.6)
\]

\[
a(u, \phi) = -\nu \int_{\Omega} \frac{\partial u_j}{\partial x_i} \frac{\partial \phi_j}{\partial x_i} d\Omega; \quad (5.2.7)
\]

\[
b(P, \phi) = -\int_{\Omega} P \frac{\partial \phi_i}{\partial x_i} d\Omega; \quad (5.2.8)
\]

\[
c(u, u, \phi) = \int_{\Omega} u_j \frac{\partial u_i}{\partial x_j} \phi_i d\Omega; \quad (5.2.9)
\]

\[
d(\sigma, \phi) = \nu \int_{\Gamma} \frac{\partial u_j}{\partial x_i} \phi_j n_i d\Gamma - \int_{\Omega} P \phi_i n_i d\Gamma; \quad (5.2.10)
\]

\[
(f, \phi) = \int_{\Omega} f_i \phi_i d\Omega; \quad (5.2.11)
\]

\[
b^*(u, \Psi) = \int_{\Omega} \frac{\partial u_i}{\partial x_i} \Psi d\Omega. \quad (5.2.12)
\]

### 5.2.2  Numerical Discretisation of the FE Fields

#### FE Discretisation

In the Garlerkin FE method [87], the variable fields are approximated with the same mathematical functions used for the trial functions and the numerical fields and their derivative fields for the velocity, the Lagrangian multiplier and the space, in each element (subspace \( \Omega_h \)), can be defined as

\[
u_i(\Omega_h) = \sum_{j=1}^{n} \phi_i^j u_i^j; \quad (5.2.13)
\]

\[
P(\Omega_h) = \sum_{j=1}^{m} \Psi^j P^j; \quad (5.2.14)
\]

\[
x_i(\Omega_h) = \sum_{j=1}^{n} \phi_i^j x_i^j; \quad (5.2.15)
\]

\[
\phi(\Omega_h) = \sum_{i=1}^{n} \phi_i; \quad (5.2.16)
\]

\[
\Psi(\Omega_h) = \sum_{i=1}^{m} \Psi_i; \quad (5.2.17)
\]

\[
\varphi(\Omega_h) = \sum_{i=1}^{n} \varphi_i, \quad (5.2.18)
\]

where \( \phi, \Psi \) and \( \varphi \) are the chosen basis functions for the velocity, Lagrangian multiplier and spatial field respectively. In all of the CFD examples given in this thesis, Talyor-Hood type elements have been chosen for \( u \) and \( P \) in order to meet the LBB (Ladyzhenskaya-Babuska-Brezzi) condition [199], which provides sufficient conditions for the stability of the numerical
computation.

**Spatial Derivatives**

The FE formulae require the calculation of first order spatial derivatives of the velocity field, which can be calculated as

\[
\frac{\partial u_i(\Omega_h)}{\partial x_j(h)} = \sum_{k=1}^{n} \frac{\partial u^k_i}{\partial x^j_k} \partial x^k_j; \quad (5.2.19)
\]

\[
\frac{\partial u_i(\Omega_h)}{\partial \xi_j} = \sum_{k=1}^{n} \frac{\partial \phi^k_i}{\partial \xi_j} u^k_i; \quad (5.2.20)
\]

\[
\frac{\partial x_i(\Omega_h)}{\partial \xi_j} = \sum_{k=1}^{n} \frac{\partial \phi^k_i}{\partial \xi_j} x^k_i, \quad (5.2.21)
\]

and \(\partial \xi_i/\partial x_j\) can be evaluated via the matrix inversion operation

\[
\frac{\partial \xi_i}{\partial x_j} = \left[ \frac{\partial x_i}{\partial \xi_j} \right]^{-1}. \quad (5.2.22)
\]

**Numerical Integration**

As the FE formulation is in an integral form, the discretised form of Equations (5.2.5) and (5.2.4) are integrated over each elemental space for the entire computational domain (\(\Omega\)), which is done by a numerical technique called Gaussian quadrature [200] that approximates the function space by a set of polynomials of a certain order. Numerical integration by Gaussian quadrature may be expressed as a weighted sum of some point values of the original function,

\[
\int_{\Omega} f(x, t)dx \approx \sum_{i=1}^{m} w_i f(x, t), \quad (5.2.23)
\]

where \(m\) is the number of Gauss points and \(w_i\) is a weighting coefficient. The minimum number of Gauss points suggested by Juan [201] is that the squares of the first order derivatives of the basis function should be integrated exactly by the quadrature, which gives \(m = 2K - 1\) for a basis of the order of \(K\) (i.e. 3 Gauss points per dimension for a quadratic function).

**Time Discretisation**

The implicit backward Euler finite difference scheme [202] is chosen for time stepping. For a fine enough spatial discretisation, a temporal convergence proportional to \(\delta t\) is expected. The scheme is also unconditionally stable for any step size. In this scheme the first-order time derivative present in the Navier-Stokes equation (5.2.4) becomes

\[
\frac{\partial u_j}{\partial t} \approx \frac{u_j^t - u_j^{t-\delta t}}{\delta t}, \quad (5.2.24)
\]
where \( \_t \) refers to the fields at the current time step while \( \delta t \) is the chosen step size. The discrete form of Equation (5.2.4) then becomes

\[
\delta t(u_t^i, \phi_i) - a(u_t^i, \phi_i) - b(u_t^i, u_t^i, \phi_i) = \delta t(u_t^i-\delta t, \phi_i) + (f_t^i, \phi_i) + d(\sigma_t, \phi_h). \tag{5.2.25}
\]

On the other hand, Equation (5.2.5) is always solved at the current time step,

\[
b^*(u_t^i, \Psi) = 0. \tag{5.2.26}
\]

### 5.2.3 Newton-Raphson Technique

As there is a non-linear system of equations, the commonly used Newton-Raphson technique is employed as the root searching method [203]. The method is based on the first order Taylor series expansion of the root \( f(u) \) around its current state \( u - \delta u \). The expansion can be written as,

\[
f(u) = f(u - \delta u) + \delta u \frac{\partial f(u - \delta u)}{\partial u} + O(\delta u^2) = 0, \tag{5.2.27}
\]

where \( O(\delta u^2) \) means the truncation of an infinite series from the second order and higher. The solution might be reached by solving the systems of equations,

\[
\frac{\partial f(u - \delta u)}{\partial u} \delta u = -f(u - \delta u), \tag{5.2.28}
\]

where a Jacobian matrix is defined as \( \frac{\partial f(u)}{\partial u} \).

Inside the Jacobian matrix \( \frac{\partial f(u - \delta u)}{\partial u} \), the contribution from the \( a \) operator is

\[
\frac{\partial a}{\partial u_i} = \nu \int_\Omega \frac{\partial \phi_i}{\partial x_j} \frac{\partial \phi_i}{\partial x_j} \, d\Omega. \tag{5.2.29}
\]

The contribution from the \( b \) operator is

\[
\frac{\partial b}{\partial P} = \int_\Omega \frac{\partial \phi_i}{\partial x_i} \Psi \, d\Omega. \tag{5.2.30}
\]

The contribution from the \( b^* \) operator is

\[
\frac{\partial b^*}{\partial u_i} = \int_\Omega \frac{\partial \phi_i}{\partial x_i} \Psi \, d\Omega. \tag{5.2.31}
\]

And the contribution from the \( c \) operator is

\[
\frac{\partial c}{\partial u_i} = \int_\Omega \phi_i \left( \frac{\partial u_i}{\partial x_j} \phi_j + \frac{\partial \phi_i}{\partial x_j} x_j \right) \, d\Omega. \tag{5.2.32}
\]
The contribution from the time derivative is
\[ \frac{\partial(u_t, \phi)}{\partial u_i} = \int_{\Omega} \frac{\partial \phi_i}{\partial t} d\Omega. \] (5.2.33)

A global line searching and bracket-tracking scheme has been implemented to improve the convergence of the Newton-Raphson method, which is described in detail in [204]. The technique is based on a third order approximation of a scale factor (\(\lambda\)) to the Newton step (\(\delta u\)) as
\[ u_{\text{new}} = u_{i-1} + \lambda \delta u, \quad 0 < \lambda < 1; \] (5.2.34)
where the objective function is modeled as the cubic function of \(\lambda\) which is solved in an iterative fashion [204].

The convergence criteria for the steady state and within each time step in transient simulations are met by satisfying any one of the three of the following conditions [196]:

1. \(L_2\) \(-\) norm of the increments \(< 1 \times 10^{-10}\): If the \(L_2\) \(-\) norm of the total residual from the right hand side drops below the tolerance;

2. \(L_2\) \(-\) norm of the increments / \(L_2\) \(-\) norm of the solution \(< 1 \times 10^{-13}\): If the \(L_2\) \(-\) norm of the total residual scaled by the \(L_2\) \(-\) norm of residual due to the boundary conditions drops below the tolerance;

3. \(L_2\) \(-\) norm of the residual of the functions \(< 1 \times 10^{-9}\): If the \(L_2\) \(-\) norm of the solution increments drop below the tolerance.

5.2.4 Test Problem 1: The Backward-Facing Step

The 2D backward-facing step has been studied both experimentally and numerically in [205] whose results are chosen to validate the CFD code used in this thesis. The computational domain is outlined in Figure 5.2(a) where the step has an expansion ratio of 2 and the upstream chamber is 10 times the length of the step height in order to allow the development of inlet the flow which is starting from a uniform inflow profiles normal to the inlet while the stress-free outlet (\(\partial u_i/\partial n_i = 0\)) is placed 100 units downstream to the step. The remaining the boundary is specified as non-slip wall (\(u = 0\)).

The problem is simulated for a range of Reynolds numbers (Re) between 100 and 2,000 on structured grids made of homogeneous quadrilateral cells where the Q-Q-L (quadratic-quadratic-linear) basis functions are chosen for the spatial \(x\), velocity \(u\) and Lagrangian multiplier \(P\) fields respectively. The results of a mesh convergence analysis for the highest Re are given in Figures 5.3(a) and (b) where the plots show that \(\| u - u^h \|\) and \(\| \partial(u - u^h)/\partial x \|\) start to behave linearly with \(h^3\) and \(h^2\) respectively from the third grid size (\(h\)). For the final solutions, the mesh is refined until both \(\| u \|_{\text{max}}\) and total \(\| \omega \|\) in the expansion region vary.
Figure 5.2: The CFD domain of the backward-facing step problem. Notice the geometries showing are not to scale. (a) The boundary conditions where the inlet is marked in red; the non-slip wall is in black and the outlet is in green. (b) The streamlines of the steady-state of flow at Re = 1,000, where the magnitude of the velocity is normalised by $u_\infty$.

by less than 0.1% compared to the results obtained from the previous grid size, shown in Figure 5.3.

The steady-state results are compared with both the experimental data collected by Armaly [205] and the numerical results conducted by Barton in [206], as shown in Figure 5.4, where the parameters of interest (e.g. separation and reattachment lengths) are defined in Figure 5.2(b). For Re = 100 to 600, the results show good agreement with the other two studies. For higher Re, experimental results begin to diverge from those predicted by the 2D numerical models, owing to the three-dimensionality of the physical flow [205]. A comparison with another numerical study presented by Kaiktsis in [207] shows that both models predict that the lengths of both the primary and secondary recirculation zone (e.g. $x_1=18.5$ in [207], 13.5 in [205] and 17.0 in the present model) at Re =2,000 is significantly longer than Armaly’s experimental results. Kaiktsis’ model also shows that this type of 2D flow is time-independent up to Re = 2,500.
Figure 5.3: The mesh convergence analysis at Re=2,000. (a) The maximum magnitude of the velocity normalised by the inlet velocity $U_\infty$ over the domain $\Omega$ versus the minimum cell size $h$. (b) The total circulation (vorticity, $\omega$) in the region after the step normalised by $U_\infty$ and step height versus the cell size.

Figure 5.4: The comparison of the steady-state solutions at Re=100-600. All measured lengths are normalised by the step height $d$. 
5.3 Finite Element Method for the Lighthill Aeroacoustic Analogy

The variational formulae based on the Galerkin FE method, are applied for solving the modified form of the Lighthill equation (Equation (5.3.8)) in the newly developed Fortran code (AcousticParallel). The solver is divided into two parts: the calculation of the acoustic source (AcousticSource) and the computation of the inhomogeneous wave equations (AcousticSolve). The two-stage strategy of the acoustic solver allows the decoupling of the numerical schemes (both spatially and temporally) for resolving the aerodynamic and acoustic fields, which can be useful for accommodating the different numerical requirements for the two problems. As demonstrated in Figure 5.5, the aerodynamic quantities of incompressible flows are used as inputs to AcousticSource for evaluating acoustic loads at each node of the FE grid which may be different from the one used for the CFD, created in a separate package. In the examples provided in this thesis, the acoustic domains either share the same numerical grid with the aerodynamic domain or an extended version of it with extra elements padded around the edge. AcousticSolve performs the numerical procedures for solving the Lighthill equations whose solution is produced in the form of normalised density $\tilde{\rho}$ as defined in Section 5.3.3.

![Figure 5.5: A schematic diagram of the procedures for the aeroacoustic solver.](image)

AcousticParallel adapts the MPICH2 library [197] and MUMPS numerical library for parallel computing. A user manual for AcousticParallel is provided in Appendix D while a pseudo-code for the package is summarised in Figure 5.6.

The two packages, AcousticSource and AcousticSolve share many of the same I/O (input/output) and FE modules (defined in Section 5.3.4) in their codes, except for the lack of a matrix solving module in the former. AcousticSource builds the nodal-based acoustic loads ($f(\tilde{T}, \phi)$ defined in Section 5.3.3), based on aerodynamic fields of the incompressible flows ($x, u, P$ defined in Section 5.2.2) and according to the three source models defined in Section 5.3.2. In AcousticSolve, the new FE formulae of modified Lighthill equation are numerically resolved as a system of linear equations. The left hand side of the equation consists of a wave operator, as defined in Section 5.3.3. Given a constant speed of sound ($c_0$), the coefficient matrix is time-invariant. However, the matrix needs to be updated for every time step after
Read mesh parameters;
Build FE fields for numerical integration;
Initialise the global time (GTIME);
DO WHILE GTIME <= Total number of time steps
  IF Source Model 1 THEN
    Calculate nodal acoustic loads for Model 1;
  ELSIF Source Model 2 THEN
    Calculate nodal acoustic loads for Model 2;
  ELSIF Source Model 3 THEN
    Calculate nodal acoustic loads for Model 3;
  END IF
  Output the solution field;
  Read mesh parameters;
  GTIME++;
END DO

Read mesh parameters;
Build FE fields for numerical integrations;
Build the left hand side of the Lighthill equation;
Build boundary conditions;
Initialise the global time (GTIME);
DO WHILE GTIME <= Total number of time steps
  Solve the system of linear equations;
  Output the solution field;
  IF GTIME < Total number of time steps THEN
    Read mesh parameters;
    Update the boundary conditions;
  END IF
  IF Add Mach number effect THEN
    Update the left hand side of the Lighthill equations;
  END IF
  GTIME++;
END DO

Figure 5.6: The pseudo-code for AcousticParallel which is made of two parts AcousticSource and AcousticSolve.

taking into account the Mach number effect (discussed in Section 5.4.5). The right hand side of the equation is made up of the acoustic loads previously calculated by AcousticSource and the acoustic boundary conditions, specified in Section 5.3.4.

5.3.1 Approximation of the Lighthill Analogy for Compressible and Incompressible Flows

At first glance, compressible flow should be the standard for modelling any aeroacoustic problem based on the Lighthill analogy, since the analogy is derived from the compressible flow system and sound waves are inherent to density perturbations. Lighthill in his original paper [187] stated that the contribution of the viscous stress to the $T_{ij}$ is probably as a slow dissipation
factor during wave propagation and hence is not important in the source region. Furthermore the difference between the \(\rho \delta_{ij}\) and \(c_0^2 \rho \delta_{ij}\) should also be unimportant in near-isothermal flows. Therefore, the principal acoustic generation is due to the fluctuating Reynolds stress and the Lighthill stress tensor can be simplified to

\[
T_{ij} \approx \rho_0 u_i u_j, \tag{5.3.1}
\]

with an error proportional to the square of the Mach number \(M^2\).

From then on, some work has been carried out to investigate the actual role of the viscous stress in the Lighthill stress tensor. Crighton [208] pointed out that the viscous contribution to the Lighthill stress is less than the Reynolds stress contribution by the order of \(M\), owing to the inefficient octopole radiation pattern. On the other hand, Morfey [209] suggests that the quadrupoles’ sound radiation can be greatly enhanced by the presence of rigid walls and the viscous shear stress is the dominant sound source in low Mach number boundary layers. This view is also supported by the numerical study conducted by [210] where the authors showed that the viscous shear stress can be an independent sound source, given that the length of the wall is much larger than the acoustic wavelength. In a study of a laminar flow over a cylinder [211], the author concluded that the viscous stress is a significant sound source in their implementation of Curle’s analogy.

Similarly for the third source term in the Lighthill stress, [212] discussed the dynamics of the enthalpy on the sound production in the context of the low Mach number jet flows. Based on numerical analysis and experimental findings, it was stated that the intensity of the jet noise can vary with \(M^6\) rather than the \(M^8\) law proposed in [187], and suggested that the discrepancy could be due to the enthalpy production related to the intensity of turbulent velocity fluctuations and that the enthalpy term also may be subject to a flow-acoustic interaction. The author stated that there is a need for a full expression of the Lighthill stress tensor for general modelling of cold and heated jets. The \(M^6\) relationship was further validated for hot jets [213]. A significant development was made by Freund who found that the enthalpy term, although relatively small, is not negligible even for nearly isothermal flows at modest Mach number [214].

In addition to the approximation proposed by Lighthill, there is often another approximation adapted by many numerical modellers in the implementation of the Lighthill analogy, which is the use of incompressible flows for the evaluation of the Reynolds stress [194, 215, 195, 216]. All of these models have achieved some degree of success in incorporating incompressible flows into acoustic simulations, however one crucial question remains about the effect of compressibility. Crow [217] presented a perturbation analysis of the incompressible approximation as an infinite series of the square products of the Mach number and showed that the leading error of such an approximation was proportional to \(M^2\). Ristorcelli derived a \(M^2\)-order correction term for the previous approximation by accounting for the interactions between the incompressible part of the Reynolds stress and the acoustic mode of the flow, assuming the ‘acoustic flow’ is irrotational.
In the next section, a new way of applying the Lighthill analogy to incompressible flows is presented.

### 5.3.2 Alternative View of the Lighthill Aeroacoustic Analogy

When incompressibility is assumed for low Mach number flows, the same mathematical steps taken by Lighthill are repeated and the following relation is reached

$$
\frac{\partial^2 \rho'}{\partial t^2} = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} \left( \rho_0 u_i u_j - \sigma_{ij} + P \delta_{ij} \right) = 0.
$$

(5.3.2)

It can be seen that the acoustic source field on the right hand side has been forced to zero due to the conservation relations, which is not surprising because acoustic waves are suppressed in the incompressible formulation. This suggests that there has to be some form of modelling of ‘compressibility’ in the incompressible solutions in order to extract sound from the flow.

Comparing the incompressible system (5.2.2) and (5.2.3) with the compressible system (5.1.1), (5.1.2) and (5.1.3) together with the Lighthill equation (Equation (5.3.2)), the following conclusions may be reached:

1. The conservation relations are unchanged between the incompressible and compressible systems, meaning that there is no extra mass, momentum or energy added or removed from the system by eliminating the constitutive relation;

2. The $P$ is no longer the ‘true’ pressure and coupled to density variations but as a Lagrangian multiplier whose sole purpose is to impose the divergence-free condition on the velocity field;

3. The density waves are suppressed in the incompressible formulation and their energy is redistributed into the Reynolds stress, the viscous stress and the Lagrangian multiplier which collectively zero the Lighthill stress field;

4. If $c_0$ is substituted for $c$ on both sides of the original Lighthill equation, and the constitutive relation is enforced on the incompressible flow results, there could be a non-zero Lighthill stress component equivalent to the difference between the Reynolds and viscous stress on the right hand side and a density wave travelling at speed $c$ on the other side of the equation.

Here, an alternative view of the Lighthill aeroacoustic analogy is presented when applied to incompressible flows by treating the third term in the Lighthill stress essentially as the artificial compressibility. Moreover, if one assumes that there is no interaction between the resurgent acoustic waves and the rest of the incompressible fields, three possible models for the Lighthill...
stress tensor can be formulated, as listed below,
\[
T_{ij} \approx \begin{cases} 
\rho_0 u_i u_j \\
-P \delta_{ij} \\
\rho_0 u_i u_j - \sigma_{ij}.
\end{cases}
\tag{5.3.3}
\]

All three models are based on incompressible flow fields and should produce the same outcomes according to the conservation relations.

After the Laplacian operation, the Lighthill stress fields are transformed into the sound source fields. The first source model consisting of only the Reynolds stress may be mathematically calculated as
\[
\frac{\partial u_i u_j}{\partial x_i \partial x_j} = \frac{\partial^2 u_i}{\partial x_i \partial x_j} u_j + \frac{\partial u_i}{\partial x_j} \frac{\partial u_j}{\partial x_i} = 0
\tag{5.3.4}
\]
where \( \omega_i \) is the vorticity vector. It can be seen that the acoustic source term related to the Reynolds stress simplifies to the first-order second invariant of the velocity gradient tensor in incompressible flow, which may be interpreted as the imbalance between the local rate of strain and rate of rotation.

The second source model involves the Lagrangian multiplier and its second-order spatial derivative becomes
\[
\frac{\partial^2 P \delta_{ij}}{\partial x_i \partial x_j} = \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + \frac{\partial^2 P}{\partial z^2}.
\tag{5.3.5}
\]
The Poisson relation between the pressure and stress may be used for the incompressible flow, derived by taking the divergence of the Navier-Stokes equations,
\[
\frac{\partial^2 P}{\partial x_i \partial x^i} = \frac{\partial (u_i u_j)}{\partial x_i \partial x_j}.
\tag{5.3.6}
\]
Therefore, in theory, such a model will yield exactly the same source field as the previous Reynolds stress model and the time derivatives of the velocity has no net contribution to the sound source.

The third source model, which combines the Reynolds stress and viscous stress, has its viscous stress decomposed as
\[
\frac{\partial^3 u_i}{\partial x_i \partial x_j \partial x_j} = \frac{\partial^2 \nabla \cdot u}{\partial x_j^2} = 0.
\tag{5.3.7}
\]
Due to the divergence-free condition imposed on the velocity field, like temporal derivatives of velocity, viscous effects seem to have no net contribution to the acoustic source field, however
the variational formulation often requires separation of volume integrals in order to impose sensible boundary conditions. Their roles in the variational form of the formulation will be investigated on the numerical examples of spinning vortex pairs in Section 5.3.5 and 2D vortex shedding from a square cylinder in Section 5.4. The Reynolds stress model is denoted as Model 1, the Lagrangian multiplier model as Model 2 and the combined Reynolds and viscous stress model as Model 3.

5.3.3 Weak Form

In the new FE formulation, the acoustic variable is normalised by dividing reference density \( \tilde{\rho} = \rho / \rho_0 \) on both sides of the equation, so the Lighthill equation becomes

\[
(\tilde{\rho}|_{tt}, \phi) - e(\tilde{\rho}, \phi) = f(\tilde{T}, \phi) + g(\tilde{\rho}, \phi), \forall \tilde{\rho}, \phi \in H^1(\Omega),
\]

where the operators \( e, f \) and \( g \) can be decomposed into,

\[
e(\tilde{\rho}, \phi) = -c_0^2 \int_{\tilde{\Omega}} \frac{\partial \tilde{\rho}}{\partial x_i} \frac{\partial \phi}{\partial x_i} d\tilde{\Omega},
\]

\[
f(\tilde{T}, \phi) = \int_{\tilde{\Omega}} \frac{\partial (u_i u_j)}{\partial x_j} \frac{\partial \phi}{\partial x_i} d\tilde{\Omega} \quad \text{(Model I)},
\]

\[
= -\int_{\tilde{\Omega}} \frac{\partial P}{\partial x_i} \frac{\partial \phi}{\partial x_i} d\tilde{\Omega} \quad \text{(Model II)},
\]

\[
= \int_{\tilde{\Omega}} \left[ \frac{\partial (u_i u_j)}{\partial x_j} - \frac{\partial^2 u_i}{\partial x_j \partial x_j} \right] \frac{\partial \phi}{\partial x_i} d\tilde{\Omega} \quad \text{(Model III)},
\]

\[
g(\tilde{\rho}, \phi) = \int_{\tilde{\Gamma}} \left[ c_0^2 \frac{\partial \tilde{\rho}}{\partial x_i} + \frac{\partial (u_i u_j)}{\partial x_j} \right] n_i \phi d\tilde{\Gamma} \quad \text{(Model I)},
\]

\[
= \int_{\tilde{\Gamma}} \left[ c_0^2 \frac{\partial \tilde{\rho}}{\partial x_i} + \frac{\partial P}{\partial x_i} \right] n_i \phi d\tilde{\Gamma} \quad \text{(Model II)},
\]

\[
= \int_{\tilde{\Gamma}} \left[ c_0^2 \frac{\partial \tilde{\rho}}{\partial x_i} + \frac{\partial (u_i u_j)}{\partial x_j} + \frac{\partial^2 u_i}{\partial x_j \partial x_j} \right] n_i \phi d\tilde{\Gamma} \quad \text{(Model III)}.
\]

5.3.4 FE Discretisation of the Acoustic Fields

Spatial Discretisation

Basis functions \( \phi(\xi_i) \) are chosen for the acoustic field \( \tilde{\rho} \),

\[
\tilde{\rho}(\Omega_h) = \sum_{i=1}^{n} \phi_i \rho_i; \quad \phi(\Omega_h) = \sum_{i=1}^{n} \phi_i;
\]

For evaluating the source terms defined in Model 3, the second order spatial derivatives for the velocity field need to be obtained. According to [219], the derivatives of \( \partial^2 \xi_i / \partial x_j^2 \) can be
derived by expanding \( \delta x_i \) as a Taylor series expansion of \( \delta \xi_i \) and \( \delta \xi_i \) as a Taylor series expansion of \( \delta x_i \), followed by substituting the latter into the former, and the second order derivatives can be obtained through the relation

\[
\frac{\partial^2 u_i}{\partial x_j \partial x_j} = \frac{\partial^2 u_i}{\partial \xi_m \partial \xi_n} \frac{\partial \xi_m}{\partial x_j} \frac{\partial \xi_n}{\partial x_j} + \frac{\partial u_i}{\partial \xi_m} \frac{\partial^2 \xi_m}{\partial x_j \partial x_j},
\]

(5.3.18)

where \( \frac{\partial^2 \xi}{\partial x_j^2} \) can be found via the equation,

\[
\frac{\partial^2 \xi_i}{\partial x_j^2} = -\frac{1}{2} \frac{\partial^2 x_l}{\partial \xi_m \partial \xi_n} \frac{\partial \xi_m}{\partial x_j} \frac{\partial \xi_n}{\partial x_j} \frac{\partial \xi_i}{\partial x_l}.
\]

(5.3.19)

**Time Discretisation**

For time stepping, Houbolt’s method [220] has been selected, which has been shown in [221, 222] to produce convergent results for large time steps and is more flexible than the second order central difference scheme when a non-uniform grid step is used. It is an implicit fourth-order finite difference scheme which is unconditionally stable for all \( \delta t \) and can be written as

\[
\frac{\partial \tilde{\rho}^t}{\partial t} = \frac{1}{6\delta t} (11\tilde{\rho}^t - 18\tilde{\rho}^{t-\delta t} + 9\tilde{\rho}^{t-2\delta t} - 2\tilde{\rho}^{t-3\delta t}),
\]

(5.3.20)

and

\[
\frac{\partial^2 \tilde{\rho}^t}{\partial t^2} = \frac{1}{\delta t^2} (2\tilde{\rho}^t - 5\tilde{\rho}^{t-\delta t} + 4\tilde{\rho}^{t-2\delta t} - \tilde{\rho}^{t-3\delta t}).
\]

(5.3.21)

This method requires the calculation of two preceding time steps before the first time step \( t_0 \) and the values at \( (t_0 - \delta t) \) and \( (t_0 - 2\delta t) \). These may be calculated via the specified first and second time derivatives at \( t_0 \) as

\[
\tilde{\rho}^{t_0-\delta t} = \delta t^2 \frac{\partial^2 \tilde{\rho}}{\partial t^2} |_{t_0} + 2\tilde{\rho}^{t_0} - \tilde{\rho}^{t_0+\delta t},
\]

(5.3.22)

and

\[
\tilde{\rho}^{t_0-2\delta t} = 6\delta t^2 \frac{\partial^2 \tilde{\rho}}{\partial t^2} |_{t_0} + 6\delta t \frac{\partial \tilde{\rho}}{\partial t} |_{t_0} + 9\tilde{\rho}^{t_0} - 8\tilde{\rho}^{t_0+\delta t},
\]

(5.3.23)

where \( \tilde{\rho}^{t_0+\delta t} \) may be calculated at the end of the first time step. The entire discretisation form of the Equation (5.3.8) can now be numerically solved as

\[
\frac{2}{(\delta t)^2} (\tilde{\rho}, \phi) + c_\phi^2 \tilde{\rho}^{t_i} (\phi) = f(T), \phi) + g(\tilde{\rho}, \phi) + \frac{5}{(\delta t)^2} (\tilde{\rho}^{t-\delta t}, \phi)
\]

\[
- \frac{4}{(\delta t)^2} (\tilde{\rho}^{t-2\delta t}, \phi) + \frac{1}{(\delta t)^2} (\tilde{\rho}^{t-3\delta t}, \phi).
\]

(5.3.24)

**Boundary Conditions for the Acoustic Domain**

As discussed in Section 5.3.2, on shared boundaries (\( \Gamma_{\text{share}} \)) between the incompressible flow and the acoustic domains, the relation \( P' - c_0^2 \tilde{\rho} = 0 \) is imposed, which may be viewed as the
‘artificial compressibility’ re-introduced to the aeroacoustic domain. Therefore, these boundary conditions can be implemented as the Neumann type for the three models defined in Equations (5.3.10), (5.3.11) and (5.3.12) as

\[
g(\Gamma_{\text{share}}) = \int_{\Gamma_{\text{share}}} \left( \frac{\partial u_i}{\partial t} + \mu \frac{\partial^2 u_i}{\partial x_j \partial x_j} \right) n_i \phi d\Gamma_{\text{share}}; \quad (\text{Model I}) \quad (5.3.25) \\
= 0; \quad (\text{Model II}) \quad (5.3.26) \\
= \int_{\Gamma_{\text{share}}} \frac{\partial u_i}{\partial t} n_i \phi d\Gamma_{\text{share}}; \quad (\text{Model III}) \quad (5.3.27)
\]

so it can be seen that there is no need for any explicit condition to be implemented for the model using the Lagrangian multiplier while the unsteady flow condition needs to be explicitly accounted for in both Models 1 and 3 while the viscous stress also needs to be added in Model 1, if it is significant on the shared boundaries.

On the other hand, the outgoing waves need to be considered on the external boundary \(\Gamma_{\text{out}}\) for the acoustic domain only where all aerodynamic quantities are deemed to be negligible (i.e. a static medium). The basic implementation for absorbing conditions follows the practice adopted in [223], which is based on a first-order approximation of the solution to the wave equation [224],

\[
\int_{\Gamma_{\text{out}}} c_0^2 \frac{\partial \tilde{\rho}}{\partial x_i} n_i \phi d\Gamma_{\text{out}} = \int_{\Gamma_{\text{out}}} c_0 \frac{\partial \tilde{\rho}}{\partial t} \phi d\Gamma_{\text{out}}. \quad (5.3.28)
\]

Applying the Houbolt scheme for the discretisation of the time-dependant boundary integral, the above equation is modified to take the form

\[
\left( \frac{2}{(\delta t)^2} + c \frac{11}{6\delta t} \right) (\tilde{\rho}^t, \phi) + \epsilon(\tilde{\rho}^t, \phi) = f(\tilde{T}^t, \phi) + g(\tilde{\rho}^t, \phi) \\
+ \left( \frac{5}{(\delta t)^2} - \frac{3}{\delta t} \right) (\tilde{\rho}^{t-\delta t}, \phi) \\
- \left( \frac{4}{(\delta t)^2} + c \frac{3}{2\delta t} \right) (\tilde{\rho}^{t-2\delta t}, \phi) \\
+ \left( \frac{1}{(\delta t)^2} - c \frac{1}{3\delta t} \right) (\tilde{\rho}^{t-3\delta t}, \phi). \quad (5.3.29)
\]
5.3.5 Test Problem 2: Spinning Vortex Pair

The classical aeroacoustic problem of a spinning vortex pair has been employed in many numerical models for the purpose of benchmark testing [225, 226, 227, 177, 223] and is used to validate the acoustic code. In the models, the associated parameters for the problem are defined in Figure 5.7, where \( \Gamma \) is called the circulation intensity and the acoustic waves are propagated at the speed of \( c_0 \), which are normalised by the rotation radius \( r_0 \), and the rotation period

\[
T = 2\pi \omega = \frac{8\pi^2 r_0^2}{\Gamma}
\]

where \( \omega \) is the angular velocity.

(a)

(b)

Figure 5.7: (a) The schematic diagram and the problem parameters. (b) The outline of the numeric grid.

The computational domain, shown in Figure 5.7(b), consists of \( 650 \times 650 \) biquadratic elements arranged in a structured grid, however all the nodes are spread in a concentric form where there is a uniform element length of \( l_e = 0.008 \sim 0.015r_0 \) in the region within \( 1.5r_0 \) to the center and gradually stretched in the radial direction towards the periphery with an exponential expansion ratio of 1.01 whose value has been determined to avoid significant artificial reflections between neighbouring elements after some preliminary trials. The overall mesh consists of 1,692,601 nodes.

The flow field is constructed as an incompressible inviscid fluid which can be calculated by a complex velocity potential function \( \Phi(z, t) \) [227, 177, 223],

\[
\Phi(z, t) = \frac{\Gamma}{2\pi i} \ln(z - b) + \frac{\Gamma}{2\pi i} \ln(z + b),
\]

where \( z = x + iy = re^{i\theta} \) and \( b = r_0e^{i\omega t} \). The velocity field can be reconstructed by differentiating the potential function with respect to \( z \)

\[
u - iv = \frac{\partial \Phi(z, t)}{\partial z} = \frac{\Gamma}{i\pi} \frac{z}{z^2 - b^2},
\]

where \( u \) and \( v \) are the real and imaginary parts of the velocity, respectively.
and from the unsteady inviscid Bernoulli’s equation, the Lagrangian multiplier can also be reconstructed via the relation

\[ P = P_0 - \rho_0 \frac{\partial \Phi}{\partial t} - \rho_0 \frac{u^2 + v^2}{2}, \]  

(5.3.32)

where \( P_0 \) is assumed to be a static field.

There is one obvious shortcoming in this expression as there exists an infinite value at the center of each vortex core with \( r_{\text{vortex}} = 0 \), caused by \( |z| = |b| \), which means the function is no longer square integrable in \( \Omega \). Therefore in the vortex center region \( (\Omega_1) \), the source field is nullified by forcing any aerodynamic quantities to zero for \( r_{\text{vortex}} \leq \lambda r_0, \lambda = 0.005 \sim 0.05 \). Overall the new problem can be formulated as

\[
f(\Omega) = \int_{r_{\text{vortex}}=0}^{r_{\text{vortex}}=\lambda r_0} f(0) d\Omega_1 + \int_{r_{\text{vortex}}=\lambda r_0}^{r_{\text{vortex}}=\infty} f(\Phi) d\Omega_2, \tag{5.3.33}
\]

where \( f(0) \) is a constant field and has the value zero in space \( \Omega_1 \), while the one involved with the velocity potential function (\( \Phi \)) governs the flow elsewhere in the domain \( \Omega_2 \). The two domains share the boundary \( \Gamma_{r_{\text{vortex}}} \).

As the flow is inviscid, only source Models 1 (Equation (5.3.10)) and 2 (Equation (5.3.11)) are employed in this problem. Towards the edges of the domain, \( u \to 0 \) and \( P \to 0 \) therefore the boundary integral \( \Gamma_{\text{far}} \) from the two source fields is considered negligible. The non-reflecting boundary condition, introduced in Section 5.3.4, is implemented in order to prevent significant reflection of the outgoing waves for both models. However, there are different approximations to the boundary integral created at truncated boundaries around each vortex centre. On \( \Gamma_{r_{\text{vortex}}} \), according to the relations defined in Equations (5.3.25) and (5.3.26), boundary integrals can be calculated as

\[
g(\Gamma_{r_{\text{vortex}}}) = \int_{\lambda r_0} \rho_0^2 \left( \frac{\partial \rho'}{\partial x_i} n_i \phi - u_j \frac{\partial u_i}{\partial x_j} n_j \phi d\Gamma_{r_0} \right) \tag{5.3.34}
\]

Model 1;

\[
g(\Gamma_{r_{\text{vortex}}}) = \int_{\lambda r_0} \rho_0^2 \left( \frac{\partial \rho'}{\partial x_i} n_i \phi - \frac{\partial P}{\partial x_i} n_i \phi d\Gamma_{r_{\text{vortex}}} \right) \tag{5.3.35}
\]

Model 2,

where the difference between Models 1 and 2 is the lack of \( \int_{\lambda r_0} (\partial u_i/\partial t) n_i \phi d\Gamma_{r_{\text{vortex}}} \) in the former.

**Approximating the Spatial Gradients**

Equation (5.3.10) requires the calculation of the first-order spatial gradients of the velocity field for the assembly of the Lighthill source. This can be done based on interpolations of the specified nodal velocity, similar to the approach taken in [223] where the velocity vectors are directly calculated from the velocity potential function \( \Phi \) at each node. In this way, the derived gradient fields are linear functions within each element which uses a quadratic basis for the velocity field. In Figure 5.8, the strength of the individual components of the velocity gradients, located at
$r = 1.02$ and $\theta = \pi/3$, are calculated by the nodal interpolation scheme and by the analytical function respectively. The result shows that although the nodal interpolation scheme predicts the correct shape of the $\partial u/\partial y$ in Figure 5.8(b) and $\partial v/\partial x$ in Figure 5.8(c), it fails to agree with the analytical results in both the shapes and the sizes of the $\partial u/\partial x$ and $\partial v/\partial y$, due to the insufficient grid resolution. This misrepresentation of the velocity gradients by the nodal interpolation method can cause a severe distortion in the predicted acoustic waves, as shown in Figure 5.12.

In Method 2, the velocity gradients are calculated analytically from the velocity potential function (5.3.30) at each Gauss point,

$$\frac{\partial u}{\partial x} - i \frac{\partial v}{\partial x} = \frac{-\Gamma}{\pi i \partial (z^2 + b^2)}; \quad (5.3.36)$$

$$\frac{\partial u}{\partial y} - i \frac{\partial v}{\partial y} = \frac{\Gamma}{\pi \partial (z^2 - b^2)}; \quad (5.3.37)$$

The resolution of the acoustic source integral has been decoupled from the grid which is only used for solving the wave propagation. The accuracy of the approximation of the source field is now only dependent on the order of the Gaussian quadrature used for the numerical integration. In Figure 5.9, a convergence analysis is conducted for the relationship between the integral of the first-order spatial gradients at $r = 1.02$ and the number of Gauss points. The results show that a quadrature of 15 Gauss points per dimension (i.e. 225 points per 2D element) can produce a reasonable accuracy ($\delta < 0.01$) for the first-order spatial derivatives of the velocity field.

Furthermore, a convergence analysis is done to study the grid resolution of the acoustic propagation for both Models 1 and 2. From Figures 5.10(a) and (b), it can be seen that as the grid is refined, the far-field signal produced by both models becomes more symmetrical to the ambient pressure, while Model 2 (e.g. the pressure source) always predicts a larger amplitude than Model 1 (e.g. velocity-based source) at all grid resolutions. There is also a small shift in the phase of the acoustic signals in relation to the grid resolution.
Figure 5.8: The calculated velocity gradients at $r = 1.02$ and $\theta = \pi/3$ by the nodal interpolation method. (a) $\partial u/\partial x$; (b) $\partial u/\partial y$; (c) $\partial v/\partial x$; (d) $\partial v/\partial y$. 
Figure 5.9: The calculated strength of the four velocity gradients combined ($\int_0^{2\pi} (\|\partial u/\partial x\| + \|\partial u/\partial y\| + \|\partial v/\partial x\| + \|\partial v/\partial y\| d\theta)$ at $r = 1.02$ by the Gauss-point-based interpolation method.

Figure 5.10: The acoustic signals measured at $r = 50r_0$ and $\theta = 0$. (a) Model 1 with a truncated vortex surface at $r_{vortex} = 0.05r_0$ and a time step, $\delta t = 0.01T$. (b) Model 2 with a truncated vortex surface at $r_{vortex} = 0.05r_0$ and a time step, $\delta t = 0.01T$. 
Approximation of Temporal Gradients

The form of Lighthill stress tensor in Equation (5.3.10) only contains spatial derivatives of the velocity field, however a close look into the fluid dynamics via the inviscid Naiver-Stokes equations suggests that such an acoustic source form carries an implicit weight of the first-order temporal derivative of the velocity, as

$$ u_j \frac{\partial u_i}{\partial x_j} = - \frac{\partial p}{\partial x_i} - \frac{\partial u_i}{\partial t}. $$  

(5.3.38)

Figure 5.11: The calculated velocity gradients at $r = 1.02$ and $\theta = \pi/3$ by the Houbolt method with $\delta t = 0.01T$. (a) $\partial u/\partial t$; (b) $\partial v/\partial t$. 
As a result, an analysis has been drawn to investigate the temporal resolution of the velocity field calculated by the numerical scheme (i.e. the Houbolt method), as shown in Figure 5.11 where the calculated values are compared to the ones derived from the analytical functions,

$$\frac{\partial u}{\partial t} - i\frac{\partial v}{\partial t} = \frac{2\omega\Gamma}{\pi} \frac{\partial (zb^2)}{\partial (z^2 - b^2)^2}. \quad (5.3.39)$$

It can be concluded based on the results, that the current numerical scheme used for wave propagation fails to accurately approximate the sharp temporal gradients for the velocity field in the vortice cores. It would require much finer temporal resolution to produce a close approximation to the true temporal gradients and reach an accurate velocity-based source field. On the other hand, the temporal gradient of the velocity involved in the Lagrangian multiplier model (Model 2) can be calculated in an explicit form,

$$\frac{\partial p}{\partial x_i} = -\rho_0 \left( \frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} \right). \quad (5.3.40)$$

From the velocity potential function $\Phi$, the temporal gradient can be calculated as

$$\frac{\partial u}{\partial t} - i\frac{\partial v}{\partial t} = \frac{2\omega\Gamma}{\pi} \frac{zb^2}{(z^2 - b^2)^2}. \quad (5.3.41)$$

Therefore in Model 2, the temporal gradients involved in the source field can be exactly accounted for at each Gauss point. In Figure 5.12, the contribution of the two temporal derivatives ($\int_0^{2\pi} (\|\partial u/\partial t\| + \|\partial v/\partial t\|) d\theta$) can be compared in a line integral calculated at $r = 1.02$ which is close to the vortex center, against the number of Gauss points in the numerical integration. The results show good convergence behavior with the number of Gauss points and a Gaussian quadrature with 15 points per dimension can produce an error less than 1% compared to the scheme which uses 20 Gauss points per dimension.

![Figure 5.12: The calculated temporal gradients of the velocity at $r = 1.02$ in Model 2. The combined temporal gradients field ($\|\partial u/\partial t\| + \|\partial v/\partial t\|$) is normalised by $\Gamma^2/r_0$.](image)

Lastly, the temporal sampling rate of the source field also has a large impact on the calcu-
lated acoustic far-field. As shown in Figure 5.13, the two temporal velocity gradients calculated at \( r = 1.02 \) are compared against the number of uniform time steps per half \( T \). In Figure 5.14, as the number of time steps increases from 10 to 50 points per half period, the calculated acoustic field increases in amplitude as well as shifts in phase for both models. Nearly infinite-valued temporal gradients (i.e. both \( \partial u_i / \partial x \) and \( \partial u_i / \partial t \to \infty \) at the \( r_{\text{vortex}} = 0 \)) are approximated by a finite number of temporal sampling points.

![Graph](image)

Figure 5.13: The sampling rate vs the temporal gradients of the velocity at \( r = 1.02 \) and \( \theta = \pi/3 \) in Model 2. (a) \( \partial u / \partial t \); (b) \( \partial v / \partial t \).

Effect of Truncated Vortices

In the two previous sections, it is observed that the two source models predict different acoustic far-field results under the same spatial and temporal resolutions. As derived earlier, the only difference between the two models lies in the treatment of the artificially truncated boundaries close to the vortex centers. Here, an analysis is drawn to study the effect of the placement of the artificially truncated surface.
Figure 5.14: The acoustic field calculated at $r = 50$ and $\theta = 0$ vs the number of time steps per half $T$. (a) The acoustic field far-field calculated by Model 1. (b) The acoustic field calculated by Model 2.
In Figure 5.15, the results show that the strength of acoustic far-field signals increase as more source domain is used in the calculation of the source field for both models. Among all the truncated surfaces tested, Model 2 predicts larger amplitude far-field than Model 1 does.

Results

The acoustic simulation was run at a constant time resolution of 100 steps per period and lasted for 20T. The radiation fields calculated by the three numerical models, at t = 15T, on grid \( l_e = 0.01r_0; r_{vortex} = 0.005r_0 \) and \( \delta t = 0.01T \), are presented in Figure 5.16 where the interleaved spiral pattern of the simulated acoustic field is predicted by all models.

From the figure, irregularity can be found on the wave patterns predicted by the source model based on the Reynolds stress for the nodal-based model, which can be explained by the poor approximation of the spatial gradients in the source region as shown in the previous discussion. The diminishing velocity field in the vortex core region generates large, static

Figure 5.15: The acoustic field calculated at \( r = 50 \) and \( \theta = 0 \) on a grid with a minimal step \( \delta = 0.01 \) and a temporal sampling rate at \( \delta t = 0.01T \). (a) The acoustic far-field calculated by Model 1. (b) The acoustic far-field calculated by Model 2.
and negative acoustic source, therefore creates a negative acoustic field in the center of each vortices for all three models. The simulated acoustic far-field shown in Figure 5.16 produces no noticeable sign of reflections from the edge of the domain (at \( x \approx 147r_0 \)), which shows the effectiveness of the absorbing boundary implementation.

Along the centerline of the domain at \( \theta = 0 \) and \( r = 50 \), the normalised acoustic field \( p'/\rho_0c_0^2 \)
is compared with the values predicted by Howe’s model [228] in Figure 5.17

$$p \approx -\rho_0 \Gamma 8\pi^2 r_0^2 \left( \frac{T c_0}{2r} \right)^{0.5} \cos[2\theta - \frac{4\pi}{T}(t - \frac{r}{c_0}) + \frac{\pi}{4}]. \quad (5.3.42)$$

Here, the far-field density-pressure relation is assumed to follow the Bernoulli equation $p' = \rho' c_0^2$.

Among the results produced by the three models, the simulated signal for the Lagrangian multiplier has the closest match to the analytical solutions, which is due to the better approximation of both the spatial and temporal gradients in the source field than the other two models, as discussed in previous sections.

![Figure 5.17](image)

Figure 5.17: The acoustic signal measured at $r = 50r_0$ in a truncated source domain at $r_{vortex} = 0.005r_0$ on a grid with minimal $\delta = 0.01r_0$ and a sampling rate of 0.01T. Model 1a is the nodal-based Reynolds stress model, Model 1b is the integral-based Reynolds stress model and Model 2 is the integral-based Lagrangian multiplier model.

From the results shown in Figure 5.17, the nodal-based Model 1 produces the worst results due to its poor ability of approximating the source field derived from sharp velocity gradients close to each vortex centres. In addition, Gaussian-based Model 2 produces far closer solutions to the analytical derivation than Gaussian-based Model 1 does. Compared to the solution produced by the former, the latter one underestimates the acoustic strength at $r = 50r_0$ by approximately half.

As derived in Equations (5.3.35), the only difference between Gaussian-based Models 1 and 2 is the additional temporal velocity gradients presented in the truncated boundary integrals near each vortex center, therefore it may be concluded that the variational form of the Lighthill equations is sensitive to the imposed boundary conditions, and $\partial(P-c_0^2\rho')/\partial x_i n_i = 0$ adopted by Model 2 is a better approximation than the $(c_0^2\partial\rho'/\partial x_i + u_j\partial u_i/\partial x_j)n_i = 0$ used by Model 1 for evaluating sound sources on surfaces in the vortex region. According to the results illustrated in Figure 5.15, both source models are expected to produce the same acoustic far-field as the truncated surface gets closer to the vortex center.
5.4 Test Problem 3: Aeroacoustic Simulation of 2D Flow Around a Square Cylinder

Analogy-based methods have been widely adopted for studying vortex generated sound such as the examples of vortex shedding from a solid cylinder. The Lighthill analogy [223, 194] is implemented for the investigation of flow-induced noise in flows over a 3D cylinder, where LES was employed for the CFD solution. In this work, the Lighthill stress tensor only included the Reynolds stress calculated from the solutions of the incompressible Navier-Stokes equations.

In other work, Curle’s analogy and the Ffowcs Williams-Hawkings analogy were modified to build the source field [229, 230, 231, 232]. While sharing the same acoustic far-field expression with Lighthill’s formulation, Curle’s analogy singles out the acoustic effect of the solid boundary by explicitly separating a boundary integral for the force acting on the fixed surface, known as the acoustic dipole, and a volumetric source made from the Lighthill stress, also known as the acoustic quadrupole. It can be shown that the dipole strength is much larger than the quadrupole for low Mach number flows [229], therefore the quadrupole may be completely removed from the acoustic modelling of this flow [230].

In [231], the Doppler effect was introduced to the solution of Curle’s equation based on the results of an earlier study conducted in [233]. In [234], the solutions from Curle’s analogy were compared with the results obtained by Powell’s analogy [190], where the acoustic source fields were represented by the Lamb vector on a 2D vortex cylinder in a confined space. The study shows that both analogies rely heavily on cancellation of sources and are prone to numerical truncation of the source domain. The Ffowcs-William Hawks analogy furthers the concepts of Curle’s analogy by making the surface integral ‘permeable’ meaning that the surface can be introduced anywhere in the source domain allowing the exchange of the mass, momentum and energy. Lockard [232] experimented with a few different integration surfaces sampled at various distances away from a 2D cylinder and reported that the vortex passing through the integral surface leads to noise production.

In the following sections, a complete two-stage simulation of aeroacoustics is applied for solving the 2D vortex shedding from a square cylinder problem, using the customised code. Firstly, the incompressible Navier-Stokes equations are numerically resolved via the DNS method, followed by the assembly of the acoustic source field. Three different source models for the Lighthill analogy, based solely on solutions of the incompressible flow as described in Section 5.3.2, are compared in the modified Lighthill analogy in variational formulations (Equation (5.3.8)).
5.4.1 Incompressible Flow Field

The rectangular CFD domain is 31 units long in the x direction and 21 units wide in the y direction with a single unit square placed 10 units from the inlet in the x direction, as shown in Figure 5.18 where the numerical parameters are also defined for the problem. All parameters are non-dimensionalised with respect to one or more of the following parameters, the inlet velocity \( u_\infty \), the diameter of the cylinder \( D \), the reference density \( \rho_0 \) and the speed of sound in a static medium \( c_0 \). Grids of three resolutions ranging from 240,000 to 1.5 million nodes, were tested for the problem and the elemental spaces were all constructed using the Q-L basis (Quadratic Lagrangian function for the spatial and velocity fields, Linear Lagrangian function for the Lagrangian multiplier field).

Figure 5.18: A schematic of the CFD domain for the vortex shedding.
The construction of the grid places the minimum cell size near the cylinder wall while the entire grid is gradually stretched in both the x and y directions away from the center surface according to the power law $\delta r_{n+1} = \delta r_n^\lambda$, $\lambda = 1.02$, as shown in Figure 5.19 for the lowest resolution grid (Grid 1).

Figure 5.19: Magnified Grid 1 used for the CFD simulation around the square cylinder. The flow is from left to right.
The second grid (Grid 2) was constructed by refining the base grid in two directions. The final grid (Grid 3) further refines the elements in the second grid only in the x direction. The flow was gradually increased from $Re = 0$ to $Re = 150$ in order to develop an instability and the flow was sustained with a uniform time step until the amplitude of the lift coefficient changed by less than 0.1%. The fully developed lift and drag coefficients calculated on the cylinder surface are plotted in Figures 5.20(a) and (b) for one shedding period.

![Figure 5.20: The time history of lift and drag coefficients measured on the square cylinder for an estimated shedding cycle, Grid 1 (Q-L basis 241,200 nodes in blue); Grid 2 (Q-L basis 962,400 nodes in green); Grid 3 (Q-L basis 1,443,000 nodes in red). (a) Lift coefficients. (b) Drag coefficients.](image)

The Strouhal number, maximum amplitude of the lift coefficient and the mean values of the drag coefficient, together with some model parameters, are compared with published data from [235], where an incompressible formulation was adopted and the results obtained from a compressible model in [233] in Table 5.1. The results produced by the three grids are in the range of those published for the three listed aerodynamic quantities. In Figures 5.21(a)-(c), the velocity gradients related to the acoustic source field are also calculated in the form of $\int_{\Gamma} (\partial u_i / \partial x_j)^2 d\Gamma$, normalised by $u_{\infty}^2 / D$ and integrated on three faces of the cylinder (front, back and top) for all three grids.
Table 5.1: Comparison between Different Flow Models

<table>
<thead>
<tr>
<th></th>
<th>Strouhal number</th>
<th>$C'_{L}$</th>
<th>$C_D$</th>
<th>Min Wall Spacing</th>
<th>$\Delta t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[235]</td>
<td>0.165</td>
<td>1.56</td>
<td>0.38</td>
<td>0.0038</td>
<td>0.025</td>
</tr>
<tr>
<td>[236]</td>
<td>0.151</td>
<td>1.40</td>
<td>0.40</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>Grid 1</td>
<td>0.162</td>
<td>1.57</td>
<td>0.37</td>
<td>0.039</td>
<td>0.017</td>
</tr>
<tr>
<td>Grid 2</td>
<td>0.162</td>
<td>1.58</td>
<td>0.37</td>
<td>0.019</td>
<td>0.017</td>
</tr>
<tr>
<td>Grid 3</td>
<td>0.162</td>
<td>1.57</td>
<td>0.37</td>
<td>0.013</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Figure 5.21: The time history of the fully developed velocity gradients fields measured on the surfaces of the square cylinder over one shedding period, Grid 1 (Q-L basis 241,200 nodes in blue); Grid 2 (Q-L basis 962,400 nodes in green); Grid 3 (Q-L basis 1,443,000 nodes in red). (a) On the front surface; (b) On the back surface; (c) On the top surface.
5.4.2 Acoustic Near-Field

Some analysis of the individual source fields used in the new FE formulation are first performed before the simulation of acoustic waves. The volumetric source fields for the three source models can be mathematically calculated as \( S_i = \frac{\partial (u_i u_j)}{\partial x_j} + \nu \frac{\partial^2 u_i}{\partial x_j \partial x_j} \) and \( \frac{\partial P}{\partial x_i} \), as defined in \( f(\tilde{T}, \phi) \) and \( g(\tilde{\rho}, \phi) \). Figures 5.22, 5.23 and 5.24 present a look at the general distribution of the acoustic source terms at \( t = 0 \), where all terms are normalised by \( u_\infty^2 / D \) and the maximum and minimum values are fixed between -0.01 and 0.01 for display purposes.

![Visualisation of the acoustic source components based on Model 1. (a) \( u \partial u / \partial x \); (b) \( v \partial u / \partial y \); (c) \( u \partial v / \partial x \); (d) \( v \partial v / \partial y \).](image)

It can be seen that acoustic source terms are concentrated around the cylinder surfaces and in the regions of vortex generation. The distribution of the viscous stress is even more concentrated near the solid surfaces and less spread in the open regions compared to the other two types of acoustic source. The figures also show that there are large regions of alternating source strength behind the cylinder, corresponding to the vortex train, particularly for Models 1 and 2. To obtain more quantitative comparisons among the different acoustic loading terms, they are integrated on unit length planes parallel to the cylinder surfaces. The integrals normalised by \( u_\infty^2 \), are shown in Figures 5.25-5.27.
The results show that for all surfaces on the cylinder, the viscous stress dominates the acoustic production and is coupled with the pressure fluctuations in this region, similar to
Figure 5.25: The individual contributions to the acoustic source by components of the Reynolds stress (blue), Lagrangian multiplier (red) and viscous stress (green), measured at planes of 1 unit length parallel to the front face at various distances. (a) Acoustic loads in the x direction at x = 0.5; (b) Acoustic loads in the y direction at x = 0.5; (c) Acoustic loads in the x direction at x = 0.6; (d) Acoustic loads in the y direction at x = 0.6; (e) Acoustic loads in the x direction at x = 1.0; (f) Acoustic loads in the y direction at x = 1.0.

the findings in [209, 237, 238], while the Reynolds stress assumes the primary role for sound production within 0.1 unit length from the surface. For planes parallel to the front and back surfaces, all stress terms calculated in the x direction oscillate at the drag frequency which is twice the frequency of oscillating force measured in the y direction (i.e. the lift frequency). The amplitude of the acoustic source fields are higher on the front face than on the back. However, the source fields in front of the cylinder also decay more rapidly than those behind the cylinder, whereas the sound source increases due to the formation of the shed vortices. On the top face of the cylinder all of the acoustic source terms oscillate at the lift frequency and those along the y direction are much stronger than those in the x direction. Another observation is that although the Reynolds and viscous stress in the same direction seem to be always oscillating at the same frequency, the two source waves are not always in phase.

Next, the source distributions above the top surface are examined in Figure 5.28, where the acoustic source terms, normalised by $u_{\infty}^2$, are evaluated on the planes orthogonal to the top face at three locations. From the figure, it is evident that there are stronger acoustic sources
Figure 5.26: The individual contributions to the acoustic source by the components of the Reynolds stress (blue), Lagrangian multiplier (red) and viscous stress (green), measured at planes of 1 unit length parallel to the back face at various distances. (a) Acoustic loads in the x direction at x = -0.5; (b) Acoustic loads in the y direction at x = -0.5; (c) Acoustic loads in the x direction at x = -0.6; (d) Acoustic loads in the y direction at x = -0.6; (e) Acoustic loads in the x direction at x = -1.0; (f) Acoustic loads in the y direction at x = -1.0.

located towards the leading face of the cylinder than on the back face and that there is a phase difference between the source waves at different locations along the top surface.

As can be seen in these results, the Reynolds stress becomes the dominant sound source everywhere in the domain except for the area near the no-slip boundary on the cylinder surface.
Figure 5.27: The individual contributions to the acoustic source by components of the Reynolds stress (blue), Lagrangian multiplier (red) and viscous stress (green), measured at planes of 1 unit length parallel to the top face at various distances. (a) Acoustic loads in the x direction at $y = 0.5$; (b) Acoustic loads in the y direction at $y = 0.5$; (c) Acoustic loads in the x direction at $y = 0.6$; (d) Acoustic loads in the y direction at $y = 0.6$; (e) Acoustic loads in the x direction at $y = 1.0$; (f) Acoustic loads in the y direction at $y = 1.0$.

Figure 5.28: The Reynolds and viscous stress along the planes vertical to the top surface of the cylinder, where the source fields are evaluated at $x = -0.5$ (blue), $x = 0$ (red) and $x = 0.5$ (green). (a) $S_x = \int_{y=0.5}^{y=10.5} (u\partial u/\partial x + v\partial u/\partial y - \nu(\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2))dL$; (b) $S_y = \int_{y=0.5}^{y=10.5} (u\partial v/\partial x + v\partial v/\partial y - \nu(\partial^2 v/\partial x^2 + \partial^2 v/\partial y^2))dL$. 

130
5.4.3 Acoustic Far-Field

The 2D acoustic domain spans approximately 200 by 200 unit lengths and is an enlargement of the flow domain now embedded in the center, as shown in Figure 5.29. Only the CFD results based on the highest resolution grid (Grid 3) were used in the acoustic simulation, as it has the best accuracy for the acoustic source. As the flow is periodic, the data acquired from one fully developed period of the CFD simulation was recycled in the acoustic simulation which was sustained for approximately 28 periods. New biquadratic elements were padded at the boundary of the original CFD grid. The element sizes were gradually stretched in both directions according to the same power law used in the construction of CFD grids at the expansion ratio $\lambda = 1.005$. As a result, the grid used in the acoustic simulation consists of 4.8M nodes and 1.2M elements. The Galerkin FE method and the variational formulation were applied for solving different forms of the Lighthill equation in a customised solver (AcousticSolve).

![Figure 5.29: The acoustic domain including two types of numerical boundary conditions: the shared boundary (red and green lines) $\Gamma_{flow} \cap \Gamma_{acoustic}$ and the non-reflective acoustic outlet (black line) $\Gamma_{acoustic} \not\subset \Gamma_{flow}$.](image)

The acoustic loads were evaluated at each node in the flow domain and the average nodal values within a shedding period were subtracted. There are no acoustic loads at newly created nodes (i.e. $T_{ij} = 0$ and $U_i = 0$), which represent a static medium.

Here, biquadratic shape functions $\phi(\xi, \eta)$ were chosen for the acoustic field, which is of the same order of the velocity field. There are now two sets of boundary conditions in the acoustic domain as specified in Figure 5.29. On the shared borders between the flow and acoustic domains ($\Gamma_{share}$), the relation $P' - c^2 \rho = 0$ is imposed. Since all the boundary conditions of the flow domain are time-invariant, the presence of the temporal gradient, defined in the boundary integrals in Model 3, is expected to carry no weight in the acoustic simulation. However, the
role of the viscous stress needs to be investigated, especially considering the regions around non-slip walls, to the resulting Model 1. On the external boundary $\Gamma_{\text{ext}}$, non-reflective conditions are imposed following the approach used in [223]. In addition, in order to avoid the artificial source due to truncation of the wake, the source field related to the vortex train was filtered by a Hanning window $\tilde{T}_{ij} = T_{ij} \times \frac{1}{2}(1 + \cos(\pi \times (x - l_{\text{damp}})/l_{\text{damp}}))$, proposed in [234]. The filtering window spans a distance of 19 units and gradually diminishes the source field towards the outlet boundary of the original CFD domain, starting at $l_{\text{damp}} = 1$ downstream to the cylinder.

The FE formulation is solved in the acoustic domain defined in Figure 5.29. The acoustic far field at $t = 174$ is calculated using all three source models as shown in Figures 5.30. All models predict the correct acoustic radiation pattern which is more or less symmetric about the horizontal axis, while the acoustic field produced by Model 1 is significantly weaker than the other two models.

![Figure 5.30: The simulated acoustic radiation fields for three Lighthill source models. The maximum and minimum values shown are fixed between 0.01 and -0.01 for display purposes. (a) The Reynolds stress model; (b) The Lagrangian multiplier model; (c) The Reynolds and viscous stress model.](image)

In Figure 5.31, the radiation fields at $r = 75$ and $r = 40$ are plotted for the models along with the DNS results given in [233]. All three source models predict a radiation field which is slightly shifted towards the upstream of the cylinder and the maximum strength of the acoustic pressure occurs at approximately $\theta = 80^\circ$. Overall, for the radiation field calculated at $r = 75$, Model 1 underestimates the maximum acoustic strength by approximately 20%. Model 2 shares almost the same radiation pattern with the third model and both produced better agreement with the DNS results than Model 1, though they still underestimate the acoustic strength ($\sim 15\%$) in the upstream sector and overshoot in the downstream sector by a smaller amount ($\sim 10\%$). This may be due to numerical errors associated with wave propagation as the radiation pattern at $r = 75$ is quite different from the pattern produced by the Lagrangian
multiplier model, calculated at \( r = 40 \), which more closely resembles the shape predicted by DNS.

### 5.4.4 Effect of Source Truncation associated with the Vortex Train

The source mechanism associated with the alternating vortex tail is further analysed by integrating the normalised Reynolds stress-related source on planes parallel to the back face of the cylinder at 4 locations, as shown in Figure 5.32. In both figures, it can be seen that the convected vorticity waves, forming at approximately 0.5 units behind the cylinder, gradually dissipate while traveling towards the exit boundary of the CFD domain. Even though the mesh is gradually stretched towards the outlet, inducing numerical dissipation in addition to the viscous dissipation, the vorticity wave is not diminished before it reaches the end of the domain. The oscillating frequency of the source wave was largely unchanged during convection, but the phase is shifted by the convection speed \( u_\infty \).

The source wave along the x direction sampled at 5 units away from the back face, is nearly 180 degrees out of phase with the source field measured at 15 units. As a result, there exists a delicate balance within the vorticity waves where the sound radiated by the preceding vortex is largely cancelled by the sound wave generated by the following vortices. This makes the entire
Figure 5.32: The normalised acoustic source due to the Reynolds stress sampled on the planes parallel to the back face of the cylinder at the distance of $x = -5.5$ (blue), $x = -10.5$ (green), $x = -15.5$ (in red) and $x = -20.5$ (light blue). (a) The acoustic loads in the x direction. (b) The acoustic loads in the y direction.

vortex train an inefficient sound source compared to the solid cylinder.

Figure 5.33: The acoustic field calculated by Model 3 without filtering the source associated with the vortex train.

This balance is artificially disrupted at the outlet boundary ($\partial u_i / \partial n_i = 0$), while it is observed that there is still a significant amount of acoustic source measured on the outlet boundary ($x = -20.5$). If left untreated, this can lead to the high amplitude of artificial sound demonstrated in [234] and shown in Figure 5.33 where the acoustic field is largely enhanced by the source located close to the original exit boundary of the CFD domain.

As a sidenote, a downstream obstacle may also have similar effect on the Lighthill source field caused by the alternating vortices in a similar way to the reflective numerical outlet.
boundary. This may explain some of the large sound production by so-called BVI (body vortex interaction) in a tandem configuration, as discussed in [236].

The effect of the length of the truncation window has been investigated, as shown in Figure 5.34. For a truncation window starting at 5 units behind the cylinder, there is only a very small difference in the acoustic field in the front sector compared to a window starting at 1 unit. There is larger increase in magnitude for the acoustic signals on the rear side of the cylinder, however the maximum change does not exceed 6%. For a truncation window starting at 10 units behind the cylinder, there is evidence to support the presence of artificial reflection from the rear sector, possibly due to more rapid damping of the source field compared to the other two models.

Figure 5.34: The acoustic field calculated at \( r = 50 \), using Model 3 with a truncation window starting at 1, 5 and 10 units behind the cylinder.
5.4.5 Mach Number Effect

Following the artificial compressibility assumption, there may be an improvement to the approximation of the constitutive law by introducing a Mach number relation based on the results of incompressible flows. The speed of sound for a perfect gas in isentropic flow may be written as a function of the Mach number

\[ c = \frac{c_0}{\sqrt{T_0}} = \frac{c_0}{\sqrt{1 + \frac{1}{2}(\kappa - 1)M^2}}, \]  

(5.4.1)

where \( \kappa \) is the adiabatic heat capacity ratio. By rewriting the Mach number as a decomposition of the incompressible and compressible parts, as \( M = \bar{M} + M' = \bar{U}/c + U'/c \), the above relation can be rearranged into,

\[ c^2 = c_0^2 - \kappa(\bar{U}^2 + 2\bar{U}U' + U'^2). \]  

(5.4.2)

It can be seen that as long as \( \bar{U} > 2U' \), the incompressible part of the deviation of the speed of sound from the reference state is larger than the changes due to compressible effects.

If \( \kappa = 1.4 \) is selected for an ideal gas such as air, a modified wave operator (\( \hat{\epsilon} \)) on the left hand side of the Lighthill equation can be formulated as

\[ \hat{\epsilon}(\tilde{\rho}, \tilde{\phi}) = -(c_0^2 - \frac{\kappa - 1}{2}\bar{u}^2) \int_{\tilde{\Omega}} \frac{\partial \tilde{\rho}}{\partial x_i} \frac{\partial \tilde{\phi}}{\partial x_i} d\tilde{\Omega}. \]  

(5.4.3)

The results using the modified wave operator in Model 3 for five angles at \( r = 50 \), are plotted in Figure 5.35. It can be seen that there is very little impact on the amplitude of the acoustic field compared to the model with a homogeneous wave operator. There is some small gain in the acoustic strength in the y direction (\( \theta = 90' \)) and some minor loss at all other angles. The other change is a small phase shift for the signals measured at all angles.
Figure 5.35: The Mach number effect, where the blue line is the acoustic signals calculated using the homogeneous wave operator while the green dashed line represents the signals produced with the $c$ calculated based on the incompressible flow solutions. The normalised acoustic field is measured at (a) $r = 0^\circ$; (b) $\theta = \pi/3$; (c) $\theta = \pi/2$; (d) $\theta = 2\pi/3$; (e) $\theta = 5\pi/6$. 
5.5 Discussion and Conclusions

In this chapter, a numerical model for computational aeroacoustics has been presented, based on incompressible flows, the finite element method and a modified Lighthill equation whose left hand side has been modified to include the effect of Mach number while various mathematical forms have been proposed and tested for modelling the acoustic source fields. Like the original form of the Lighthill equation which is derived from the fundamental principles of fluid dynamics without approximations of any kind [187], the modified Lighthill analogy should produce the same acoustic field as the ones calculated by DNS methods for isentropic flows, when provided a good approximation of $c$.

The role of the viscous stress as an acoustic source has been investigated in Section 5.4. Based on the results of the numerical simulation, the lone Reynolds stress source model is found to under-estimate the strength of the source field, which suggests that the viscous stress can be a significant acoustic source at least in the regions close to the solid surface. Outside the viscous boundary layer, however the acoustic source fields is shown to be dominant by the Reynolds stress.

It has been found in test problems 2 and 3 that the temporal gradient of the velocity field may be only relevant to the acoustic production at the boundary. When there is strong temporal gradients at the truncated surfaces of source regions, such as case of the spinning vortex pair (Section 5.3.5), the velocity-based source models tend to under-estimate the strength of the acoustic near field, compared to the Lagrangian Multiplier-based source model. On the other hand, both the Reynolds and viscous stress model (Model 3) and the Lagrangian multiplier model (Model 2) have produced the same acoustic far-field in the numerical simulation of vortex shedding from a square cylinder, where there is no significant temporal gradients at the boundary of the source field. The results from both numerical studies support the initial hypothesis on the role of Lagrangian multiplier $P$ as a ‘reservoir’ for storing acoustic energy in an incompressible formulation and support the idea that the resurgent acoustic field does not interact with the main flow field in low Mach numbers to a large extent.

It is worthwhile to discuss the impact of the Mach number which is implemented as a part of the wave operator (i.e. the propagation field) in the new formulation. Similar treatment has also been adapted in the numerical studies of vortex shedding from a 2D circular cylinder for modelling the Doppler effect which is said to be one of the main causes of the polarity of the radiation field whose maximum strength is shown to be tilted towards the front sector of the cylinder ($\theta_{\text{max}} \sim 78.5^\circ$) while phase differences exists between acoustic waves at different angles at the same radius [233]. The authors argued that the Curle’s analogy is derived based on a static far field, therefore a correction related to the speed of sound $c(\theta) = c_\infty(1 - M\cos(\theta))$, needs to be introduced to properly represent the acoustic propagation in a moving fluid [233].

In the model there is also clear evidence of polarity of the far-field ($\theta_{\text{max}} \approx 70^\circ$ at $r = 40$) and phase differences among acoustic signals at different angles even before the added correction...
of the speed of sound $c$, as shown in Figures 5.31 and 5.35 respectively. The result is likely to be due to the uneven distribution of the acoustic loads on different faces of the solid body, which vary in strength, frequency and phase. From Figures 5.25 to 5.27, it can be found that there are larger acoustic loads on the front face than on the back of the cylinder while there is the acoustic sources due to the viscous stress operating at the drag frequency on both front and back faces but not acting on the top face. In Figure 5.28, the acoustic loads are stronger towards the front end of the cylinder and that there is a phase difference between the source waves calculated at different distances to the leading edge. As a result, there are two small sectors around the x-axis ($\theta = 0^\circ, 180^\circ$) where the acoustic far-field is dominated by the drag frequency and the maximum acoustic strength shifts towards the front sector of the propagation field, which is supported by the findings in [233] for a circular cylinder. In other words, the new modified Lighthill equation is capable of not only predicting the acoustic signals in stagnated far fields but also accounting for the Doppler effect in a dynamic media.

The physical shape of the solid body is an important factor in determining the property of the acoustic far field along with those aerodynamic conditions (e.g. Re and Mach numbers). Consequently, it may not be sufficient to model the acoustic near field as a point source even in the case that the predicted acoustic wave length is much larger than the physical size of the source region. Additional evidence to support the view, can be found by comparing the acoustic far-field produced by the vortex shedding from a square [236] and a circular cylinder [233], the polarity of of the former is shown to be shifted more towards the front face than the later under the same aerodynamic conditions, while the maximum strength of the acoustic far-field is larger for the circular cylinder than for the square one, probably due to the larger lift force acting on the surface of the former ($C_L \sim 0.5$ for the circular vs $C_L \sim 0.4$ for the square).

### 5.6 Summary

The analogy-based aeroacoustic simulations focussed on 2D numerical problems with the customised solver. The FE model comprises two separate procedures where the incompressible Navier-Stokes equations are numerically resolved in order to provide the acoustic source field for solving the Lighthill equation in variational forms. Three variations of the Lighthill acoustic source field are implemented, including the Reynolds stress model, the Reynolds and viscous stress model, and the Lagrangian multiplier model. The CFD code has been validated on the 2D backward facing step problem while the simulated acoustic far-fields are compared with the reference data in the numerical studies of the 2D spinning vortex pair and 2D vortex shedding from a square cylinder.
Chapter 6

Aeroacoustic Modelling of Fricatives
/s/ and /ʃ/

In [49], Shadle raised the question ‘What controls the nature of the fricative sounds?’ and presented a series of mathematical and experimental work in a search of the answer. Shadle proposed two types of acoustic sources (obstacle and non-obstacle or wall sources) and made the conclusion that the most influential factors in determining the acoustic quality of the fricative sounds are the presence of an obstacle (i.e. teeth), the length of the cavity anterior to the most constricted area and the flow rate [49, 52]. In this chapter, the newly developed aeroacoustic model, introduced in Chapter 5, is applied to simulate the two fricative sounds /s/ and /ʃ/ on Shadle’s simplified mechanical models. The reason for choosing Shadle’s mechanical models is two-fold. First, the mechanical models for the /s/ and /ʃ/ have been extensively studied and the aeroacoustical measurements are available for comparison with numerical models. Second, the simplified mechanical models have been shown to produce realistic physical properties for the fricatives sounds [49].

Section 6.1 gives a brief review of past studies of fricatives and discusses the subject in terms of the articulation, aerodynamic and acoustic properties. The mathematical modelling is presented in Section 6.2 where three pairs of 2D models are numerically simulated to produce acoustic spectra. The results and findings are discussed in Section 6.3.

6.1 Physical Properties of Fricatives /s/ and /ʃ/

Before and after Shadle’s study, experimental work and mathematical modelling have been developed to facilitate the understanding of the physical nature of the production of fricative consonants. As introduced in Chapters 2 and 3, technologies like X-ray [11], MRI [239, 240, 19], EMA [127, 131], EPG (Electropalatography) [241, 242] and the like have been widely employed in studies of the articulation of fricatives. Aerodynamic and acoustic quantities have been measured in numerous studies on either human subjects [243, 55, 49], or from mechanical models of the human vocal tract [46, 49, 53, 52, 244, 245]. On the other hand, information
obtained from experiments can not fully explain the physical nature of the sound, therefore mathematical models, mainly based on the source-filter theory [246, 11, 46, 1, 49, 47, 247], amongst many others, have been used to produce synthesised spectra for fricatives. In the next few sections, a brief review of the past studies is made for the two fricative consonants in terms of their articulation, aerodynamic and acoustical properties. Here, /s/ is defined as the unvoiced alveolar fricative as the leading consonant in English word ‘seem’ and /ʃ/ is defined as the unvoiced palatal-alar fricative as the leading consonant in the word ‘sheet’.

6.1.1 Articulation

The two fricatives of interest are sometimes called ‘stridents’ because in both articulations, there are narrow constrictions of the VT posterior to the lips, directing airstreams to the teeth [239]. The constriction (i.e. the minimum cross-sectional area in the VT) is formed between the tongue and alveolar ridge for the /s/ and between the tongue and hard palate in the /ʃ/ configuration [239]. The airflow is prevented from leaking laterally by the tongue-palate contacts [158]. The constriction divides the VT roughly into the anterior and posterior cavities. The anterior cavity of /s/ is both shorter in length and smaller in volume than the one for /ʃ/, owing to the more anterior position of the constriction. In addition, sublingual cavities have been observed for the /ʃ/ configurations on human speakers [239]. The most anterior part of the tract has a more circular cross-sectional shape than the /s/ configuration, largely due to the effect of lip rounding [239]. Immediately posterior to the constriction site, /ʃ/ shows a more gradual expansion of the cross-sectional areas than in the /s/ case. Compared to /s/, the /ʃ/ configuration also has a larger cross-sectional area in the upper pharynx region where the VT is widest for both fricatives, due to the rise of the posterior part of the tongue in the latter case [239]. Further back towards the glottis, the configurations of the two fricatives have been shown to share similar area functions for the same speaker [239, 19]. Both fricatives have been described as ‘voiceless’, meaning no glottal vibration during the phonation and the glottis remains somewhat more open than during vowel production [246].

The above description gives a general picture of the articulation of the two sounds. However, the articulation does vary from speaker to speaker. In [241], the study shows that one test speaker produced /s/ through a 6-8mm wide groove near the front of the alveolar ridge while the other produced the same sound by placing the groove near the back of the ridge. For /ʃ/, both speakers produced 10-12mm grooves behind the posterior border of the alveolar ridge. Jackson et. al [240] also reported that their measured constriction area for /s/ (1cm²) is much larger than the ones measured in [239] (0.1-0.3cm²). It is also worth noting that there is experimental evidence of intra-speaker difference in the articulation for the same fricative and one of the major contributors is the linguistic context where the fricative is placed (e.g. preceding or following vowels) [241].

For the purpose of excluding personal variations on the articulation and on the sound
quality, a number of parametised VT models have been proposed for both experimental and mathematical modelling in [248, 246, 49, 52, 158, 249, 250, 251] among others, where a two-chamber model (constriction-front cavity) has been detailed in [246]; three-chamber geometric models including a back cavity and an obstacle, acting as the lower teeth, have been used in both of the experimental measurement and mathematical models built in [49, 52], and upper teeth have been added into the models in [250, 158].

6.1.2 Aerodynamics

The acoustic mechanism for the voiceless fricatives are not well understood [49, 53]. Nevertheless, it is known that there is no sound source due to glottal excitation, such as is the case in voiced sounds [1]. It is also believed that sound is induced by airflows related to unsteady aerodynamic events casued by forcing the airstream through narrow constrictions [46]. Stevens in [2] summarises the mechanism of sound generation from a jet passing through a narrow constriction: 1) the velocity fluctuations downstream to the constriction site; 2) the jet interacting with a solid surface; 3) the velocity fluctuations within the constriction site.

Regarding the production of /s/ and /ʃ/, it was initially postulated that the source of sound originates from random fluctuation of pressure around the constriction, induced by the turbulence when the flow exceeds a critical Reynolds number (e.g. $160 < Re < 1200$ for circular constrictions) [252]. Later, more evidence was revealed by both experimental [49, 53, 52] and mathematical investigations [54], that the interactions of the turbulent air with the surrounding walls, such as the teeth immediately anterior to the most constricted region, generate much more acoustic energy than the turbulent air itself. In [247], a more analytical framework was developed to explain the extraction of acoustic energy from non-acoustical aerodynamic processes for the unvoiced sounds. The model is built on vortex sound theory [228] which expresses the sound generation as interactions of the velocity and vorticity. Due to the process known as ‘boundary layer separation’ after the injection of jet out of the constriction, there are vortical structures formed in the anterior cavity. The convected vortices induce forces on the obstacles (rigid surfaces) in the path of the jet, hence enhance the production of acoustic energy. Therefore, the sound source can be mathematically calculated via the convolution of a function representing the temporal behavior of the acoustical energy radiated by a single vortex, with another function describing the arrival of vortices in the source region. In this framework, the characteristics of the acoustic source field are determined by both the properties of the jet vorticity (e.g. jet speed, distribution of vorticity etc.) and the shape of the VT anterior to the constriction [247]. A similar mathematical model also based on the vortex sound theory, is presented in [158], describing the sound generation by the jet vorticity passing through a pair of incisors.

The aforementioned studies all employed empirically derived source-filter models optimised to fit experimentally collected data. As mentioned in Chapter one, such source-filter type
models are easy to implement, but they give very little insight into the nature of the aeroacoustic sound generation process for speech production. A few DNS models have attempted to simulate fricative sounds. Among them, 2D simplified geometric models have been investigated in [253, 254] where the author finds that the aperture of the constriction makes a large impact on the turbulent jet formed downstream. A range of CFD techniques are tested on a 2D /ʃ/ model based on a sagittal profile of a realistic VT [50] where the results show that RANS methods are unable to produce the physical spectrum while an acoustic analogy coupled with incompressible model seems to produce the best outcome. The dynamics of vortices downstream to the constriction, has been investigated in [249], using 3D LES simulation on simplified 3D geometries. The study finds that the spatio-temporal distribution of the vorticity field is sensitive to the shapes of anterior cavity in terms of the shape of the constriction and the distance between the jet and wall. A more realistic 3D geometric model for /s/ based on the data collected from MRI scans is used in 3D LES simulations in [51].

A range of flow parameters have been proposed to simulate the physiological conditions of the VT flow for the voiceless fricatives, which are listed in Table 6.1. It seems that the physiological flow involved in a human fricative is in the low Mach number range (< 0.1), therefore the use of incompressible flow equations may be justified for numerical simulation, such as in the models presented in [249, 51, 254].

<table>
<thead>
<tr>
<th>Reference</th>
<th>Flowrate (cm$^3$)</th>
<th>Constriction Size (cm$^2$)</th>
<th>Max. Re</th>
<th>Max. M</th>
</tr>
</thead>
<tbody>
<tr>
<td>[46]</td>
<td>200-500</td>
<td>0.05-0.2</td>
<td>3,300-17,000</td>
<td>0.3</td>
</tr>
<tr>
<td>[52]</td>
<td>160-420</td>
<td>0.08</td>
<td>4,300-11,200</td>
<td>0.06-0.15</td>
</tr>
<tr>
<td>[249]</td>
<td>332</td>
<td>0.125</td>
<td>5,000</td>
<td>0.08</td>
</tr>
<tr>
<td>[251]</td>
<td>180</td>
<td>0.06</td>
<td>5,000</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Regarding the boundary conditions, the subglottal pressure is assumed to be reasonably stable under normal circumstances [46], therefore a constant subglottal pressure [54] or a steady flow rate [49, 251] can be applied at the glottal end. The VT wall has often been treated as a rigid dry surface in most of the computational models [50, 51, 254] and in those mechanical models [49, 53, 52] while it is in fact damp and compliant in physiological conditions, which may induce additional acoustic energy loss at certain frequencies [52]. As for the far-field condition, both experimental and mathematical models often assume quiet and static media (i.e. ambient air) where the ‘listening point’ is placed [52, 251].

### 6.1.3 Acoustics

Spectral properties of fricatives vary depending on speaker and phonetic context [246]. Even for the same speaker, the spectra of the same fricative might show variations due to a change in the articulation (e.g. jaw positions, etc.) or a change in aerodynamic conditions (e.g. intra-oral pressure, etc.) [158]. Regardless, they can normally be distinguished by English speakers in
the same phonetic context, therefore, the cues on which this differentiation occurs can only be based on acoustical stimulus [243].

From the preception test conducted in [246], both /s/ and /ʃ/ are shown to have broad spectra where the peak frequency for /s/ is in the range (3,500-6,400Hz) while the peak frequency for /ʃ/ is in the range (2,200-2,700Hz). According to the study done by Shadle [49], the dynamic range (the difference between maximum and minimum amplitudes of the spectrum) is approximately 20dB for the /s/ and 37.1dB for the /ʃ/ among the test speakers. Further spectral characteristics of the two fricatives have been summarised in [255] in terms of spectral moments, dynamic amplitude and slope above maximum amplitude.

Attempts have been made to simulate or synthesise spectra for the two fricatives either in experiments [49, 53, 52] and via numerical models [47, 54, 158]. Among them, Shadle concluded that there is no evidence of source-filter interaction in fricative-like mechanical models [49], hence the ‘source’ (i.e. acoustic production) and ‘filter’ (i.e. acoustic propagation) can be modelled independently.

**Acoustic Source**

In the literature, aeroacoustic analogies, separating the flow domain (near field for evaluating the aeroacoustic source) and acoustic domain (the far field for wave propagation), have been widely adopted in the studies of human speech sounds [193, 194, 50, 253, 195]. The acoustic source may be examined in terms of three essential elements: the location, the amplitude and the spectral shape.

The location of the source for fricative sounds has been a long standing subject of research. It became first known that the noise source is located somewhere in the VT rather than at the glottis [246]. Later, Stevens noticed that turbulence is generally generated in rapid flows near a surface within or immediately anterior to an articulated supra-glottal constriction, which induces ‘fricative’ noise as an excitation of the VT system [46]. Based on the experimental results of fricative-like mechanical models, Shadle pointed out that the presence of an obstacle (rigid surface), such as teeth directly in the path of the airflow downstream to the constriction greatly enhances the sound production compared to the same configuration without such obstacles, thus the pressure fluctuation at the teeth is assumed to be the best source model for fricatives /s/ and /ʃ/ [49]. This view is further supported by the experimental results on more realistic geometric models of VT where the ‘obstacle’ in the /ʃ/ was identified as the lower anterior teeth [53, 52].

However, such views are challenged in [54] where the authors applied numerical modelling to analyse the possible locations of the source in the form of acoustic monopoles, representing the volume velocity fluctuations, as a point source at the end of the constriction and dipoles, representing the fluctuation of force acting on the obstacle, as a uniformly distributed source around the obstacle. The study used 1D area functions derived from MRI scans to represent
the VT geometry and concluded that there is likely to be a distributed dipole source spanning the VT walls and teeth anterior the the constriction in the /ʃ/ configuration, while the sound production seems to be concentrated at the teeth in the /s/ case. Following this the upper and lower incisors were modelled as two normal planes with a gap in-between [158], the study found that the overall sound pressure is twice the pressure fluctuation in the turbulent boundary layer of the impinging jet and the transformation of the hydrodynamic energy into sound occurred predominantly at geometrical irregularities, particularly at sharp edges where the Green’s function is singular. The DNS study was performed on a 3D MRI-derived VT model of /s/ where the authors found that the highest production of the acoustic energy in the form of Powell’s acoustic source [190], is possibly located close to the surfaces of lower anterior teeth [51].

Regarding modelling the strength of the fricative source, Fant [11] developed a mathematical relation for a pressure source whose strength is proportional to \( Re^2 \), which implies a positive relation between the constriction size and flow rate with the strength of the sound source. In contrast, based on findings from both the experimental and theoretical work, Stevens [46] reported that the source pressure level of the turbulence noise is only weakly dependent on the size of the constriction and is largely dependent on the pressure drop across the constriction \( \Delta P \). Given a reasonably stable subglottal pressure during the production of a fricative, the turbulent \( \Delta P \) across a constriction can be estimated via a relation involving the volume flow rate, constriction area and the length of the constriction [47].

As mentioned previously, Shadle recognised two distinctive source mechanisms in models of fricatives where the /s/ and /ʃ/ are classified as ‘obstacle’ noise as opposed to the ‘wall’ noise supposedly in the vicinities of the constriction. It was found that the the ‘obstacle’ type is more efficient in converting aerodynamic energy into sound, with a higher rate of change of sound pressure with volume velocity compared to the ‘wall’ source [53]. A theoretical framework for explaining the role of an obstacle in increasing acoustic efficiency was developed [247] which relates the amplitude of the sound source to the size of the vortices, the convection speed and the radial change of the VT geometry.

The spectral shape of sound source for /s/ has been modelled as a flat peak at 800-4,000Hz followed by a -6dB/octave drop, while peak frequencies of 300-2,000Hz followed by a -12dB/octave drop has been proposed for the /ʃ/ in source-filter models [11]. Parametised geometric models have been employed in measuring sound sources in the studies presented in [49] where the measured spectra were inverse filtered with the known Greens function of the test geometries. One potential weakness of this method is the pre-determination of source location [48] which Shadle admits, although it works well for the ‘obstacle’ noise, does not produce satisfactory results for the ‘non-obstacle’ case (geometries with a constriction only).

Another way of predicting source spectra is via mathematical optimisations such as the parametric model developed in [54] which involves three parameters (peak frequency, roll-off rate below the peak frequency and roll-off rate above the peak frequency) which are optimised to fit the data of speech sounds, The study suggested a peak frequency for /ʃ/ in range of
2,100-4,100Hz, roll-off rate below the peak frequency 2dB/octave, roll-off rate above the peak frequency 0-16dB/octave and the peak frequency for the /s/ is 2,600-4,700Hz, 1-3 dB/octave roll-off rate below the peak, 0-12dB/octave roll-off rate above the peak.

**Acoustic Propagation**

The propagation properties of the VT in the fricative consonants have been modelled in a similar way to those in vowel modelling. 1D area functions of fricatives have been measured in MRI during sustained articulations [11, 239, 240, 19] and the VT is viewed acoustically, as a series of concatenated tubes with uniform diameters in each section. However, unlike the vowels, the fricatives have the source somewhere above the glottis, hence altering the transfer function of the tract.

An idealised model for fricative consonants was presented in [246] where the source is placed at the anterior end of the constriction, only the constriction and an anterior cavity were considered in the transfer function where the poles (i.e. resonance) are calculated as half of the length of the constriction and a quarter of the length of the anterior section, while the zeros (i.e. anti-resonance) evaluated as a quarter of the length of the constriction. The results imply that the /s/ configuration has a shorter anterior cavity and shorter constriction sites than the /f/ configuration, therefore the spectrum of /s/ has a peak situated at a higher frequency than /f/. As the constriction site moves towards the lips, the first pole of the VT transfer function increases in frequency [46]. This is confirmed by the numerical study performed in [54], where the authors reported a shift of the dipole source from 0.9cm from lips to a location 2.1cm anterior to the oral constriction, causing the frequency of the zero from about 2,000Hz to 3,700Hz. The effect of the sublingual cavity in the oral cavity, observed on the VT for /f/, has been studied in [54] where it is found to have an impact of a 100-350Hz decrease in the frequency of the lowest front cavity resonance. Subglottal coupling at the glottal opening was discussed in [47], where additional resonance frequencies at 1kHz and 2.2kHz were added.

For each individual section of the VT, filtering properties such as wall compliance, heat conduction through the wall, wall friction and viscous damping in the far-field, have been incorporated into the mathematical models based on simplified physical relations which often involve parameters derived from experiment measurements [46, 49, 47]. On the other hand, numerical models based on CFD techniques often employ simplified propagation conditions for the VT domain, such as acoustically hard walls, no viscous damping or heat loss (i.e. isentropic condition) [249, 51, 254].

Overall, various studies [50, 53] suggest that three-dimensionality of the flow is required for accurate modelling of the aeroacoustic source for the fricatives while a 1D area function is sufficient for predicting the filtering properties [53] of the VT.
6.2 Aeroacoustic Simulations on Simplified 2D Geometries for Fricatives /s/ and /ʃ/

The numerical model used in the aeroacoustic simulation of the fricative sounds is built on the two-stage simulation solving incompressible CFD and the Lighthill aeroacoustic analogy, as presented in Chapter 5. The decision to test the model on simplified 2D VT models is motivated by the following questions:

1. How well can the modified Lighthill analogy combined with an incompressible flow model simulate the fricative sounds?
2. What are the geometric and numerical approximations required to accurately simulate realistic fricative sounds?

Although several numerical models have been employed in aeroacoustic analogies [50, 253, 254] for VT configurations for fricatives, none has adopted a formulation based on volume integrals, but rather rely on pre-determined ‘source surfaces’ such as the teeth surface downstream to the constriction [254], or the constriction, lips and teeth surfaces [50]. Such representations might not be comprehensive enough to accurately evaluate the source field. As discussed in the numerical study of vortex shedding from a square cylinder in Chapter 5, there is a large amount of cancellation of the source field within the boundary layer of the cylinder as well as in the vortex train behind the cylinder. This view is also supported by the study in [232], where the author found that different intergration surfaces applied to the Ffowcs-Williams Hawkings equation, lead to different calculated acoustic far-fields, possibly due to the truncated quadruple source moving through the integrated surface.

More evidence revealed by LES on an MRI-based /s/ model in [51] with an analysis of Powell’s source field \( (u \times \omega) \) shows that the highest valued sound source is not likely to be located on the surface of anterior teeth since the vorticity level is not the highest in this region, although in theory a source surface could cover the entire flow domain, and therefore capture all the sound production region. In practise, the simulated CFD domain may impose limitations on finding such a surface, as is the case of the alternating vortex train artificially terminated at the edge of the CFD domain in the vortex shedding from a square cylinder example in Chapter 5. To summarise, a surface-based integral method is a truncation of the overall source field defined in acoustic analogies. As a result this could lead to insufficient modelling of the source field for fricative sounds.

There also could be numerical advantages in using a variational form of the Lighthill analogy over those based on integrals of a ‘source surface’. As discussed in [256], the source terms obtained by typical CFD techniques usually do not represent acoustic pressure fluctuations accurately enough, especially on solid boundaries. Compare Equations (5.1.9) and (5.3.2), on the boundaries of source domain, the integral-based Ffowcs-Williams Hawkings analogy yields
\[ \partial(F_{ij}(\delta(S))/\partial x_j + \partial(Q_{i}\delta(S))/\partial t), \] which means that the gradients of pressure and viscous stress need to be explicitly evaluated on the VT walls. There is no such requirement when bridging the source and acoustic domain as the variational formulation of the Lighthill analogy gives \[ \partial(P - c_0^2 \rho')/\partial x_i n_i = 0 \] on the shared boundary.

Many fricative models including the ones presented in [50, 254, 51, 249], involve turbulence modelling by LES whose function is to separate the flow domain into large scales and small scales, via a filtering procedure. The method is based on the assumptions first proposed by Kolmogorov [257] which states that the large scale of the flow is the primary source of the turbulent kinetic energy while the small scales are isotropic in their character and only responsible for viscous dissipations. Based on the assumptions, those so-called ‘subgrid-scale’ flows can be modelled statistically, reducing the computational demands on the numerical solver, while the energy-bearing large scale flow can still be calculated in a similar fashion to the DNS method [168]. However, when modelling fricative sounds and dealing with wall-bounded flows, the velocity gradients near the no-slip VT boundaries could become extremely large at high Reynolds number and the results strongly depend on the grid resolution in the viscous sublayer especially in LES [168]. Given that in LES the aerodynamic quantities in the boundary layer are often modelled by subgrid-scale models (e.g. the Smagorinsky eddy-viscosity model [258]), rather than directly calculated by solving the governing relations, it can be reasonably expected that the volumetric source adopted by the variational form of the Lighthill analogy to have better accuracy than the surface-based source model used by other acoustic analogies under the same grid resolution used by LES in wall-bounded aeroacoustic modelling of fricative sounds.

Given the modelling framework, both the geometric and numerical requirements need to be considered in order to successfully simulate realistic sounds. Ideally, it is desirable to model VT geometries as anatomically accurately as possible. From Chapter 4, it has been learnt that there are two significant differences between the two modelled fricative geometries and the sustained articulations in MRI, namely the maximum constricted areas and the placement of the constriction. As indicated in Figure 4.18, the most constricted areas in the ‘Talking Head’ models are much less than the measured versions (i.e. 43mm$^2$ in linear and 117mm$^2$ in cubic /s/ models respectively, vs 9mm$^2$ in MRI; 41mm$^2$ in linear and 45mm$^2$ in cubic /ʃ/ models respectively, vs 18mm$^2$ in MRI) and the constrictions sites are approximately 8mm anterior to the corresponding sites in MRI for both models. The potential cause for such differences were discussed in previous chapters, whereas the impact of discrepancies are of interest and investigated in the aeroacoustic simulation of the two fricative sounds and whether the modelled geometries are adequate for simulating realistic fricative sounds.

The two mid-sagittal profiles of the two cubic models are shown in Figures 6.1(a) for the /s/ and 6.1(b) for the /ʃ/. In both cases, the VT shapes are shown to have relatively small variations (< 50%) in diameter until reaching the main constricted region which is located close to the lower anterior teeth in the /s/ case and between the tongue and gums of the upper teeth for the /ʃ/ model. The maximum cross-sectional area ratio for both 3D models are close
to 12. Downstream and anterior to the constriction sites, the /s/ model has the upper teeth sitting in close proximity (< 3mm) as shown in Figure 6.1(c), while outlines of both upper and lower teeth can be seen approximately 13mm away from the most constricted area for model /ʃ/, forming a second constricted zone, as shown in Figure 6.1(d). The distances between the teeth and the constriction calculated from the customised fricative models seem to be in good agreement with Shadle’s mechanical models, as demonstrated in Figure 6.2 where the VT has been divided into three concatenated tubes, a back cavity, a constricted section and another cavity anterior to the end of the constriction, with a semi-circular obstacle, representing the teeth, placed normal to the central axis, and the distances between the obstacle and constriction is 5mm for the /s/ and 15mm for /ʃ/ model [49].

Figure 6.1: 2D mid-sagittal profiles of the cubic fricative models. (a) The mid-sagittal shape of the cubic model of the /s/. (b) The mid-sagittal shape of the cubic model of the /ʃ/. (c) A close-up view of the constricted region of the /s/ model. (d) A close-up view of the constricted region of the /ʃ/ model.

On the other hand, the location of the source region is investigated so that the computational cost may be reduced by narrowing the size of the CFD domain. As for the wall-bounded 3D DNS modelling, it is estimated that the level of numerical complexity is proportional to $\sim Re^{3.4}$ while methods like LES might reduce the numerical requirements for the simulation, the computational cost is still proportional to $\sim Re^{2.4}$ [259]. The LES simulation on a simplified
geometric model of /s/ in [249], consisting of 3.3M cells in the mesh, ran at approximately 605s per time step. A full-length 3D /f/ model, created by extruding a 2D sagittal contour of a realistic VT, was simulated by LES on a grid made of 1.3M cells and took approximately 304s/step [50], where the author realised that the results by LES are limited by the mesh being too coarse. A truncated 3D model of the /s/ based on MRI scans, covering part of the oral cavity and adjacent open space, containing up to 71.2M cells is presented in [51]. Considering that the aeroacoustic modelling of fricatives usually requires tens of thousands of time steps in order to built up a spectrum, the size of the simulation domain needs to be limited in order to complete the simulation in a feasible time.

As both geometric features (i.e. the constriction ratio and the length of the frontal cavity) can be reflected in 2D geometric models, as shown in Figures 6.2 for Shadle’s mechanical models, a great deal of simplification can be achieved by simulating fricative sounds in a 2D domain. According to the study done in [50], the results produced by the 2D models do not match well with Shadle’s experimental data partly due to the disrepencies in aerodynamic conditions caused by lack of variation in the 3rd dimension and the inability of the 2D CFD to simulate real turbulence. Nonetheless 2D models have been employed in many numerical studies of human phonation for the purpose of identifying sound production mechanism of the glottis [160, 193]; investigating the impacts of glottis shapes [193], the subglottal pressure and oscillation frequencies [161] on the sound production. Likewise, similar 2D mathematical models can be useful for the purpose of validating the fricative geometries.

Furthermore, the numerical requirements for simulating fricatives are investigated in the study. As introduced in Chapter 5, the application of the aeroacoustic analogy allows the solution of the flow independent of the acoustic field, thus potentially reducing the size of the simulation domain for the CFD. The same can be done with temporal resolution. The time scale used for the CFD simulation might not be sufficient for solving the acoustic field. The reason could be due to the distributed nature of the source and the numerical dissipation of the acoustic propagation. The difference between 2D and 3D wave propagation (i.e. \( r \) in \( \mathbb{R}^2 \) vs \( r^2 \) in \( \mathbb{R}^3 \)) may also be considered and the 2D results are normalised in order to match better to the 3D geometry.

In summary, the aeroacoustic modelling work introduced in this chapter is designed as a pilot study for future 3D numerical simulations for the customised fricative models presented in Chapters 4 and 5. The study aims to use Shadle’s simplified geometric models of fricatives in a 2D domain to investigate the mechanism of sound production and propagation with an emphasis on the likely impacts of the constriction ratio and the length of the anterior cavity. The results of the study will determine the future modelling requirements for simulating realistic fricative sounds.
6.2.1 Problem Description

Considering the similarities between the 2D shapes of the fricative models of the ‘Talking Head’ and Shadle’s mechanical models, the aeroacoustic equations are numerically solved on a 2D sagittal profile of the original 3D fricative configurations proposed in [49] in the new analogy-based mathematical models, as shown in Figure 6.2. In the models, the shape of the VT can be conceptualised as a three-chamber tube, where the domain is divided into a back cavity, a constriction zone, a frontal cavity and a section at the end of the tube for representing the open space. There are six test variations of these geometric models in the study presented below, where an obstacle in the form of a 2mm wide rectangular rod, is sitting vertically downstream of the constriction in some of the configurations for representing the teeth. Similar to the third numerical example presented in Chapter 5, the aeroacoustic modelling is divided into a CFD part where the incompressible Navier-Stokes equations (5.2.3) are numerically resolved for extracting the acoustic source field, followed by a CAA modelling where the modified Lighthill equation (5.3.8) is numerically solved for the acoustic far field.

\[
\begin{align*}
\rho_0 &= 0.00118\, g/cm^3; \\
c &= 34480\, cm/s; \\
\nu &= 15\, cm^2/s; \\
u_0 &= 850\, cm/s;
\end{align*}
\]

Figure 6.2: The computational domain for Shadle’s fricative models. The dimensional parameters are listed in Table 6.2.

The CFD simulation starts with creating a stationary solution by artificially increasing the viscosity of the fluid and gradually increasing the inlet flowrate. The solution becomes the starting point for a transient simulation of the incompressible Navier-Stokes equations at the intended \( Re \) number with relatively large time steps for establishing the flow. The dimensions of the CFD domain for all the six test examples are listed in Table 6.2. In all cases, the length
of the back cavity is shortened to 25mm to reduce the domain size, since there is no evidence of significant sound production reported in [49], which is also supported by the findings from this numerical study.

The shape and size of the open-space is adjusted to the path of the outgoing vortex trains and the grid is gradually stretched to allow dissipation of the vorticity. The boundary conditions are the same for all six examples, which are constant flowrates at the inlet (outlined in red solid lines in Figure 6.2); no-slip VT walls (outlined in black solid lines in Figure 6.2 and no-stress Neumann type conditions at the edge of the open-space (outlined in blue solid line in Figure 6.2). Before the flow exits the domain, numerical damping factors, in the form of gradually increased viscosity, are introduced in the ‘buffer zone’, in order to dissipate the vortices to avoid pressure reflections caused by truncations of the flow, similar to the sponge layer method described in [260]. In all test examples, the buffer zones are located in the open-space region approximately 100mm away from the edge of the domain and the source field from the buffer zone is discarded in the following acoustic simulation.

Following this, another run of transient simulation is conducted with smaller time steps for the purpose of evaluating acoustic sources. As outlined in Figure 6.2 by dashed lines, the Lighthill equation is solved on an extended grid of the CFD domain for resolving the acoustic far-field. In this case, new elements are added to the CFD inlet for extending the back cavity to the full original length as in Shadle’s model and the open-space is enlarged to include the ‘listening point’ which is located along the central axis at 26cm from the end of the tract \( r_{\text{listening}} \). The placement of the ‘listening point’ is consistent with the design in Shadle’s experiment. Considering the difference in the radiation conditions between 2D and 3D models, the far-field spectrum is scaled by \( Hb/2r_{\text{listen}}\pi \) which takes account of the intensity of the acoustic source (\( r \) in 2D and \( r^2 \) in 3D) and the radiation relation (\( r^2 \) in 2D and \( r^3 \) in 3D). The boundary conditions defined for the CAA simulations for all six test examples are the same, which are the constitutive relations for all the shared borders with the CFD domain and non-reflecting conditions for the far-field (shown as dashed lines in Figure 6.2), as defined in Section 5.3.4.

The meshes for all six examples are created as structured grids made of quadrilateral elements. Along the x-axis, the cells are gradually compressed from the inlet to the entry of the constriction section with a ratio of 0.95; then \( \delta x \) is constant throughout the constriction and compressed with the ratio of 0.95 when entering the front cavity and towards the posterior (leading) edge of the obstacle in Models 3-6 whereas the cells are stretched with a ratio of 1.05 along the x-axis for the non-obstacle cases (Models 1-2). Behind the obstacle the cells are stretched with the ratio 1.05 along the x-axis until reaching the edge of the open-space. The grid cells are axi-symmetric to the y-axis. Starting from the centerline (i.e. \( y = 0 \)), the cells are first compressed with a ratio of 0.95 so that a finer layer of cells can be placed near the walls in the constriction region. After this, the cells are expanded with a ratio of 1.05 to half way between the wall of the constriction and the wall of the front and back cavities (i.e. the two
cavities have the same diameters) then compressed again with a ratio of 0.95 until reaching the wall of the two cavities. In the open-space, the grid is stretched with the ratio of 1.05 in the y-direction until the edge of the domain.

A view of the coarse mesh used for Model 3 is shown in Figure 6.3 and the parameters of all six meshes are summerised in Table 6.3. Taylor-Hood elements are used for the FE fields, which means that Quadratic basis functions are used for spatial, velocity and acoustic fields while the Linear type is used for the Lagrangian multiplier.

The first set of examples are designed to investigate the role of the front cavity with respect to its length (i.e. the distance between the constriction and the lips). Both of Shadle’s fricative models have a constriction ratio of 16 based on the mid-sagittal contours of the 3D shapes, which is close to the calculated values from the 3D fricative models (~ 12). Shadle argues that the lengthening of the front cavity will lower the resonance frequency as well as narrowing the

<table>
<thead>
<tr>
<th>Name</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_{b1}(\text{mm}) )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>( L_{b2}(\text{mm}) )</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>( L_{c}(\text{cm}) )</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>( L_{f1}(\text{mm}) )</td>
<td>15</td>
<td>30</td>
<td>5</td>
<td>15</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>( L_{f2}(\text{mm}) )</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>( L_{o1}(\text{mm}) )</td>
<td>600</td>
<td>600</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>( L_{o2}(\text{mm}) )</td>
<td>0</td>
<td>0</td>
<td>300</td>
<td>300</td>
<td>350</td>
<td>350</td>
</tr>
<tr>
<td>( L_{l}(\text{mm}) )</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
<td>260</td>
</tr>
<tr>
<td>( H_{b}(\text{mm}) )</td>
<td>25.4</td>
<td>25.4</td>
<td>25.4</td>
<td>25.4</td>
<td>25.4</td>
<td>25.4</td>
</tr>
<tr>
<td>( H_{c}(\text{mm}) )</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>( H_{o1}(\text{mm}) )</td>
<td>400</td>
<td>250</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>( H_{o2}(\text{mm}) )</td>
<td>400</td>
<td>250</td>
<td>150</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
</tbody>
</table>

Table 6.3: Grid Parameters for the Simulated Domains

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Nodes CFD</td>
<td>394,821</td>
<td>378,741</td>
<td>503,041</td>
<td>592,841</td>
<td>1,602,457</td>
<td>1,782,225</td>
</tr>
<tr>
<td>No. Nodes CAA</td>
<td>412,871</td>
<td>396,791</td>
<td>739,851</td>
<td>836,851</td>
<td>2,073,447</td>
<td>2,253,215</td>
</tr>
<tr>
<td>No. Elements</td>
<td>98,200</td>
<td>94,200</td>
<td>125,140</td>
<td>147,540</td>
<td>399,440</td>
<td>444,336</td>
</tr>
<tr>
<td>No. Elements</td>
<td>102,700</td>
<td>98,700</td>
<td>129,640</td>
<td>208,440</td>
<td>517,080</td>
<td>561,976</td>
</tr>
<tr>
<td>( \delta x ) near the obstacle wall (mm)</td>
<td>nil</td>
<td>nil</td>
<td>0.018</td>
<td>0.015</td>
<td>0.028</td>
<td>0.027</td>
</tr>
<tr>
<td>( \delta y ) near the constriction wall (mm)</td>
<td>0.036</td>
<td>0.036</td>
<td>0.036</td>
<td>0.036</td>
<td>0.0045</td>
<td>0.0045</td>
</tr>
<tr>
<td>( \delta t ) at the listening point (mm)</td>
<td>3.2</td>
<td>3.2</td>
<td>2.5</td>
<td>2.5</td>
<td>1.24</td>
<td>1.24</td>
</tr>
<tr>
<td>( \delta t ) CFD coarse (ms)</td>
<td>0.1</td>
<td>0.1</td>
<td>0.05</td>
<td>0.05</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>( \delta t ) CFD fine (ms)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>( \delta t ) CAA (ms)</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>
formant’s bandwidth [49]. However, this would require an acoustic source to occur close to the exit end of the constriction. As reported in other studies [54, 249], the aeroacoustic source due to separation of the boundary layer of jets is likely to be distributed in space. As shown in the MRI measurements, both of the simulated fricative shapes have more anterior positions for the main constriction sites due to a more forward articulation of the tongue. Thus, the actual role of the length of the front cavity becomes a subject of the investigation. The length of the front cavity is 15mm for Shadle’s /s/ configuration, which is approximately the same as the one in the simulated /s/ model (~18mm), and is 30mm for Shadle’s /ʃ/ configuration, similar to the calculated value from the simulated /ʃ/ model (~32mm).

The second pair of examples focusses on the effect of an obstacle which is placed 5mm away from the constriction for the /s/ and 15mm for the /ʃ/ in Shadle’s obstacle models. The obstacle is supposed to mimic the presence of the teeth, which can be observed in similar distances (9mm for /s/ and 19mm for /ʃ/) anterior to the constriction site in the simulated fricative models. The purpose of the modelling is to investigate the sound production mechanism with the presence of a downstream obstacle (i.e. teeth) and the location of the sound source. Although the 2D shape is exactly the sagittal profile of the 3D mechanical model, the contraction ratio of the 2D models is only a quater of the value measured from 3D geometry due to contractions in the 3rd dimension. Unlike a 3D circular constriction of a radius of \( r \), which has an area contraction(expansion) ratio of \( H_c^2/H_b^2 \), the 2D medial model only has a ratio of \( H_c/H_b \). As
a result, the jet speed is reduced by a factor of $H_c$ in the 2D solutions compared to the 3D version, given the flow is nearly incompressible.

Coincidentally, the FE models of fricatives also have smaller constriction ratios ($\sim 12$) compared to the measured shapes by MRI ($25 - 48$). Therefore, the decision was made to normalise the constriction ratio in the third pair of examples by reducing $H_c$ to reflect the same area ratio as in the 3D model (i.e. 64). As mentioned in many other studies [247, 158, 54, 253], the speed of the jet is a crucial element in defining the sound source properties of fricatives. Hence, the simulated acoustic field is expected to better match the experimental data than the two previous cases. However, the change does not affect the Reynolds number of the 2D flow as it does in the 3D cases.

### 6.2.2 Case 1: The Effect of the Length of the Front Cavity

Models 1 and 2 represent the non-obstacle version of Shadle’s simplified geometric models of fricatives. Between the two models, the only difference is the length of the front cavity which is 15mm for /s/ and 30mm for /ʃ/ [49].

**Flow Domain**

**Steady-State** The streamlines for the stationary flow for Models 1 and 2 are shown in Figures 6.4(a) and (b) respectively. In general, the flow pattern of the two models are very similar, which shows separation upon leaving the constriction and the symmetrical recirculation zones formed between the walls of the front cavity and the jet and a smaller pair formed in the back cavity just before the entrance to the constriction. In each case, the recirculation zone stretches to the full length of the upper and lower walls of the front cavity. In model cases, the jet gradually slows down and widens as it enters the open space due to conservation of momentum.

The pressure drop across constrictions of various sizes and shapes has been studied in [49], whose results have been used to derive the mathematical relations for sound sources in an empirical model. A mesh analysis has been carried out to investigate the change of the pressure drop across the constriction with respect to mesh resolutions during the artificial steady-state. The study is not a direct measurement of the mesh performance of the flow at the modelled $Re$ number, however it can serve as an indication of the minimum requirements for modelling the flow. The mesh is refined by subdividing existing elements equally in both directions (i.e each element into 4 new elements), starting from the coarsest one. The results are listed in Table 6.4, which indicates that the pressure drop across the constriction section is approximately 1% or less between the first and second refinement. Therefore, the steady-state solution on Mesh 2 is chosen as the basis for the further transient simulation.
Figure 6.4: The streamlines of the artificial steady-state, where the magnitude of the velocity field is normalised by $u_0$. (a) Model 1, (b) Model 2.

Table 6.4: Mesh Analysis of the Pressure Drop across the Constriction

<table>
<thead>
<tr>
<th></th>
<th>$\nu$ Factor</th>
<th>Num Elements</th>
<th>$P_c/\rho_0u_0^2H_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>33.34</td>
<td>24,550</td>
<td>-32.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>98,200</td>
<td>-30.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>392,800</td>
<td>-30.32</td>
</tr>
<tr>
<td>Model 2</td>
<td>33.34</td>
<td>23,550</td>
<td>-31.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>94,200</td>
<td>-30.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>376,800</td>
<td>-30.31</td>
</tr>
</tbody>
</table>

**Transient Simulations with Large Time Steps** Next, the flow is simulated with the specified viscosity and run as a time-dependent CFD with a relatively large time step $\delta t$. The transient simulation allows the development of the flow and an example of streamlines of the developed flow is shown in Figure 6.5 where the velocity field at each grid point is calculated as the mean vector over the period of 10ms. From the figure, a major difference in the jet behavior between Models 1 and 2 can be observed, which is that the vortex train, caused by the breaking up of the jet, is largely passing freely into the open space in Model 1 while the vortices interact with the walls of the front cavity in Model 2.

As most of the vortical structures originate in the front cavity and cause pressure fluctuations along their path, the temporal behavior of the pressure drop in the front cavity is analysed in Figures 6.6(a) and (b) to show the development of the flow. The simulated flow lasts for 3,000 time steps or 300ms, which are divided into three periods for the frequency analysis. As can be seen from the frequency analysis in Figures 6.6(a) and (b), the slow signals ($< 1,000\text{Hz}$) of the pressure fluctuation in the front cavity decay more rapidly than those higher frequencies, as
the flow develops for both Models 1 and 2 and there is significantly less variation in the signal after the first period.

**Transient Simulations with Smaller Time Steps** To evaluate the sound source to be used for the acoustic simulation, the temporal resolution of the flow needs to be investigated in order to generate a comprehensive representation of a source for a wide range of frequencies in the acoustic analogy. The Lighthill aeroacoustic source is integrated on cross-sections at the end of each tract, which is used as an indicator for the acoustic source convected into the open space from the space within the tract.

In Figure 6.7, three sizes of $\delta t$ are chosen to study the temporal behavior of the acoustic source calculated at the end of the tracts. The results show that the solutions produced by $\delta t = 0.01ms$ are adequate for capturing the form of the source wave in both models but still under-estimate the strength when compared to solutions produced by half its size. A decision was made so that a balance between the total simulation time, which is related to the total number of time steps and the lowest frequency in the spectrum, and the accuracy of the results. In the end, $0.01ms$ is chosen for building the source field in the acoustic simulation for Models 1 and 2.
Figure 6.6: Evolution of the pressure drop in the front cavity with $\delta t = 0.1ms$, where the magnitude of the pressure is normalised by $\rho_0u_0^2H_b$. (a) Model 1 ($t = 0 - 300ms$); (b) Model 2 ($t = 0 - 300ms$).
Figure 6.7: Temporal history of the Lighthill source calculated at the end of the tract, where the magnitude of the stress is normalised by $u_0^2/H_b$ and at the temporal resolution of $\delta t = 0.025, 0.01$ and 0.005 ms. (a) Model 1; (b) Model 2.
Acoustic Domain

Acoustic Near Field The fine resolution transient CFD simulation lasts for 2,500 time steps or 25ms following the run with large time steps. In Figures 6.8, the acoustic near field is represented by the time-averaged source fields of Models 1 and 2. Recall the relations (5.3.3) derived in Chapter 5, the Lighthill acoustic source field $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ can be decomposed into $(\nabla u)^2 - \omega^2$.

Based on the results, it can be seen that the majority of the acoustic source seems to be distributed spatially downstream of the constriction site for both models, however the spread of the source is wider for Model 2 than Model 1 in the front cavity. The maximum acoustic production occurs outside the tract and within the first 20mm from the end of the constriction in Model 1 while the average level of acoustic source within the front cavity is higher for Model 2. In the primary acoustic production region, the total acoustic source shows good correlation with the vorticity/enslurophy waves in most of locations in both models.

To get a more quantitative analysis of the spatial distribution of the source field, the two mathematical forms are integrated over each section of the models and the results are plotted in Figure 6.9.

The results show that there is no significant amount of acoustic production prior to front cavities for both models. On the other hand, Model 2 has more acoustic production and vorticity in the front cavity than in the open space, while the opposite is true for Model 1. Apart from the difference in the strength of the source field in the front cavities, the spectral shape of the source field for Model 2 appears to have a smaller roll-off rate towards higher frequencies ($2,000Hz$) than the source waves for Model 1.

A close look inside the front cavities, as shown in Figure 6.10, reveals that a longer front cavity (Model 2) generates more acoustic source on solid walls than a short front cavity (Model 1), likely due to the increased interactions between the break-off vortices and the wall. Another noticeable difference between the two models is the enhanced role of the back wall in the acoustic production for Model 2 than for Model 1, which implies stronger backward travelling vortice waves for a longer front cavity.
Figure 6.8: The time-averaged acoustic source fields of Models 1 and 2, where the magnitude of the source field is normalised by $u_0^2/H_b^2$ and shown in the logarithmic scale of base 10. (a) The acoustic source field due to the vorticity $\omega^2$ for Model 1 ($t = 315 - 325\,ms$); (b) the full Lighthill source field $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 1 ($t = 315 - 325\,ms$); (c) the acoustic source field due to the vorticity $\omega^2$ for Model 2 ($t = 315 - 325\,ms$); (d) the full Lighthill source field $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 2 ($t = 315 - 325\,ms$).
Figure 6.9: The acoustic source fields evaluated in each sections of Models 1-2, where the magnitude of the source field is normalised by $u_0^2$. (a) The acoustic source due to vorticity $\omega^2$ for Model 1; (b) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 1; (c) the acoustic source due to vorticity $\omega^2$ for Model 2; (d) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 2.

Figure 6.10: The acoustic source fields evaluated on the solid surfaces in the front cavities of Models 1-2, where the magnitude of the source field is normalised by $u_0^2/H_b$. (a) The acoustic source due to vorticity $\omega^2$ for Model 1; (b) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 1; (c) the acoustic source due to vorticity $\omega^2$ for Model 2; (d) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 1.
Acoustic Far Field  The acoustic simulation is done by solving the modified Lighthill equation on the extended grid to the CFD domain, at the same spatial resolution as the fine temporal transient CFD simulation. The resulting acoustic field, as pictured in Figure 6.11, has a weak radiation field in the open space along with some localised high and low pressure regions associated with the convected vortices. The ‘listening’ point is located 250mm away from the end of each tract, the same as in Shadle’s models. The resulting spectra, shown in Figure 6.12, have 2,500 sampling points and have a resolution of 40Hz which is the same as in Shadle’s analysis. However, there is no averaging among multiple sampling sequences as Shadle did in her study, due to the computational cost of the numerical simulations.

![Figure 6.11](image.png)

Figure 6.11: The simulated acoustic field where the acoustic pressure is normalised by $\rho_0 c_0^2$. (a) Model 1 at $t = 325ms$; (b) Model 2 at $t = 325ms$.

Based on the results given in Figure 6.12, it can be observed that the introduction of viscous effects in the source field does not alter the spectral shape of the far-field but rather uniformly distributes energy across the spectra. Interestingly, the addition of the viscous source reduces the overall amplitude of the signal for Model 1 while it enhances the strength of the far field in Model 2. There are large differences between the simulated spectra and Shadle’s experiment data for both models. In particular, the strength of the simulated spectra are 10-20dB less than the measured ones. Both simulated spectra show little variation in signals above 3kHz unlike the measured ones. The only common features shared between simulated and measured spectra for both Models 1 and 2 is the energy peak located around 1,200Hz. Comparing the two simulated spectra, Model 2 has more energy in frequencies below 3,000Hz than Model 1.
Figure 6.12: The spectra of acoustic fields evaluated at 250mm away from the end of the tract. (a) Model 1; (b) Model 2.
6.2.3 Case 2: 2D /s/ and /ʃ/ Models with Obstacles

In Models 3 and 4, a 2mm wide obstacle is added in the front cavity for mimicking the teeth in the oral cavity. According to Shadle, the /s/ model (Model 3) has the obstacle placed 5mm downstream of the constriction site while the obstacle is located 10mm further downstream as a result of a longer front cavity in the /ʃ/ model [49].

Flow Domain

Steady-State   The artificial steady-state solutions for Models 3 and 4 are reached in the similar manner to the previous two non-obstacle cases, except for an increased artificial viscosity used for stabilising the solution, as indicated in Table 6.5. The general flow pattern of the static flows are calculated as streamlines and presented in Figures 6.13(a) for Model 3 and (b) for Model 4. From the figures, it can be seen that the introduction of an obstacle generates more vortical structures in the front cavity, especially in the regions between its leading edge and the end of the constriction.

![Streamlines](image)

Figure 6.13: The steamlines of the artificial steady-state, where the magnitude of the velocity field is normalised by $u_0$. (a) Model 3; (b) Model 4.

The mesh analysis, listed in Table 6.5, shows similar relations between the mesh resolution and the pressure drop for the constriction for both models. Considering the complexity of the flow, meshes of the highest resolution have been chosen as the basis for further transient simulations.

Transient Flow Field with Large Time Steps   Starting with the solution of the artificial steady state, the flow is simulated with the specified viscosity and solved as a time-dependent
Table 6.5: Mesh Analysis of the Pressure Drop across the Constriction

<table>
<thead>
<tr>
<th>Model</th>
<th>Viscous Factor</th>
<th>Num Elem</th>
<th>$P_c/\rho_0u_0^2H_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 3</td>
<td>66.67</td>
<td>31,285</td>
<td>-59.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>125,140</td>
<td>-58.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500,560</td>
<td>-57.77</td>
</tr>
<tr>
<td>Model 4</td>
<td>66.67</td>
<td>36,885</td>
<td>-59.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>147,540</td>
<td>-58.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>590,160</td>
<td>-57.79</td>
</tr>
</tbody>
</table>

CFD. After experimenting with different time step sizes for the transient simulation, it has been found that a smaller step size than the one used for Models 1 and 2 needs to be applied to the two new models. So $\delta t = 0.05ms$ is chosen to build up the flow and the resolved streamlines at the end of the run are plotted in Figures 6.14 for both models.

![Streamlines](image1)

Figure 6.14: The streamlines of the averaged transient flow, where the magnitude of the velocity field is normalised by $u_0$. (a) Model 3 ($t = 165 - 175ms$); (b) Model 4 ($t = 165 - 175ms$).

Some major differences between the two flow models can be observed in Figure 6.14. Firstly, the high speed zone is split in two regions in Models 4 (inside the jet and on the obstacle) while the high speed zone is confined within the jet in Model 3. Secondly, the vortical structures within the front cavity are more compact in the Model 3 compared to the ones in Model 4. Lastly, the manner of interaction between the jet and obstacle is remarkably different between the two models, where the jet can be seen largely intact when hitting on the leading edge of the obstacle in Model 3 but already broken into vortices before reaching the obstacle in the case of $/f/$, resulting in lots of interactions between vortices with the surrounding walls of the front cavity other than the surfaces of the obstacle.

Again, an analysis has been presented in Figures 6.15(a) and (b) to show the development...
of the flow in terms of the pressure fluctuations across the front cavity in both models. The transient simulation for both models lasts for 150ms and is divided into three even periods for the analysis. From the results, it can be seen that the pressure fluctuation changes significantly after the first periods for both models. Between second and third periods, variations less than 0.5dB occur mainly for frequencies below 2,500Hz for Model 4 and more so for Model 3.

**Transient Simulation with Smaller Time Steps** The results of the temporal analysis of the acoustic sources calculated at the end of the tract, are presented in Figure 6.16 for the two models. In both cases, there is less than 15% variation between the solutions calculated at $\delta t = 0.01ms$ and $\delta t = 0.005ms$. Therefore, the decision was made to use $\delta t = 0.01ms$ for building the aeroacoustic source field.
Figure 6.16: Temporal history of the Lighthill source at the end of the tract for $\delta t = 0.025$, 0.01 and 0.005 ms, where the magnitude of the source is normalised by $u_0^2/H_b$. (a) Model 3; (b) Model 4.
Acoustic Domain

Acoustic Near Field  The distribution of the Lighthill aeroacoustic source fields are evaluated for Models 3 and 4 in Figures 6.17. The results show that the presence of the obstacle drastically changes the distribution of the source field from the non-obstacle models. The leading edge of the obstacle appears to be one of the main source regions for both models. On the other hand, the source field for Model 4 is more spatially distributed in the front cavity than in Model 3.

Figure 6.17: The time-averaged acoustic source fields of Models 3 and 4, where the magnitude of the source field is normalised by $u_0^2/H^2$ and shown in the logarithmic scale of base 10. (a) The acoustic source field due to the vorticity $\omega^2$ for Model 3 ($t = 165 - 175ms$); (b) the full Lighthill source field $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 3 ($t = 165 - 175ms$); (c) the acoustic source field due to the vorticity $\omega^2$ for Model 4 ($t = 165 - 175ms$); (d) the full Lighthill source field $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 4 ($t = 165 - 175ms$).
Similar to the findings in Models 1 and 2, it is seen that there is no significant acoustic production prior to the front cavity for the new pair of models, as shown by the results presented in Figure 6.18. Furthermore, the contributions from the front cavity to the total acoustic production seems to be more dominant in Model 4 than in Model 3.

Figure 6.18: The acoustic source fields evaluated at each sections of Models 3-4, where the magnitude of the source field is normalised by \( u_i^2 \). (a) The acoustic source due to vorticity \( \omega^2 \) for Model 3; (b) the full acoustic source \( \partial u_i/\partial x_j \cdot \partial u_j/\partial x_i \) for Model 3; (c) the acoustic source due to vorticity \( \omega^2 \) for Model 4; (d) the full acoustic source \( \partial u_i/\partial x_j \cdot \partial u_j/\partial x_i \) for Model 4.

Figure 6.19 shows the results for the acoustic source distributed on solid walls of the front cavities for Models 3 and 4, where the solid surfaces are divided into three groups (the back wall, upper wall and obstacle surfaces, as marked in Figure 6.2). In both models, the strongest acoustic source is located near the surface of the obstacle, however the source field for Model 4 seems to be more concentrated around the obstacle than in Model 3. Investigating the relationship between the vorticity field (i.e. enstrophy waves) and the acoustic source field, it is seen that high vorticity usually leads to strong acoustic production, with a noticeable exception at the upper wall of the front cavity where the acoustic production is suppressed by equally strong gradient fields for both models.

**Acoustic Far Field** The acoustic simulation for Models 3 and 4 are similar to the two previous models. The normalised acoustic fields are shown in Figures 6.20(a) for Model 3 and 6.20(b) for Model 4. From the figures, it can be seen that both simulated acoustic fields show sparse plane wave patterns in the open space, emerging approximately 100mm away from the end of the tract.

The spectra of the two models are produced by transforming 2,500 sampling points which gives 40Hz resolution.
Figure 6.19: The acoustic source fields evaluated on the solid surfaces in the front cavities of Models 3-4, where the magnitude of the source field is normalised by $u_0^2/H_b$. (a) The acoustic source due to vorticity $\omega^2$ for Model 3; (b) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 3; (c) the acoustic source due to vorticity $\omega^2$ for Model 4; (d) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 4.

Figure 6.20: The simulated acoustic field where the acoustic pressure is normalised by $\rho_0 c_0^2$. (a) Model 3 at $t = 175ms$; (b) Model 4 at $t = 175ms$.

From the results presented in Figure 6.21, it can be observed that neither of the two models resemble the far-field spectrum of Shadle’s measurements, while both models significantly under-estimate the strength of the acoustic far-field. In fact, the two 2D configurations produce very similar acoustic far-fields. Comparing to the non-obstacle models (1 and 2), the presence
of an obstacle, downstream to the constriction, produces more acoustic energy in the far field (10-20dB). The introduction of viscous stress in the source field does not significantly alter the spectra, implying Reynolds stress dominates the acoustic production at the specified $Re$. As a result, the viscous source is removed from the source models for the next pair of configurations which are expected to have even larger local $Re$ number in the flow.
6.2.4 Case 3: The Modified 2D /s/ and /ʃ/ Models with Obstacles

In the last pair of examples, the /s/ and /ʃ/ configurations are modified to match the jet speed to Shadle’s 3D models, by narrowing the constriction to 0.8mm in diameter. The remaining parameters for Models 5 and 6 are the same as Model 3 and 4 respectively.

Flow Domain

Steady-State The general flow pattern for the artificial steady state solutions of Models 5 and 6, shown as streamlines in Figures 6.22, are quite similar to those states of Model 3 and 4, but has a much higher jet speed. The artificial viscosity level has to be raised considerably in order to achieve the static equilibrium.

![Streamlines of artificial steady-state](image1)

Figure 6.22: The streamlines of the artificial steady-state, where the magnitude of the velocity field normalised by $u_0$. (a) Model 5; (b) Model 6.

The mesh analysis, listed in Table 6.6, shows that the pressure drop across the constriction is insensitive to the three test mesh resolutions, likely due to the artificially low $Re$ numbers for both models. Considering the larger viscous factors and larger pressure drops across the constriction than in the two previous models, which implies that there are far more unsteady energy in Models 5 and 6, the decision was made to use the meshes with the highest resolutions (1.6M nodes for Model 5 and 1.8M nodes for Model 6) for the transient simulations.

Transient Flow with Large Time Steps The first transient simulation on Models 5 and 6 are carried out at $\delta t = 0.01ms$. The selected $\delta t$ has to accommodate the development of the flow, and maintain a solution convergence within 5 or 6 iterations per time step. The general
Table 6.6: Mesh Analysis of the Pressure Drop across the Constriction

<table>
<thead>
<tr>
<th>Model</th>
<th>Viscous Factor</th>
<th>Num Elem</th>
<th>$P_c/\rho_0u_0^2H_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 5</td>
<td>200</td>
<td>24,965</td>
<td>-272.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99,860</td>
<td>-271.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>399,440</td>
<td>-271.02</td>
</tr>
<tr>
<td>Model 6</td>
<td>200</td>
<td>27,771</td>
<td>-272.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>111,084</td>
<td>-271.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>444,336</td>
<td>-271.02</td>
</tr>
</tbody>
</table>

Flow profiles for Models 5 and 6 are shown as streamlines calculated at the end of the simulation in Figures 6.23(a) and (b) respectively.

Figure 6.23: The streamlines of the averaged transient flow, where the magnitude of the velocity field is normalised by $u_0$. (a) Model 5 ($t = 40 - 45ms$); (b) Model 6 ($t = 40 - 45ms$).

The streamlines of Models 5 and 6 show some similarities with the flow patterns of Models 3 and 4 respectively shown in Figures 6.14(a) and (b), where the /s/ configurations (Model 3 and 5) have the jet broken up on the obstacle and the vortex train travels upwards close to the walls of the baffle after leaving the tract, while in the /f/ models (Models 4 and 6) the vorticies are generated before the jet reaches the obstacle and the vortex trains leaves the tract in a more forward path compared to the previous cases. Apart from the increased speed of flow due to narrower constrictions, the new models (5 and 6) generate more rapid and complex vortex dynamics downstream to the constriction than the un-modified models (3 and 4).

Figure 6.24 shows the development of the pressure fluctuation within the front cavity during the transient run, which is divided into three equal periods of 10ms each. After the frequency analysis, it is evident that the two flows undergo different evolutions from the similar artificial
steady state; Model 6 is shown to have a much larger roll-off rate and lower peak frequency than Model 5.

Figure 6.24: Evolution of the pressure drop in the front cavity with $\delta t = 0.01 ms$, where the magnitude of the pressure is normalised by $\rho_0 u_0^2 H_b$. (a) Model 5; (b) Model 6.

**Transient Simulation with Smaller Time Steps** Three different temporal resolutions have been tested for the calculated aeroacoustic source at the outlet of the tract, the results are shown in Figure 6.25. For both models, reducing $\delta t$ from 0.005ms to 0.0025ms has less than 15% change in peak values. Considering the primary frequency range of interest for fricative analysis falls between 1,000-10,000Hz in Shadle’s analysis [49], a temporal resolution of 0.005ms has 20 sampling points per wavelength for a 10,000Hz signal, therefore it is considered to be adequate for CAA simulations [169]. Given a fixed amount of computational resource, it is better to use relatively large $\delta t$ to build longer periods so that a finer frequency resolution can be achieved than using a smaller $\delta t$ and shorter periods.
Figure 6.25: Temporal history of the acoustic source calculated at the end of the tract, where the magnitude is normalised by $u_0^2/H_b$. (a) Model 5; (b) Model 6.
Acoustic Domain

Acoustic Near Field  The time-averaged acoustic source fields for Models 5 and 6 are shown in Figures 6.26(a) and 6.26(b) respectively. From the figures, it can be seen that the distribution of the acoustic source is concentrated in Model 5 more than in Model 6. The spread of the primary source region in Model 5 appears to be between the obstacle and the end of the constriction, while two separate regions in the front cavity (the area near the end of the constriction and the area around the obstacle) seems to be significant for acoustic production in Model 6.

Figure 6.27 further quantifies the level of increase of acoustic source (\(\sim 20\,dB\)) in all sections for Models 5 and 6, comparing to Models with smaller constriction ratios (Models 3 and 4). The reduced constriction ratio appears to have the effect of increasing acoustic production in the front cavity for both Models 5 and 6 with respect to Model 3 and 4, while the front cavity is still the dominant source region in both configurations, it is coupled with the strongest production of vorticity over the entire domain.

From the results presented in Figure 6.28, the source waves on obstacles behave significantly differently between the two models. The obstacle appears to be a high frequency (\(5,000\,Hz\)) source in Model 5 while having a more rapid roll-off rate in Model 6, hence it becoming a relatively low frequency sound generator. Not only there are inter-model differences, within the same model, the source waves behaves differently according to their spatial locations. For instance, the upper wall in Model 5 is a relatively weak low frequency generator compared to the obstacle, while a similar role is performed by the back wall in Model 6.
Figure 6.26: The time-averaged acoustic source fields for Models 5 and 6, where the magnitude of the field is normalised by $u_0^2/H_b^2$ and shown in the logarithmic scale of base 10. (a) The source due to vorticity $\omega^2$ for Model 5 ($t = 40 - 45ms$); (b) the full source field $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 5 ($t = 40 - 45ms$); (c) the source due to vorticity $\omega^2$ for Model 6 ($t = 40 - 45ms$); (d) the full source field $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 6 ($t = 40 - 45ms$).
Figure 6.27: The acoustic source fields integrated over each section along the tract in Models 5 and 6, where the magnitude of the field is normalised by $u_0^2$. (a) The source due to vorticity $\omega^2$ for Model 5; (b) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 5; (c) the source due to vorticity $\omega^2$ for Model 6; (d) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 6.

Figure 6.28: The acoustic source fields evaluated on the solid surfaces in the front cavities of Models 5-6, where the magnitude of the source field is normalised by $u_0^2/H_b$. (a) The acoustic source due to vorticity $\omega^2$ for Model 5; (b) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 5; (c) the acoustic source due to vorticity $\omega^2$ for Model 6; (d) the full acoustic source $\partial u_i/\partial x_j \cdot \partial u_j/\partial x_i$ for Model 6.
Acoustic Far Field The acoustic simulation lasts for 15ms with $\delta t = 0.005 ms$. The acoustic fields for the two models at the end of the run are illustrated in Figures 6.29(a) for Model 5 and 6.29(b) for Model 6. From the images, plane wave patterns can be seen emerging in the open space approximately 100mm away from the end of the tract for both models, however the wave pattern in Model 5 is much more compact than those shown in Model 6.

Figure 6.29: The simulated acoustic field where the acoustic pressure is normalised by $\rho_0 c_0^2$. (a) Model 1 at $t = 45 ms$; (b) Model 2 at $t = 45 ms$.

The spectra produced by 2,500 sampling points, for both models are plotted in Figures 6.30(a) and 6.30(b) respectively against Shadle’s experimental measurements.

A few observations may be made from the above results. Firstly, the amplitude of the two spectra of the modified 2D models increases significantly from the models of smaller constriction ratio (Models 3 and 4) and reaches the average level of Shadle’s measurements, however the spectral shapes of the two models do not match the experimental results. Secondly, the two simulated spectra show significant differences between themselves. Model 5 shows a flat plateau from 1,000-5,000Hz followed by a gradual roll-off rate towards 10,000Hz (approximately 4dB/1kHz). In contrast, the spectrum of Model 6 starts a steeper decay from 2,000Hz onwards (approximately 5dB/1kHz). The peak value of Model 6 is approximately 10dB larger in magnitude than the ones in Model 5. Overall, Model 6 (i.e. the /f/ configuration) shows a higher peak at around 1,900Hz compared to the peaks shown in Model 5 (i.e. the /s/ configuration).

6.3 Discussion

Among the simulation results of Shadle’s fricative models, the obstacle models (Figure 6.21) are shown to be less sensitive to the viscous effect than the non-obstacle ones (Figure 6.12). This
suggests that the Reynolds stress is mainly responsible for the increased acoustic production in the so-called ‘obstacle noise’ sources [49] which is not located directly on the surface of the obstacle but rather in the vicinity around it. In all the test examples, the acoustic source due to viscous effects, are shown to have the same spectral behavior as the ones induced by the Reynolds stress.

The simulated spectra for all Shadle’s models fail to match the experiment measurements even after the adjustments for the jet speed and radiation conditions. One of the causes is likely to be in the nature of the 2D vortex dynamics, related to a process known as the inverse energy cascade [261]. In 3D, the nonlinear effect tends to break eddies (vortices) of large scales into small eddies (vortices) which then transfers energy into even smaller ones until reaching the visco-thermal scale where the energy is eventually dissipated as heat [262]. Such a process is called a direct energy cascade. However it is reversed in scales for 2D vortex dynamics via mechanisms like ‘vortex thinning’ [263].

As observed from the study of the spinning vortex pair in Chapter 5, each vortex represents
an acoustic source region which generates local pressure fluctuations as well as radiating acoustic energy into the far-field. The character of the radiated sound is closely related to the properties of the vortex (e.g. angular velocity, circulation intensity, etc.). As the energy is transferred from small scale vortices to larger ones, the angular velocity of the larger scales needs to be reduced due to the conservation of energy, which has the effect of lowering the frequencies of the radiated sound. As the cascade goes on, the large scale vortices can not be fully dissipated by fluid viscosity before entering the buffer zone where the source field is truncated in the acoustic simulation. Hence, the acoustic energy originally extracted from the flow via small scale vortices is eventually accumulated at the large scale level in the 2D aeroacoustic models.

As demonstrated in [50], a comparison between the 2D and 3D /f/ models reveals a shift of the energy toward higher frequency range in the far-field spectra produced by the 3D models. Their results also show that the spectra of the 3D models have a slower roll-off rate among frequencies above 4,000Hz than those produced by 2D models. On the other hand, the physical mechanisms for the extraction of acoustic energy from the flow are not expected to be drastically different between the modified 2D and 3D models. The main reason is that the production of vortices which is the main source for acoustic energy, are confined to relatively small spaces not affected by the energy cascade.

Considering the source properties in the two fricative models with respect to their calculated far-field spectra, there are two dominant source regions (one on the back wall of the front cavity and one on the surface of the obstacle) in Model 6 (/f/ in Figure 6.28(d)) while only one (on the surface of the obstacle) in Model 5 (/s/ in Figure 6.28(c)). The source field calculated on the back wall of the front cavity in Model 6 represents a low frequency generator compared to the one located on the surface of the obstacle which has a smaller roll-off rate towards the high frequency range (∼1dB/kHz for 4-10kHz). From the results of CFD simulations, different interaction patterns between the fluid and the solid surfaces are observed, where the jet is largely intact before reaching the obstacle and vortices are broken regularly off the jet core in the region above or behind the obstacle in Model 5 as shown in Figure 6.26(a), while the jet has already broken into vortices before reaching the obstacle in Model 6 as shown in Figure 6.26(d). Such differences in aerodynamics result in different sound production mechanisms which may be described as ‘jet-obstacle interaction’ for the /s/ model and ‘vortex-wall-obstacle interaction’ for the /f/ model.

Comparing Models 4 (Figure 6.19(d)) and 6 (Figure 6.28(d)), it is seen that the speed of the jet plays an important role in deciding the relative strength of the two source mechanisms. At a low jet speed, the obstacle source dominates the overall sound production in the /f/ model whose far-field spectrum (Figure 6.21(d)) is very similar to the one produced by the low-speed model of /s/ (Model 3 in Figure 6.21(c)). As the jet speed increases, the strength of the source field calculated on the back wall increases more than the amount added to the obstacle source, as shown in Models 5 (Figure 6.28(b)) and 6 (Figure 6.28(d)). This is likely due to the increased proximity of the vortices to the back wall caused by the high speed jet.
As the primary production region of vortices in the two obstacle models of the /ʃ/ (Models 4 in Figure 6.19(c) and 6 in Figure 6.28(c)) lie between the back wall of the front cavity and the obstacle, the jet oscillates between the upper and lower walls while breaking into swirls of vortices. Once created, vortices travel towards both the anterior and posterior sides of the front cavity. Similar observations on 3D models have been made in [249] for a non-obstacle configuration with an elliptic shaped constriction, where the author reported that simulated vortices bend back and forth along the jet axis. According to the simulation results, it also seems that the backward traveling vorticity waves have a slower speed than those travelling forward, therefore the interaction between vortices and the back wall occurs at a slower rate than the ones with the downstream obstacle. Hence, the source on the back side of the front cavity becomes a relatively low frequency sound source, with a higher roll-off rate towards the high frequencies than the one located on the surfaces of the obstacle.

The role of the back wall as a sound source is also revealed in the non-obstacle models (Models 1 in Figure 6.10(b) and 2 in Figure 6.10(d)). With a short front cavity, Model 1 has the jet broken up into vortices outside the tract. As a result, there are very low levels of vorticity and acoustic production on the back wall of the cavity compared to the upper and lower walls of the front cavity, as shown in Figure 6.10(a). Given a longer front cavity (Model 2), the jet oscillates inside the front cavity, leading to increased interactions between the backwards travelling vortices and the back wall which then becomes one of the primary source regions for the entire model, as shown in Figure 6.10(c). In this case, the longer front cavity may increase the flow resistance at the end of the tract, which in turn encourages the break-up of the jet and subsequent backward travelling vortices.

The oscillations between adjacent frequencies are much larger in the simulated far-field spectra than the ones given in Shadle’s data. This is most likely caused by lack of data averaging in the current model. In Shadle’s work [49], the acoustic spectra are produced by averaging 16 segments each made of 1024 sampling points with a frequency spacing of 40Hz, which results in 2dB confidence limit for the spectral amplitude. Due to computational constraints, there was no averaging performed on the model data set and thus no confidence limits are available.

Returning to the two main differences between the MRI and modelled fricative geometries and their likely aeroacoustical impact, the reduced constriction size in MRI is likely a compensatory measure for a lower flowrate during the sustained articulation compared to a more ‘normal’ speech environment in the EMA experiment. From the numerical studies, the reduced constriction size can be used to maintain a relatively high jet speed which has dual effects of both increasing the efficiency of extracting acoustic energy from the flow as well as enhancing the role of the back wall as one of the primary source regions for the /ʃ/ model.

There were more retracted placement of the constriction sites in both MRI configurations compared to the modelled ones. It was found that the most important factor in separating the /s/ and /ʃ/ is the distance between the constriction and obstacle (teeth), which leads to different sound production mechanisms and resulting far-field spectra. It is possible that during
MRI scans the subject had moved the jaw in such a way so that the lower teeth were placed more or less the same distance from the constriction, similar to the modelled values (∼5mm for the /s/ and ∼15mm for the /ʃ/). From the measured area functions of /ʃ/, a second constriction can be seen between the lower teeth and gum posterior to the first constricted zone between the tongue and hard palate, where the gap between them is approximately 20mm in both MRI and modelled versions. In the /s/ case, the MRI image resolution (7mm) is not sufficient to reveal the gap between the constriction and the teeth.

6.4 Summary

A two-stage aeroacoustic simulations have been conducted on 2D models of the fricative /s/ and /ʃ/ derived from Shadle’s experiments in [49], where the complex VT geometries were conceptualised into a three-chamber configuration (a back cavity, a constriction site and a front cavity) for the two fricatives. The CFD modelling is divided into three steps, including an artificial steady-state, a transient simulation with large time steps and a transient simulation with small time steps. Different numerical conditions are studied at each step of the CFD modelling. The solutions from the last CFD modelling step are used in the modified Lighthill equation for solving the acoustic far-field in the second stage of the aeroacoustic modelling. Various geometric factors including the length of the front cavity, the presence of an obstacle and the constriction ratio, are investigated with respect to their impact on the acoustic fields. The simulated acoustic far-field spectra are compared and discussed against the measurements collected in the original experiment.
Chapter 7

Conclusions

This work presents two main findings. In the first part (Chapters 2-4), an anatomically-based and customised finite element model of the human vocal system ("Talking Head") has been developed that combines static MRI scans & fast EMA and video recordings, thus providing both high spatial and high temporal resolution for modelling speech dynamics. Physiological constrains are added in the kinematic simulation of articulators, in the form of the Sobolev penalty functions, for improving the physical realism of the data-driven model while avoiding the complexity of simulating biomechanics.

The new overlapping scheme for the segmentation algorithm is shown to be more effective and robust than a conventional non-overlapping grid particularly for curved surfaces. The least-squares fitting algorithm is shown to be both accurate and efficient in generating high-order FE volumetric models when combined with the subdivided fitting strategy.

In the second part of the thesis (Chapters 5-6), new results for the computational modelling of fricative sounds are presented, using the newly developed aeroacoustic model based on the Lighthill analogy and incompressible flow. In the 2D backward facing step problem, the numerically resolved steady-state solutions show good agreement with the published data. The results from the 2D spinning vortex pair show that the solutions produced by the Lagrangian multiplier model match the analytical solution better than the Reynolds stress model, given the same numerical resolution. From the results of 2D vortex shedding from a square cylinder, it is evident that, while the Reynolds stress is the main contributor to the acoustic far-field, the viscous stress plays a significant role in the sound source at the solid surface. As predicted by the conservation principles, the solutions produced by the Lagrangian multiplier model are nearly identical to the ones calculated by the combined Reynolds and viscous stress. Attention needs to be paid to address the artificial termination of the convected vortex at the numerical boundary. The Hanning window approach requires a sufficiently large distance to ensure that there is no significant reflection due to abrupt truncation of the alternating source field. The introduction of Mach number effects calculated using the incompressible flows does not have a large impact on either amplitude nor the phase of the acoustic waves.

From the results of the aeroacoustic simulation on Shadle’s fricative models, it shows that,
there is no evidence to support any significant acoustic production in regions posterior to the constriction. In the absence of an obstacle, increasing the length of the front cavity increases the total acoustic production for the model, owing to enhanced interactions between the vortices and the walls of the front cavity. From the simulations of models with an obstacle, the results confirm the findings in [49] which states that the presence of an obstacle (i.e. teeth) downstream to the constriction, increases the overall acoustic production and the space around the obstacle is coincident with the strongest sound source for both the /s/ and /ʃ/ configurations. The leading edge of the obstacle is shown to be the place of strongest acoustic production in all obstacle configurations regardless of the length of the front cavity. However, for a configuration with longer gaps between the constriction and obstacle, there emerges a second primary sound production area close to the back wall of the front cavity. Reducing the constriction diameter has the effect of increasing the jet speed which has a dual effect on improving the overall efficiency of acoustic production while shifting the energy peaks towards higher frequencies in the far-field spectrum.
Chapter 8

Future Work

One of the limitations of the current articulatory model, as demonstrated in Chapters 3 and 4, is that there is no direct measurement of the dynamics of the pharynx and larynx cavities. As a result, there are only static models or passive movements of the articulators in the current implementation. However, the modelling framework described in Chapter 3 is such that dynamic measurements by other techniques such as ultrasound [38, 264] or fast MRI [71] can be translated into material points used by the host-mesh fitting method for improving the accuracy of the articulator models.

Although the host-mesh distortion technique results in accurate soft tissue movements, potential improvements may be made to eliminate the need for a host mesh. The purpose of placing a geometry of interest (i.e. the ‘slave’ model) which has a relatively complex shape and a high DOFs, into a simpler geometry (i.e. the ‘host’ mesh) is to bring gradual changes to the whole geometry according to a few highly localised measurements such like the tongue model where all measurements collected by EMA came from the top surface. In the models, Sobolev penalty functions have been introduced to add constraints to the deformation of the host. It could be a more straightforward practice to apply such constraints directly to the geometry since the design of the host mesh has to be tailored to suit the ‘slave’ models. In addition, it may be possible to introduce better biomechanical constraints to the tongue and cheek-lips movement in order to improve the modelling accuracy with respect to physiological realism. Instead of picking Sobolev weights by a trial-and-error process, optimisation procedures may be introduced via fitting the tongue shape to some envelope movements (e.g. elongation and retraction) determined by a continuum-mechanical model like the one described in [33].

It is useful to make a quantification on the effect of supine and upright speaking postures on the difference between the vocal tract shapes measured by MRI and EMA. A possible evaluation tool is through the acoustic analysis (e.g. Linear Prediction Coding) based on the area functions calculated from both the model and MRI, which should be effective, at least for vowels.

Regarding modelling the vocal tract, future improvements could be made into designing automated algorithms for creating initial meshes to better approximating the segmented VT.
shapes by the flood-fill algorithm discussed in Chapter 4. Instead of only allowing the plane normals to advance in a single vector, skeletonisation methods, such as the ones introduced in [265, 266] for analysis of human airway trees, may be more capable of representing those side branches which may be connected to the main airway. As for creating initial meshes for the least-squares fitting method, adaptive refinements should be implemented in deciding the resolution for the initial mesh based on the complexity of the actual geometry. The criteria for refinement could be based on the local error projections for individual elements.

For simulating voiced sounds and stops, an Arbitrary-Lagrangian-Eulerian (ALE) formulation, such as the one introduced in [199] for fluid dynamics, needs to be implemented for solving the acoustic domain. This requires the mesh to adapt new shapes while perserving the topology [194]. This can be achieved by refitting the existing cubic mesh to the new data cloud of the new VT shapes. As long as the change is gradual, similar fitting accuracy to the previous mesh can be expected. Furthermore, an ALE form of the Lighthill Equations are derived in Appendix E.

For all the 2D CFD models used in this study, it has been found that the DNS method is sufficient to produce stable converged solutions for the simulated flows (Re < 1,500). However the predicted Re number is higher for the 3D fricative models due to the extra variation of the cross sections in the third dimension, therefore a turbulence model, such as the LES method, is likely to be a necessary requirement for 3D CFD.

In Chapters 5 and 6, an alternative view of the application of the Lighthill aeroacoustic analogy for incompressible flows has been presented and the roles of different types of acoustic source models are investigated in a number of numerical studies. So far the model does not allow interactions between the resurgent acoustic waves and the mean flow. However, in some applications (e.g. whistling) such interactions may be important [49] for modelling voice. A model is further proposed to incorporate some forms of flow-acoustic interactions in a two-equation model via separation of variables. The new framework is based on field decomposition of the incompressible mean and compressible perturbations where a velocity potential is introduced for reducing the dimensionality of the system. Details of the derivation are presented in Appendix F.

Based on the results of the numerical studies presented in Chapter 6, it is evident that 3D modelling is an essential requirement for simulating realistic fricative sounds though measures need to be taken to restrict the computational time. It has also been learned from the 2D models, that the acoustic production for the fricatives is likely concentrated in parts of the VT immediately downstream to the most constricted region in the oral cavity while the flow can be largely considered as steady and silent until reaching the end of the constriction. Therefore, it is justifiable to truncate the 3D VT geometry to include only the constriction and part of the tract anterior to the constriction for simulating the acoustic source field (the near field). For those obstacle-models, all 2D results show that the overall strength of the aeroacoustic source decays as the vortices leave the tract. Hence, a steady flow rate or a static pressure
may be imposed at the inlet (the starting point of the constriction) and a buffer zone can be implemented right at the end of the VT, which significantly reduces the domain size for the 3D CFD.
Appendices
Appendix A

List of Words and Sentences Used in the EMA Experiment

Table A.1: A List of Speech Sentences Used in the EMA Experiment

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Spoke</th>
</tr>
</thead>
<tbody>
<tr>
<td>When sunlight strikes raindrops in the air, they act like a prism and form a rainbow.</td>
<td>56,57,222</td>
</tr>
<tr>
<td>Where were you while we were away?</td>
<td>58,59</td>
</tr>
<tr>
<td>Coconut cream pie makes a nice dessert.</td>
<td>60</td>
</tr>
<tr>
<td>That pickpocket was caught red-handed.</td>
<td>61</td>
</tr>
<tr>
<td>The eastern coast is a place for pure pleasure and excitement.</td>
<td>62</td>
</tr>
<tr>
<td>The prowler wore a ski mask for disguise.</td>
<td>64</td>
</tr>
<tr>
<td>Challenge each general’s intelligence.</td>
<td>65</td>
</tr>
<tr>
<td>Spring Street is straight ahead.</td>
<td>66</td>
</tr>
<tr>
<td>Too much curiosity can get you into trouble.</td>
<td>67</td>
</tr>
<tr>
<td>The paper boy bought two apples and three ices.</td>
<td>68</td>
</tr>
<tr>
<td>Clear pronunciation is appreciated.</td>
<td>69</td>
</tr>
<tr>
<td>Please sing just the club song.</td>
<td>70</td>
</tr>
<tr>
<td>Shell shock caused by shrapnel is sometimes cured through group therapy.</td>
<td>71</td>
</tr>
</tbody>
</table>
Table A.2: A List of Speech Words Used in the EMA Experiment (Part 1)

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Spoke</th>
<th>Stops</th>
<th>Spoke</th>
<th>Fricatives</th>
<th>Spoke</th>
<th>Numbers</th>
<th>Spoke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heed</td>
<td>2,224</td>
<td>Tea</td>
<td>19,249</td>
<td>Seem</td>
<td>48</td>
<td>Zero</td>
<td>72</td>
</tr>
<tr>
<td>Hid</td>
<td>3,225</td>
<td>Peace</td>
<td>20,250</td>
<td>Same</td>
<td>49</td>
<td>One</td>
<td>73</td>
</tr>
<tr>
<td>Head</td>
<td>4,226</td>
<td>Keith</td>
<td>21,251</td>
<td>Summer</td>
<td>50</td>
<td>Two</td>
<td>74</td>
</tr>
<tr>
<td>Had</td>
<td>5,227</td>
<td>Tar</td>
<td>22</td>
<td>Soon</td>
<td>51</td>
<td>Three</td>
<td>75</td>
</tr>
<tr>
<td>Hard</td>
<td>6,228</td>
<td>Path</td>
<td>23</td>
<td>Sawn</td>
<td>52</td>
<td>Four</td>
<td>76</td>
</tr>
<tr>
<td>Heard</td>
<td>7,229</td>
<td>Castle</td>
<td>24</td>
<td>Said</td>
<td>53</td>
<td>Five</td>
<td>77</td>
</tr>
<tr>
<td>Hoard</td>
<td>8,230</td>
<td>Talk</td>
<td>25</td>
<td>Sod</td>
<td>54</td>
<td>Six</td>
<td>78</td>
</tr>
<tr>
<td>Hod</td>
<td>9,231</td>
<td>Pour</td>
<td>26</td>
<td>Singer</td>
<td>55,83</td>
<td>Seven</td>
<td>79</td>
</tr>
<tr>
<td>Hud</td>
<td>10,232</td>
<td>Core</td>
<td>28</td>
<td>Sugar</td>
<td>84</td>
<td>Eight</td>
<td>80</td>
</tr>
<tr>
<td>Who'd</td>
<td>11,233</td>
<td>Beater</td>
<td>29</td>
<td>Sheet</td>
<td>85</td>
<td>Nine</td>
<td>81,82</td>
</tr>
<tr>
<td>Hay</td>
<td>12,235</td>
<td>Beeper</td>
<td>30</td>
<td>Sharp</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hi</td>
<td>13,236</td>
<td>Beaker</td>
<td>31</td>
<td>Shape</td>
<td>87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoy</td>
<td>14,237</td>
<td>Data</td>
<td>32</td>
<td>Shop</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoe</td>
<td>15,238</td>
<td>Darpa</td>
<td>33</td>
<td>Shame</td>
<td>89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How</td>
<td>16,239</td>
<td>Darker</td>
<td>34</td>
<td>Shoot</td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hear</td>
<td>17,240</td>
<td>Porter</td>
<td>35</td>
<td>Shed</td>
<td>91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hair</td>
<td>18,241</td>
<td>Pauper</td>
<td>36</td>
<td>Shawn</td>
<td>92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Porker</td>
<td>37</td>
<td>Cheese</td>
<td>93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Deet</td>
<td>38</td>
<td>Choose</td>
<td>94</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Deep</td>
<td>39</td>
<td>Chat</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Deek</td>
<td>40</td>
<td>Char</td>
<td>96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bart</td>
<td>42</td>
<td>Chore</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Barp</td>
<td>43</td>
<td>Chair</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bark</td>
<td>44</td>
<td>Change</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tort</td>
<td>45</td>
<td>Beach</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Torp</td>
<td>46</td>
<td>Parch</td>
<td>102</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Talk</td>
<td>47</td>
<td>Pooch</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Match</td>
<td>104</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>March</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fillet</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Borsch</td>
<td>107</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fashion</td>
<td>108</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feed</td>
<td>109</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Food</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Farm</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fail</td>
<td>112</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A.3: A List of Speech Words Used in the EMA Experiment (Part 2)

<table>
<thead>
<tr>
<th>Nasals/Liquids</th>
<th>Spoke</th>
<th>Clusters</th>
<th>Spoke</th>
<th>Consonants</th>
<th>Spoke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lima</td>
<td>113</td>
<td>Sleep</td>
<td>151</td>
<td>Game</td>
<td>195</td>
</tr>
<tr>
<td>Larry</td>
<td>114</td>
<td>Sledge</td>
<td>152</td>
<td>Gone</td>
<td>196</td>
</tr>
<tr>
<td>Loom</td>
<td>115</td>
<td>Slaw</td>
<td>153</td>
<td>Goon</td>
<td>197</td>
</tr>
<tr>
<td>Lucky</td>
<td>116</td>
<td>Slew</td>
<td>154</td>
<td>Gun</td>
<td>198</td>
</tr>
<tr>
<td>Larch</td>
<td>117</td>
<td>Slight</td>
<td>155</td>
<td>Gorge</td>
<td>199</td>
</tr>
<tr>
<td>Megan</td>
<td>118</td>
<td>Slower</td>
<td>156</td>
<td>Jane</td>
<td>200</td>
</tr>
<tr>
<td>Making</td>
<td>119</td>
<td>Salad</td>
<td>157</td>
<td>June</td>
<td>202</td>
</tr>
<tr>
<td>Marble</td>
<td>122</td>
<td>Silent</td>
<td>158</td>
<td>Jean</td>
<td>203</td>
</tr>
<tr>
<td>Model</td>
<td>123</td>
<td>Solo</td>
<td>160</td>
<td>Judge</td>
<td>204</td>
</tr>
<tr>
<td>Mood</td>
<td>124</td>
<td>Sealant</td>
<td>161</td>
<td>Join</td>
<td>205</td>
</tr>
<tr>
<td>Needling</td>
<td>126</td>
<td>Screwing</td>
<td>162</td>
<td>Even</td>
<td>206</td>
</tr>
<tr>
<td>Noodle</td>
<td>127</td>
<td>Clue</td>
<td>163</td>
<td>Oven</td>
<td>207</td>
</tr>
<tr>
<td>Narchos</td>
<td>128</td>
<td>Blue</td>
<td>164</td>
<td>Oeuvre</td>
<td>208</td>
</tr>
<tr>
<td>Nanny</td>
<td>129</td>
<td>Prune</td>
<td>165</td>
<td>Zoo</td>
<td>209</td>
</tr>
<tr>
<td>New</td>
<td>130</td>
<td>Stew</td>
<td>166</td>
<td>Dozen</td>
<td>210</td>
</tr>
<tr>
<td>You</td>
<td>131</td>
<td>Crew</td>
<td>167</td>
<td>Zen</td>
<td>211</td>
</tr>
<tr>
<td>Yard</td>
<td>132</td>
<td>True</td>
<td>168</td>
<td>Visit</td>
<td>212</td>
</tr>
<tr>
<td>Yeast</td>
<td>133</td>
<td>Grew</td>
<td>169</td>
<td>Buzz</td>
<td>213</td>
</tr>
<tr>
<td>Your</td>
<td>134</td>
<td>Flew</td>
<td>170</td>
<td>That</td>
<td>214</td>
</tr>
<tr>
<td>We</td>
<td>135</td>
<td>Glue</td>
<td>171</td>
<td>Though</td>
<td>215</td>
</tr>
<tr>
<td>Warm</td>
<td>136</td>
<td>Plume</td>
<td>172</td>
<td>Thought</td>
<td>216</td>
</tr>
<tr>
<td>Wonder</td>
<td>137</td>
<td>You</td>
<td>173</td>
<td>Theta</td>
<td>217</td>
</tr>
<tr>
<td>Woohoo</td>
<td>138</td>
<td>Bleat</td>
<td>174</td>
<td>Three</td>
<td>218</td>
</tr>
<tr>
<td>Wasting</td>
<td>140</td>
<td>Brew</td>
<td>175</td>
<td>Throw</td>
<td>219</td>
</tr>
<tr>
<td>Read</td>
<td>141</td>
<td>Drew</td>
<td>176</td>
<td>Other</td>
<td>220</td>
</tr>
<tr>
<td>Right</td>
<td>142</td>
<td>Drawing</td>
<td>177</td>
<td>Nothing</td>
<td>221</td>
</tr>
<tr>
<td>Raster</td>
<td>143</td>
<td>Brawling</td>
<td>178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rather</td>
<td>144</td>
<td>True</td>
<td>179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rort</td>
<td>145</td>
<td>Through</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rude</td>
<td>146</td>
<td>Splayed</td>
<td>181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>147</td>
<td>Splendid</td>
<td>182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>148</td>
<td>Spry</td>
<td>183</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mush</td>
<td>149</td>
<td>Spraying</td>
<td>184</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spring</td>
<td>185</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sprung</td>
<td>186</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sprawl</td>
<td>187</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spleen</td>
<td>188</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strayed</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Street</td>
<td>190</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Struck</td>
<td>191</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stroke</td>
<td>192</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Soldier</td>
<td>193</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dodger</td>
<td>194</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Error Analysis on Articulatory Models in Speech Configurations

B.1 Error Analysis of the Tongue Models

Figure B.1: The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /s/.
Figure B.2: The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /ʃ/.

Figure B.3: The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /iː/.
Figure B.4: The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /aː/. 

Figure B.5: The comparison between the imaged (sustained articulation in MRI) and the simulated tongue shape using the data derived from EMA in phoneme /ɔː/. 

B.2 Error Analysis of the Cheek-Lips Models
Figure B.6: The comparison between the image and the simulated lips shape using the data derived from video cameras in phoneme /s/.

Figure B.7: The comparison between the image and the simulated lips shape using the data derived from video cameras in phoneme /ʃ/. 
Figure B.8: The comparison between the image and the simulated lips shape using the data derived from video cameras in phoneme /iː/. 

Figure B.9: The comparison between the image and the simulated lips shape using the data derived from video cameras in phoneme /aː/. 
Figure B.10: The comparison between the image and the simulated lips shape using the data derived from video cameras in phoneme /ɔː/. 
Appendix C

Segmented Vocal Tract Surface for Six Configurations

Figure C.1: Segmentation of the vocal tract surface (in dots) in the rest breathing state by 1×1×1cm uniform cells (in black). (a) The front view. (b) The side view.
Figure C.2: Segmentation of the vocal tract surface (in dots) in the fricative /s/ by 0.8×0.8×0.8cm uniform cells (in black). (a) The front view. (b) The side view.

Figure C.3: Segmentation of the vocal tract surface (in dots) in the fricative /ʃ/ by 0.75×0.75×0.75cm uniform cells (in black). (a) The front view. (b) The side view.
Figure C.4: Segmentation of the vocal tract surface (in dots) in the vowel /i:/ by 1×1×1cm uniform cells (in black). (a) The front view. (b) The side view.

Figure C.5: Segmentation of the vocal tract surface (in dots) in the vowel /a:/ by 1×1×1cm uniform cells (in black). (a) The front view. (b) The side view.
Figure C.6: Segmentation of the vocal tract surface (in yellow) in the vowel /ɔː/ by $1 \times 1 \times 1$ cm uniform cells (in black). (a) The front view. (b) The side view.
Appendix D

User Guide for AcousticParallel

D.1 Compile the Program

AcousticParallel requires MPICH (Message-Passing Interface Chameleon) library [197], FORTRAN90 compiler and MUMPS (Multifrontal Massively Parallel Solver) numerical library [198] for the compilation and execution. There are two parts of program: AcousticSource and AcousticSolve. The former performs the evaluation of the acoustic loads while AcousticSolve performs the numerical procedures for solving inhomogeneous wave equations.

D.2 Running the Program

Both packages can be executed from the command line with flags to specify the function. For instance,

```
mpiexec -n 4 ./AcousticParallel.out -LIGHTHILL lighthill.P;
mpiexec -n 4 ./AcousticParallel.out -ACOUSTIC acoustic.P;
```

where `-n` indicates the number of processes; Header `-LIGHTHILL` specifies a job for AcousticSource while `-ACOUSTIC` is a flag for AcousticSolve. File `*.P` is a master script for defining the problem. The code of AcousticParallel can be divided into five major modules of I/O, FE, Boundary, Matrices and Solvers which can be explained by their internal and external commands. In the following sections, subroutines which has been parallelised, are denoted by `*`.

D.3 IO Module

Many of the IO (input/output) functions are shared between AcousticSource and AcousticSolve. As presented in Section D.3, the IO part of the code can be categorised into five parts: problem definition; FE domain; problem parameters, boundary conditions and solution output. There
are three problem types (‘!Problem-type’) in AcousticSource (‘1’ for Reynolds stress model; ‘2’ for Lagrangian multiplier model and ‘3’ for Reynolds and viscous stress model) and two types in AcousticSolve (‘1’ for Temporal solver and ‘2’ for Harmonic solver).

External Commands

All the following I/O commands are defined in the master script (.P file) and start with the symbol ‘!’. All the numerical schemes are specified in the numerical file (.num) where a list of options are provided in Section D.4. The node file (.X) contains all the node-based values for FE fields, including the initial condition for any acoustic simulation, while topology of the mesh is provided in the element file (.T). In the case of a Lighthill problem, a list of .X files should be provided in a list file (.node). The boundary topology is defined in the .B file and a list of file names for the output should be provided in .out file.

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>!Problem-type</td>
<td>PROBLEM TYPES</td>
</tr>
<tr>
<td>!NUM-file</td>
<td>NUMERICAL FILE</td>
</tr>
<tr>
<td>!Mesh-nodes-list</td>
<td>LIST OF NODE FILES</td>
</tr>
<tr>
<td>!Mesh-nodes</td>
<td>NODE FILE</td>
</tr>
<tr>
<td>!Mesh-elements</td>
<td>ELEMENT FILE</td>
</tr>
<tr>
<td>!Mesh-boundaries</td>
<td>BOUNDARY FILE</td>
</tr>
<tr>
<td>!Initial-density</td>
<td>INITIAL SOLUTION FILE</td>
</tr>
<tr>
<td>!Outfile</td>
<td>LIST OF OUTPUT FILES</td>
</tr>
<tr>
<td>!Viscosity</td>
<td>VISCOSITY OF FLUID</td>
</tr>
<tr>
<td>!Density</td>
<td>DENSITY OF FLUID</td>
</tr>
<tr>
<td>!Sound-velocity</td>
<td>SPEED OF SOUND</td>
</tr>
<tr>
<td>!Sample-time</td>
<td>TIME/FREQUENCY STEP</td>
</tr>
<tr>
<td>!Start-iteration</td>
<td>FIRST TIME/FREQUENCY</td>
</tr>
<tr>
<td>!Time-step</td>
<td>LAST TIME/FREQUENCY</td>
</tr>
</tbody>
</table>

Internal Subroutines

In the current code, the I/O functions are not parallelised, thus each processor is directed to read and save a copy of the same input files. Both AcousticSource and AcousticSolve share the
same I/O subroutines.

\begin{verbatim}
READ_PROBLEM     READ AND DEFINE PROBLEM TYPES
READ_FEDOMAIN    READ MESH NODES AND TOPOLOGY
READ_PARAMETERS  READ PROBLEM PARAMETERS
READ_BCONDITIONS READ BOUNDARY CONDITIONS
PRINT_OUTPUT     OUTPUT FILES
INTERIM_UPDATE   UPDATE BETWEEN TIME STEPS
\end{verbatim}

D.4 FE Module

The FE module builds user defined Finite Element fields, including assembly of basis functions and evaluating Gaussian quadratures.

External Commands

All external commands related to the FE modules, shared by both AcousticSource and AcousticSolve, are defined in .num files.

\begin{verbatim}
domain_dimension!  DOMAIN DIMENSION
   no_fields!       NUMBER OF FIELDS
   call_letters!    FIELD NAME
   no_ele_faces!    NUMBER OF FACES PER ELEMENT
   no_ele_nodes_f!  NUMBER OF NODES PER FACE
   hexa_basis!      HEXAGONAL BASIS FUNCTIONS
   dim_field_*!     DIMENSION OF FIELDS
   no_basis_*!      NUMBER OF BASIS PER FIELD
   *_basis!         COEFFICIENTS FOR BASIS FUNCTIONS
   *_P_basis!       POWER COEFFICIENTS FOR BASIS FUNCTIONS
   *_xi_coordinates! XI COORDINATES FOR BASIS FUNCTIONS
   no_gauss!        NUMBER OF GAUSS POINTS
   gauss_points!    GAUSSIAN POINTS
   gauss_weights!   GAUSSIAN WEIGHTS
   volume!          VOLUME OF EACH ELEMENT
   surface_area!    SURFACE AREA OF EACH FACE
   basis_ordering_*! ORDERS OF BASIS FUNCTIONS
\end{verbatim}
Internal Subroutines
There are two main subroutines in the FE modules, shared by both AcousticSource and AcousticSolve.

- LOAD_BASE_SET BUILD BASIS FUNCTIONS
- LOAD_INTEGRALS EVALUATE GAUSSIAN QUADRATURES

D.5 Boundary Module
The boundary module is only implemented in AcousticSolve. There are three types of acoustic boundary conditions implemented in the current program. Each processor has a copy of the boundary conditions of the global domain.

External Commands
The boundary conditions are defined in the master script .P, using tags ‘!Boundary-number’ and ‘!Boundary-type’. The topology of the boundary patch should be provided in the file .B, read under the tag ‘!Mesh-boundaries’.

- !Boundary-number NUMBER OF BOUNDARY CONDITIONS
- !Boundary-type TYPES OF BOUNDARY CONDITIONS
  - Dirichlet_G DIRICHLET CONDITION
  - NormalStress_G NEUMANN CONDITION
  - Absorb ABSORB CONDITION

Internal Subroutines
The module contains both the spatial and temporal boundary conditions, including providing initial guess for the selected temporal schemes at the beginning of the simulation. Subroutine ‘PRE_BOUND2’ directs calls to individual functions for specifying various types of boundary
conditions, according to user specified orders.

BUILD_INITIAL_GUESS  BUILD INITIAL GUESS
PRE_BOUND2 MAIN SUBROUTINE
BOUNDARY_DIRICHLET BUILD DIRICHLET CONDITIONS
BOUNDARY_STRESS BUILD NEUMANN CONDITIONS
BOUNDARY_ABSORB BUILD ABSORB CONDITIONS

D.6 Matrix Module

The matrix module in AcousticSource is used to build nodal acoustic loads for the source model defined in ‘Problem-type’.

External Commands

There are no external commands for AcousticSource in the matrix assembly module as the types of source models are defined in ‘!Problem-type’. Flag ‘!LIGHTHILL’ specifies a Lighthill formulation for AcousticSolve. There are four different temporal schemes implemented in AcousticSolve.

!LIGHTHILL SPECIFY LIGHTHILL PROBLEM
!IMPLICIT 2ND – ORDER IMPLICIT SCHEME
!EXPLICIT 2ND – ORDER EXPLICIT SCHEME
!PREDICTOR-CORRECTOR 2ND – ORDER NEWMARK SCHEME
!HOUBOLT 4TH – ORDER HOUBOLT SCHEME

Internal Subroutines

The matrix part of the code for both AcousticSource and AcousticSolve are parallelised via MPICH2 commands and achieved by dividing total number of mesh elements equally over all
the processors.

*SET_MATRIX (AcousticSource & AcousticSolve) SET UP MATRIX STRUCTURES
*ASSEMBLE_LIGHT (AcousticSource) ASSEMBLE LIGHTHILL NODAL SOURCE
*PRE_SOURCE (AcousticSource) CALCULATE FE FIELDS
*LIGHTHILL_T (AcousticSource) CALCULATE NODAL ACOUSTIC SOURCE
*LIGHTHILL_V (AcousticSource) CALCULATE NODAL ACOUSTIC LOADS
*ASSEMBLE_LEFT (AcousticSolve) ASSEMBLE LEFT SIDE MATRIX
*ASSEMBLE_RIGHT (AcousticSolve) ASSEMBLE RIGHT SIDE VECTOR
*LAPLACIAN_R (AcousticSolve) BUILD LAPLACIAN OPERATOR
*TIME_R (AcousticSolve) BUILD TEMPORAL DERIVATIVES
*PRE_LIGHTHILL (AcousticSolve) CALCULATE FE FIELDS
*BUILD_SOLVE_MATRIX (AcousticSolve) BUILD COEFFICIENT MATRIX

D.7 Solver Module

Only AcousticSolve contains a solver module for solving the system of linear equations.

External Commands

For modelling Machs number effect, the wave operator needs to be updated according to the flow parameters.

!LU SPECIFY LU SOLVER
!MUMPS SPECIFY MUMPS SOLVER
!MACH-EFFECT ADD MACH EFFECT

Internal Subroutines

The serial LU decomposition method is implemented to solve small systems of equations, based on the code provided in [204]. MUMPS numerical solver is used to solve large systems with
distributed matrices.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>*SOLVE_A</td>
<td>MAIN SOLVER ROUTINE</td>
</tr>
<tr>
<td>*UPDATE_LU_RHS</td>
<td>UPDATE RIGHT HAND SIDE VECTOR</td>
</tr>
<tr>
<td>*SOLVE_MUMPS</td>
<td>SOLVING ROUTINE WITH MUMPS</td>
</tr>
<tr>
<td>PRE LU</td>
<td>BUILD LU MATRIX</td>
</tr>
<tr>
<td>LU_DECOMPOSITION</td>
<td>LU DECOMPOSITION</td>
</tr>
<tr>
<td>LU_SOLVE</td>
<td>LU INVERSION</td>
</tr>
</tbody>
</table>
Appendix E

ALE Form of The Lighthill Equation

According to [199], the ALE forms of the compressible Navier-Stokes equations may be written as

\[
\frac{\partial \rho}{\partial t} + U_i \frac{\partial \rho}{\partial x_i} + \rho \frac{\partial u_i}{\partial t} = 0, \tag{E.0.1}
\]

\[
\rho \frac{\partial u_i}{\partial t} + \rho U_j \frac{\partial u_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial \sigma_{ij}}{\partial x_j}, \tag{E.0.2}
\]

where \( U_i = u_i - \hat{u}_i \), which is the velocity of the fluid measured at the referential state and \( \hat{u}_i \) is the velocity of the mesh, measured at a referential state. The convection velocity in ALE form may be interpreted as the velocity of the fluid measured with respect to the mesh.

Now, multiply \( u_i \) to Equation E.0.1 and add this to the momentum Equation E.0.2, the new form emerges as

\[
\frac{\partial (\rho u_i)}{\partial t} = -\frac{\partial (\rho u_i U_j)}{\partial x_j} - \frac{\partial p}{\partial x_i} + \frac{\partial \sigma_{ij}}{\partial x_j} - \rho u_i \frac{\partial \hat{u}_j}{\partial x_j}. \tag{E.0.3}
\]

Differentiating Equation E.0.1 with respect to \( t \) and taking the divergence of above equation, the new equations become

\[
\frac{\partial^2 \rho}{\partial t^2} + \frac{\partial (\rho u_i)}{\partial x_i \partial t} - \hat{u}_i \frac{\partial \rho}{\partial x_i} - \rho \frac{\partial \hat{u}_i}{\partial t} = 0, \tag{E.0.4}
\]

\[
\frac{\partial (\rho u_i)}{\partial x_i \partial t} + \frac{\partial (\rho u_i U_j)}{\partial x_i x_j} + \frac{\partial^2 p}{\partial x_i^2} - \frac{\partial \sigma_{ij}}{\partial x_i} = -\frac{\partial p}{\partial x_j} u_j \frac{\partial \hat{u}_i}{\partial x_i} - \rho \frac{\partial u_i}{\partial x_j} \frac{\partial \hat{u}_j}{\partial x_i} - \rho u_j \frac{\partial^2 \hat{u}_i}{\partial x_i \partial x_j}. \tag{E.0.5}
\]

Subtracting Equation E.0.5 from Equation E.0.4 and subtract \( c_0^2 \nabla^2 \rho \) from the both sides, we reach

\[
\frac{\partial^2 \rho}{\partial t^2} - c_0^2 \frac{\partial^2 \rho}{\partial x_i^2} = \frac{\partial (\rho u_i U_j + p \delta_{ij} - c_0^2 \rho \delta_{ij} - \sigma_{ij})}{\partial x_i x_j} + \hat{u}_i \frac{\partial \rho}{\partial t} \frac{\partial \hat{u}_i}{\partial x_i} \tag{E.0.6}
\]

\[
+ \frac{\partial \rho}{\partial x_i} \frac{\partial \hat{u}_i}{\partial t} + u_j \frac{\partial \rho}{\partial x_j} \frac{\partial \hat{u}_i}{\partial x_i} + \rho \frac{\partial u_j}{\partial x_j} \frac{\partial \hat{u}_i}{\partial x_i} + \rho u_j \frac{\partial^2 \hat{u}_i}{\partial x_i \partial x_j}. \tag{E.0.7}
\]
As we can see from above, in a stagnated mesh where $\hat{u}_i = 0$, the above form returns the original Lighthill equation. Furthermore, if we assume the source field on the left hand side of the equation is nearly incompressible and $P - c_0^2 \rho' = 0$, then the above equation can be simplified into

$$\frac{\partial^2 \rho'}{\partial t^2} - c_0^2 \frac{\partial^2 \rho'}{\partial x_i^2} = \rho_0 \frac{\partial (u_i U_j - \sigma_{ij})}{\partial x_i x_j} + \rho_0 u_j \frac{\partial^2 \hat{u}_i}{\partial x_i x_j}$$  \hspace{1cm} (E.0.8)
Appendix F

New Formulation for Aeroacoustic Equations

F.1 Two-Equation Model for Modelling Flow-Acoustic Interactions

The aeroacoustic model introduced in Chapter 5 relies on assumptions that the compressible Lighthill stress is largely stored in the Lagrangian multiplier field in an incompressible formulation for low Mach number isentropic flows. Furthermore, there is no interaction between the resurgent acoustic waves and the mean flow. However, in some applications (e.g., whistling) such interactions may be important [49]. Here we take another approach to derive an acoustic formulation.

We first decompose the aerodynamic quantities into incompressible means ($\bar{u}, \bar{P}, \rho_0$) and compressible perturbations ($u', p', \rho'$) as

$$u = \bar{u} + u'; \quad p = \bar{P} + p'; \quad \rho = \rho_0 + \rho'.$$  \hspace{1cm} (F.1.1)

Following Lighthill’s derivation, we take the divergence of the Navier-Stokes equations and eliminate all terms involving only incompressible quantities and reach

$$\frac{\partial (\rho u_i)}{\partial x_i} + \frac{\partial^2 p'}{\partial x_i^2} = -\frac{\partial (\rho u_i u_j)}{\partial x_i \partial x_j} + \frac{\partial (\rho_0 \bar{u}_i \bar{u}_j)}{\partial x_i \partial x_j} + \nu \frac{\partial^3 u_i'}{\partial x_i \partial x_j^2}.$$  \hspace{1cm} (F.1.2)

We can also differentiate the continuity equation and write it in perturbation form as

$$\frac{\partial^2 \rho'}{\partial t^2} + \frac{\partial (\rho u_i)}{\partial x_i \partial t} = 0.$$  \hspace{1cm} (F.1.3)

Substitute $\partial (\rho u_i)/\partial x_i \partial t$ with $-\partial^2 \rho'/\partial t^2$ in above equations and applying the constitutive rela-
tion $p' = c^2 ho'$, we have
\[
\frac{\partial^2 p'}{\partial t^2} - \frac{\partial^2 c^2 \rho'}{\partial x_i \partial x_i} = \frac{\partial (pu_i u_j)}{\partial x_i \partial x_j} - \frac{\partial (\rho_0 \bar{u}_i \bar{u}_j)}{\partial x_i \partial x_j} - \nu \frac{\partial^2 u'_i}{\partial x_i \partial x_j \partial x_j}.
\] (F.1.4)

If we assume that the compressible component is much smaller than its incompressible counterpart, which is generally the case for low Mach number flows, we then can linearise the above equation into a new form
\[
\frac{\partial^2 p'}{\partial t^2} - c^2_0 \frac{\partial^2 \rho'}{\partial x_i^2} = \frac{\partial (\rho' \bar{u}_i \bar{u}_j)}{\partial x_i \partial x_j} + 2 \rho_0 \frac{\partial u'_i \bar{u}_j}{\partial x_j \partial x_i} - \nu \frac{\partial^3 u'_i}{\partial x_i \partial x_j \partial x_j^2}.
\] (F.1.5)

A further assumption may be made that the compressible component of the velocity field is, to leading order, irrotational (i.e. $u' = \vec{\nabla} \phi$) [217, 218]. To set the value of $c$ according to the incompressible flow values, we may form a system of two equations as
\[
\frac{\partial \rho'}{\partial t} + \bar{u}_i \frac{\partial \rho'}{\partial x_i} = -\rho_0 \frac{\partial^2 \phi}{\partial x_i^2},
\] (F.1.6)
\[
\frac{\partial^2 \rho'}{\partial t^2} - c^2_0 \frac{\partial^2 \rho'}{\partial x_i^2} - (\bar{u}_i \bar{u}_j) \frac{\partial^2 \rho'}{\partial x_i \partial x_j} - 2 \frac{\partial (\bar{u}_i \bar{u}_j)}{\partial x_j} \frac{\partial \rho'}{\partial x_i} - \frac{\partial (\bar{u}_i \bar{u}_j)}{\partial x_i \partial x_j} \rho' = 2 \rho_0 \frac{\partial (\vec{\nabla} \phi) \bar{u}_j}{\partial x_i \partial x_j} + \nu \frac{\partial^4 \phi}{\partial x_i^2 \partial x_j^2}.
\] (F.1.7)

We can see that the introduction of the velocity potential $\phi$ for the perturbation flow reduces the dimensionality of the problem. At the same time, as it increases the order of both equations by one, it means that we will have up to third-order spatial derivatives in the variational formulation. The method for calculating the third-order spatial derivatives are given in Appendix E.2.

On the other hand, we must note the underlying assumptions of this formulation which can only be applied to the source field (i.e. where the incompressible components are dominant), so that the original form of the Lighthill analogy may be reassembled with the updated source field in order to resolve the acoustic far-field. For the boundary conditions, we may use the reflective type $\partial \rho' / \partial n = 0$ for the solid walls and absorbing conditions for outward going waves.

\section*{F.2 Evaluation of the Third-order Directives}

The evaluation of the source model involving the Lagrangian multiplier requires the calculation of two 2D spatial derivatives which can be found in [219]. We need to derive an expression for the third-order spatial derivatives in the viscous source field. Here, we are only presenting the derivation for the first two third-order derivatives $\partial^3 \phi / \partial x^3$ and $\partial^3 \phi / \partial x \partial y^2$ as the process is similar for the other two.

214
The derivatives may be decomposed into a Taylor series expansion as
∂3φ
∂ 3 φ ∂ξ 3
∂ 3 φ ∂ξ 2 ∂η
∂ 2 φ ∂ξ ∂ 2 ξ
=
(
)
+
(
)
+
2
×
∂x3
∂ξ 3 ∂x
∂ξ 2 ∂η ∂x ∂x
∂ξ 2 ∂x ∂x2
∂ 3 φ ∂ξ ∂η
∂ 3 φ ∂ξ ∂η 2
+ 2 × 2 ( )2
+2×
( )
∂ξ η ∂x ∂x
∂ξ∂η 2 ∂x ∂x
∂ 3 φ ∂ξ ∂η 2
∂ 2 φ ∂ 2 ξ ∂η
+ 2×
(
)
+
2
×
∂ξ∂η 2 ∂x ∂x
∂ξ∂η ∂x2 ∂x
∂ 2 φ ∂ξ ∂ 2 η
∂ 3 φ ∂ξ ∂η 2 ∂ 3 φ ∂η 3
+ 2×
+
( ) + 3( )
∂ξ∂η ∂x ∂x2 ∂η 2 ∂ξ ∂x ∂x
∂η ∂x
2
2
2
2
2
∂ φ ∂η ∂ η ∂ φ ∂ ξ ∂ξ
∂ φ ∂ 2 ξ ∂η
+ 2× 2
+
+
∂η ∂x ∂x2 ∂ξ 2 ∂x2 ∂x ∂ξ∂η ∂x2 ∂x
∂ 2 φ ∂ 2 η ∂ξ ∂ 2 φ ∂ 2 η ∂η ∂φ ∂ 3 η
+
+
+
;
∂ξ∂η ∂x2 ∂x ∂η 2 ∂x2 ∂x ∂η ∂x3

∂ 3φ
∂ 3 φ ∂ξ ∂ξ 2
∂ 3 φ ∂η ∂ξ 2
∂ 2 φ ∂ξ ∂ 2 ξ
=
(
)
+
(
)
+
2
×
∂x∂y 2
∂ξ 3 ∂x ∂η
∂ξ 2 ∂η ∂x ∂y
∂ξ 2 ∂y ∂x∂y
∂ 3 φ ∂ξ ∂η ∂ξ
∂ 3 φ ∂η ∂η ∂ξ
∂ 2 φ ∂ 2 η ∂ξ
+
+
+
∂ξ 2 ∂η ∂x ∂y ∂y ∂ξ∂η 2 ∂x ∂y ∂y ∂ξ∂η ∂x∂y ∂y
∂ 2 φ ∂η ∂ 2 ξ
∂ 2 φ ∂ 2 ξ ∂ξ
∂ 2 φ ∂ 2 ξ ∂η
+
+ 2 2
+
∂ξ∂η ∂y ∂x∂y ∂ξ ∂y ∂x ∂ξ∂η ∂y 2 ∂x
∂φ ∂ 3 ξ
∂ 3 φ ∂ξ ∂ξ ∂η
∂ 3 φ ∂ξ ∂η ∂η
+
+
+
∂ξ ∂x∂y 2 ∂ξ 2 ∂η ∂x ∂y ∂y ∂ξ∂η 2 ∂y ∂x ∂y
∂ 3 φ ∂ξ ∂η 2 ∂ 3 u ∂η ∂η 2
∂ 2 φ ∂η ∂ 2 η
+
(
)
+
(
)
+
2
×
∂ξ∂η 2 ∂x ∂y
∂η 3 ∂x ∂y
∂η 2 ∂y ∂x∂y
∂ 2 φ ∂ξ ∂ 2 η
∂ 2 φ ∂ξ ∂ 2 η ∂u ∂ 3 η
+
+
+
.
∂ξ∂η ∂x ∂y 2 ∂ξ∂η ∂x ∂y 2 ∂η ∂x∂y 2

(F.2.1)

(F.2.2)

Following the convention of [219], we express x as a Taylor series of ξ and η as
4x(ξ, η) = a1 4 ξ + a2 4 η + a3 (4ξ)2 + a4 4 ξ 4 η + a5 (4η)2
+ a6 (4ξ)2 4 η + a7 4 ξ(4η)2 + a8 (4ξ)2 (4η)2
+ a9 (4ξ)3 + a10 (4η)3 ;

(F.2.3)

4y(ξ, η) = b1 4 ξ + b2 4 η + b3 (4ξ)2 + b4 4 ξ 4 η + b5 (4η)2
+ b6 (4ξ)2 4 η + b7 4 ξ(4η)2 + b8 (4ξ)2 (4η)2
+ b9 (4ξ)3 + b10 (4η)3 ,

215

(F.2.4)


where the coefficients are defined to be

\[
\begin{align*}
    a_1 &= \frac{\partial x}{\partial \xi}, a_2 = \frac{\partial x}{\partial \eta}, a_3 = \frac{1}{2} \frac{\partial x^2}{\partial \xi^2}, a_4 = \frac{\partial^2 x}{\partial \xi \partial \eta}, a_5 = \frac{1}{2} \frac{\partial^2 x}{\partial \eta^2}, \\
    a_6 &= \frac{1}{2} \frac{\partial^3 x}{\partial \xi^2 \partial \eta}, a_7 = \frac{1}{2} \frac{\partial^3 x}{\partial \xi \partial \eta^2}, a_8 = \frac{1}{4} \frac{\partial^4 x}{\partial \xi^2 \partial \eta^2}, a_9 = \frac{1}{6} \frac{\partial^4 x}{\partial \eta^3}, \\
    a_{10} &= \frac{1}{6} \frac{\partial^5 x}{\partial \eta^3};
\end{align*}
\]  

(F.2.5)

\[
\begin{align*}
    b_1 &= \frac{\partial y}{\partial \xi}, b_2 = \frac{\partial y}{\partial \eta}, b_3 = \frac{1}{2} \frac{\partial y^2}{\partial \xi^2}, b_4 = \frac{\partial^2 y}{\partial \xi \partial \eta}, b_5 = \frac{1}{2} \frac{\partial^2 y}{\partial \eta^2}, \\
    b_6 &= \frac{1}{2} \frac{\partial^3 y}{\partial \xi^2 \partial \eta}, b_7 = \frac{1}{2} \frac{\partial^3 y}{\partial \xi \partial \eta^2}, b_8 = \frac{1}{4} \frac{\partial^4 y}{\partial \xi^2 \partial \eta^2}, b_9 = \frac{1}{6} \frac{\partial^4 y}{\partial \eta^3}, \\
    b_{10} &= \frac{1}{6} \frac{\partial^5 y}{\partial \eta^3}.
\end{align*}
\]  

(F.2.6)

The local variables \( \xi \) and \( \eta \) can also be written in the form of a Taylor series

\[
\begin{align*}
    \triangle \xi(\triangle x, \triangle y) &= \alpha_1 \triangle x + \alpha_2 \triangle y + \alpha_3 (\triangle x)^2 \\
    &+ \alpha_4 \triangle x \triangle y + \alpha_5 (\triangle y)^2 + \alpha_6 (\triangle x)^3 \\
    &+ \alpha_7 (\triangle x)^2 \triangle y + \alpha_8 \triangle x (\triangle y)^2 + \alpha_9 (\triangle y)^3; \\
\end{align*}
\]  

(F.2.7)

\[
\begin{align*}
    \triangle \eta(\triangle x, \triangle y) &= \beta_1 \triangle x + \beta_2 \triangle y + \beta_3 (\triangle x)^2 \\
    &+ \beta_4 \triangle x \triangle y + \beta_5 (\triangle y)^2 + \beta_6 (\triangle x)^3 \\
    &+ \beta_7 (\triangle x)^2 \triangle y + \beta_8 \triangle x (\triangle y)^2 + \beta_9 (\triangle y)^3.
\end{align*}
\]  

(F.2.8)

where the coefficients are defined as

\[
\begin{align*}
    \alpha_1 &= \frac{\partial \xi}{\partial x}, \alpha_2 = \frac{\partial \xi}{\partial y}, \alpha_3 = \frac{1}{2} \frac{\partial^2 \xi}{\partial x^2}, \alpha_4 = \frac{\partial^2 \xi}{\partial x \partial y}, \alpha_5 = \frac{1}{2} \frac{\partial^2 \xi}{\partial y^2}, \\
    \alpha_6 &= \frac{1}{6} \frac{\partial^3 \xi}{\partial x^3}, \alpha_7 = \frac{1}{2} \frac{\partial^3 \xi}{\partial x \partial y^2}, \alpha_8 = \frac{1}{2} \frac{\partial^3 \xi}{\partial x^2 \partial y}, \alpha_9 = \frac{1}{6} \frac{\partial^3 \xi}{\partial y^3}; \\
\end{align*}
\]  

(F.2.9)

\[
\begin{align*}
    \beta_1 &= \frac{\partial \eta}{\partial x}, \beta_2 = \frac{\partial \eta}{\partial y}, \beta_3 = \frac{1}{2} \frac{\partial^2 \eta}{\partial x^2}, \beta_4 = \frac{\partial^2 \eta}{\partial x \partial y}, \beta_5 = \frac{1}{2} \frac{\partial^2 \eta}{\partial y^2}, \\
    \beta_6 &= \frac{1}{6} \frac{\partial^3 \eta}{\partial x^3}, \beta_7 = \frac{1}{2} \frac{\partial^3 \eta}{\partial x^2 \partial y}, \beta_8 = \frac{1}{2} \frac{\partial^3 \eta}{\partial x \partial y^2}, \beta_9 = \frac{1}{6} \frac{\partial^3 \eta}{\partial y^3}.
\end{align*}
\]  

(F.2.10)

After substituting Equations 48 and 49 into Equations 44 and 45 respectively, we can obtain
Finally, we derive the coefficients for $\alpha_i, \alpha_j, \beta_i, \beta_j$

$$\alpha_1, \alpha_2, \beta_1, \beta_2$$

$$\alpha_1 = \frac{b_2}{J}, \alpha_2 = -\frac{a_2}{J}, \beta_1 = -\frac{b_1}{J}, \beta_2 = \frac{a_1}{J}, \quad (F.2.11)$$

where $J$ is the determinant of the Jacobian matrix

$$J = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}. \quad (F.2.12)$$

Following this, we further derive coefficients for the second derivative terms,

$$\alpha_3 = \frac{(b_3\alpha_1^2 + b_4\alpha_1\beta_1 + b_5\beta_1^2)b_2 - (a_3\alpha_1^2 + a_4\alpha_1\beta_1 + a_5\beta_1^2)b_1}{J}, \quad (F.2.13)$$

$$\beta_3 = \frac{(a_3\alpha_1^2 + a_4\alpha_1\beta_1 + a_5\beta_1^2)b_1 - (b_3\alpha_1^2 + b_4\alpha_1\beta_1 + b_5\beta_1^2)a_1}{J}, \quad (F.2.14)$$

$$\alpha_4 = \frac{(2b_3\alpha_1\alpha_2 + b_4\alpha_2\beta_1 + b_4\alpha_2\beta_1 + 2b_5\alpha_1\beta_2)a_2}{J} - \frac{(2a_3\alpha_1\alpha_2 + a_4\alpha_1\beta_2 + a_4\alpha_2\beta_1 + 2a_5\alpha_1\beta_2)b_1}{J}, \quad (F.2.15)$$

$$\beta_4 = \frac{(2a_3\alpha_1\alpha_2 + a_4\alpha_1\beta_2 + a_4\alpha_2\beta_1 + 2a_5\alpha_1\beta_2)a_1}{J} - \frac{(2b_3\alpha_1\alpha_2 + b_4\alpha_1\beta_2 + b_4\alpha_2\beta_1 + 2b_5\alpha_1\beta_2)a_1}{J}, \quad (F.2.16)$$

$$\alpha_5 = \frac{(b_3\alpha_2^2 + b_4\alpha_2\beta_2 + b_5\beta_2^2)a_2 - (a_3\alpha_2^2 + a_4\alpha_2\beta_2 + a_5\beta_2^2)b_2}{J}, \quad (F.2.17)$$

$$\beta_5 = \frac{(a_3\alpha_2^2 + a_4\alpha_2\beta_2 + a_5\beta_2^2)b_1 - (b_3\alpha_2^2 + b_4\alpha_2\beta_2 + b_5\beta_2^2)a_1}{J}, \quad (F.2.18)$$

Finally, we derive the coefficients for $\alpha_6, \alpha_7, \alpha_8, \alpha_9$ and $\beta_6, \beta_7, \beta_8, \beta_9$

$$\alpha_6 = \frac{(2b_3\alpha_1\alpha_3 + b_4\alpha_1\beta_3 + b_4\alpha_3\beta_1 + 2b_5\alpha_1\beta_3)a_2}{J} + \frac{(b_6\alpha_1^2\beta_1 + b_7\alpha_1\beta_1^2 + b_6\alpha_1^3 + b_10\beta_1^3)a_2}{J} - \frac{(2a_3\alpha_1\alpha_3 + a_4\alpha_1\beta_3 + a_4\alpha_3\beta_1 + 2a_5\alpha_1\beta_3)b_2}{J} + \frac{(a_6\alpha_1^2\beta_1 + a_7\alpha_1\beta_1^2 + a_9\alpha_1^2 + a_10\beta_1^2)b_2}{J}, \quad (F.2.19)$$

$$\beta_6 = \frac{(2a_3\alpha_1\alpha_3 + a_4\alpha_1\beta_3 + a_4\alpha_3\beta_1 + 2a_5\alpha_1\beta_3)b_1}{J} + \frac{(a_6\alpha_1^2\beta_1 + a_7\alpha_1\beta_1^2 + a_9\alpha_1^2 + a_10\beta_1^2)b_1}{J} - \frac{(2b_3\alpha_1\alpha_3 + b_4\alpha_1\beta_3 + b_4\alpha_3\beta_1 + 2b_5\alpha_1\beta_3)a_1}{J} + \frac{(b_6\alpha_1^2\beta_1 + b_7\alpha_1\beta_1^2 + b_9\alpha_1^2 + b_10\beta_1^2)a_1}{J}, \quad (F.2.20)$$
\[
\alpha_7 = \frac{b_3 \alpha_1 a_4 + b_3 \alpha_2 a_3 + b_4 \alpha_1 \beta_4 + b_4 \alpha_2 \beta_3 + b_4 \alpha_3 \beta_2) a_2 \beta}{J} + \frac{(b_4 \alpha_2 \beta_1 + b_5 \beta_1 \beta_4 + b_5 \beta_2 \beta_3 + b_5 \alpha_1^2 \beta_2 + b_7 \alpha_2 \beta_1^2) a_2}{J} - \frac{(a_3 \alpha_1 a_4 + a_3 \alpha_2 a_3 + a_4 \alpha_1 \beta_4 + a_4 \alpha_2 \beta_3 + a_4 \alpha_3 \beta_2) b_2}{J} + \frac{(a_4 \alpha_1 \beta_4 + a_5 \beta_1 \beta_4 + a_5 \beta_2 \beta_3 + a_6 \alpha_1^2 \beta_2 + a_7 \alpha_1 \beta_1^2) b_2}{J},
\]
\[
\beta_7 = \frac{(a_3 \alpha_1 a_4 + a_3 \alpha_2 a_3 + a_4 \alpha_1 \beta_4 + a_4 \alpha_2 \beta_3 + a_4 \alpha_3 \beta_2) a_1}{J} + \frac{(a_4 \alpha_1 \beta_4 + a_5 \beta_1 \beta_4 + a_5 \beta_2 \beta_3 + a_6 \alpha_1^2 \beta_2 + a_7 \alpha_2 \beta_1^2) b_1}{J} - \frac{(b_3 \alpha_1 a_4 + b_3 \alpha_2 a_3 + b_4 \alpha_1 \beta_4 + b_4 \alpha_2 \beta_3 + b_4 \alpha_3 \beta_2) a_1}{J} + \frac{(b_4 \alpha_2 \beta_1 + b_5 \beta_1 \beta_4 + b_5 \beta_2 \beta_3 + b_6 \alpha_1^2 \beta_2 + b_7 \alpha_2 \beta_1^2) a_1}{J},
\]
\[
\alpha_8 = \frac{(b_3 \alpha_1 a_5 + b_3 \alpha_2 a_4 + b_4 \alpha_1 \beta_5 + b_4 \alpha_2 \beta_4 + b_4 \alpha_4 \beta_2) a_2 \beta}{J} + \frac{(b_4 \alpha_5 \beta_1 + b_5 \beta_1 \beta_5 + b_5 \beta_2 \beta_4 + b_5 \alpha_1^2 \beta_1 + b_7 \alpha_2 \beta_1^2) a_2}{J} - \frac{(a_3 \alpha_1 a_5 + a_3 \alpha_2 a_4 + a_4 \alpha_1 \beta_5 + a_4 \alpha_2 \beta_4 + a_4 \alpha_4 \beta_2) b_2}{J} + \frac{(a_4 \alpha_5 \beta_1 + a_5 \beta_1 \beta_5 + a_5 \beta_2 \beta_4 + a_6 \alpha_2^2 \beta_1 + a_7 \alpha_2 \beta_1^2) b_2}{J},
\]
\[
\beta_8 = \frac{(a_3 \alpha_1 a_5 + a_3 \alpha_2 a_4 + a_4 \alpha_1 \beta_5 + a_4 \alpha_2 \beta_4 + a_4 \alpha_4 \beta_2) a_1}{J} + \frac{(a_4 \alpha_5 \beta_1 + a_5 \beta_1 \beta_5 + a_5 \beta_2 \beta_4 + a_6 \alpha_2^2 \beta_1 + a_7 \alpha_2 \beta_1^2) b_1}{J} - \frac{(b_3 \alpha_1 a_5 + b_3 \alpha_2 a_4 + b_4 \alpha_1 \beta_5 + b_4 \alpha_2 \beta_4 + b_4 \alpha_4 \beta_2) a_1}{J} + \frac{(b_4 \alpha_5 \beta_1 + b_5 \beta_1 \beta_5 + b_5 \beta_2 \beta_4 + b_6 \alpha_2^2 \beta_1 + b_7 \alpha_2 \beta_1^2) a_1}{J},
\]

\text{(F.2.21)}
\text{(F.2.22)}
\text{(F.2.23)}
\text{(F.2.24)}
\[\alpha_9 = \frac{(2b_3a_2a_5 + b_4a_2b_5 + b_4a_5b_2 + 2b_5b_2) a_2}{J} + \frac{(b_6a_2^2b_2 + b_7a_2b_2^2 + b_9a_2^3 + b_9b_2^3) a_2}{J} - \frac{(2a_3a_2a_5 + a_4a_2b_5 + a_4a_5b_2 + 2a_5b_2b_5) b_2}{J} + \frac{(a_6a_2^2b_2 + a_7a_2b_2^2 + a_9a_2^3 + a_9b_2^3) b_2}{J}, \quad (F.2.25)\]

\[\beta_9 = \frac{(2a_3a_2a_5 + a_4a_2b_5 + a_4a_5b_2 + 2a_5b_2b_5) b_1}{J} + \frac{(a_6a_2^2b_2 + a_7a_2b_2^2 + a_9a_2^3 + a_9b_2^3) b_1}{J} - \frac{(2b_3a_2a_5 + b_4a_2b_5 + b_4a_5b_2 + 2b_5b_2b_5) a_1}{J} + \frac{(b_6a_2^2b_2 + b_7a_2b_2^2 + b_9a_2^3 + b_9b_2^3) a_1}{J}. \quad (F.2.26)\]
Bibliography


236


[253] J Cisonni, S Wada, K Nozaki, and A Hirtum. Towards sibilant [s] physical modeling: numerical study of the influence of the aperture of a tooth-shaped constriction on the


