Identifying the Site of Upper Airway Collapse in Obstructive Sleep Apnoea Patients Using Snore Signals

Arun Sebastian
Faculty of Engineering
The University of Sydney

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Sydney 2021
THIS THESIS IS DEDICATED
TO MY PARENTS
TO MY WIFE
TO MY KID
TO MY BROTHER AND SISTERS
WITH ALL MY LOVE
Statement of originality

This thesis is submitted to the University of Sydney in fulfilment of the requirement for the degree of Doctor of Philosophy. This thesis has not been submitted for any degree or other purposes. This is to certify that to the best of my knowledge, the content of this thesis is my own work.
I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Arun Sebastian
Acknowledgements

This thesis would not have been possible without the incredible support from many people, who have been with me over the past few years as I was working on my PhD program in Sydney.

Firstly, I would like to thank my supervisors Prof Philip de Chazal and Prof Peter Cistulli, for their enormous and unending support, patience, motivation, and comprehensive knowledge. Without their valuable guidance, this thesis would never have finished. Working with Philip helped me to keep my morale and confidence so high that it helped me to finish my PhD. I could not have imagined having a better advisor and mentor for my PhD project. I would like to thank Gary Cohen for his help in collecting the data and valuable suggestions.

I would like to thank all my friends at the University of Sydney, especially Dr. Nadi Sadr, Xian Long, Nadim and all the amazing members of the SLEEP research group. I thank all my friends, especially Manu and Arun, who have kept encouraging and supporting me throughout the journey and beyond. Also, I would like to thank Johanna from The Expert Editor for proofreading.

Most of all, I am grateful for the love, encouragement and support from my parents, Sebastian and Thresiamma, my brother, Amal, and my sisters, Dona and Rani, throughout my entire life.

Last but not least, I would like to thank my loving wife and my best friend, Daliya, and my loving son, Frederic, for their enormous love, support, encouragement and believing in me.
Abstract

This study aimed to determine the site-of-collapse in the upper airway using a simple and non-invasive method by processing the audio signal recorded during natural sleep. The information regarding the site-of-collapse could assist clinicians in choosing the most appropriate treatment for obstructive sleep apnoea (OSA) patients.

OSA is a highly prevalent and heterogeneous sleep-related breathing disorder characterised by repetitive partial or complete obstruction of the upper airway. OSA can lead to adverse consequences for health, quality of life and cognitive functioning. Conventional treatments for OSA include positive airway pressure devices, oral appliance (OA) therapy, and surgery. The choice of therapy is guided by the severity of OSA and patient preference. Even though different treatments are available, efficacy is highly variable between patients, creating the need to predict treatment outcome. Studies have shown that information regarding the site of airway obstruction may play a role in predicting treatment outcome and therefore assist clinicians in choosing the most appropriate treatment. This is especially true for patients who have “tongue-base” airway collapse, as they appear to be more likely to gain a large therapeutic benefit from OAs. Conventional methods in determining the site-of-collapse involve the use of an endoscope or a pressure catheter during drug-induced or natural sleep. Unfortunately, these methods are not well tolerated by patients due to their invasive nature, time/expense of the tests, and inconsistency of the obstruction site identified with natural sleep and drug-induced sleep, which in turn limit their clinical application. Therefore, the crux in identifying the site-of-collapse is to find a simple, non-invasive method with minimal impact on patients during natural sleep.

Snoring is a clinical hallmark of OSA and is caused by the vibration of different tissues due to the narrowing of the upper airway as the turbulent air passes through. Previous studies have shown that acoustic analysis of snoring can assist in the diagnosis of OSA and in estimating OSA severity. This thesis explores the acoustic properties of snoring sounds and their association with the site-of-collapse in the upper airway for OSA patients. It aimed to predict the site-of-upper airway collapse using audio signals recorded during natural sleep. The signals are processed with a system developed using machine learning. In particular, the proposed method is based on an automated and unsupervised machine learning algorithm.
to identify the predominant site-of-collapse during hypopnoea events (incomplete airway obstruction resulting in arousals or oxygen desaturation).

First, an audio database was generated using the nocturnal audio recordings of 58 OSA patients who attended a full night sleep study using polysomnography for OSA diagnosis. The audio signal was synchronised with polysomnography data. A spectral subtraction method was deployed to remove background noise from the audio recordings to improve the signal-to-noise ratio. The probable site-of-collapse was determined by manual analysis of the shape of the airflow signal during hypopnoea events, using a published method that had recently been shown to correlate well with the site-of-collapse. To validate the reliability of labelling, it was performed twice with the help of an experienced sleep physician. Simple snore (other than snoring during hypopnoeas) and other sound labelling were also conducted based on visual and auditory inspection of the audio recordings.

Second, an automatic machine learning algorithm using a Linear Discriminant Analysis (LDA) classifier was deployed to extract OSA related snore events from the nocturnal audio recordings. Results indicated that the algorithm achieved an accuracy of 87% for identifying snore events from the audio recordings and an accuracy of 72% for identifying OSA related snore events from the snore events.

Third, a novel approach for identifying the site-of-collapse in OSA patients was developed using automated classifiers. A model using LDA and a model using a Gaussian Mixture Model (GMM) were developed and the performance of the two models was compared. The model predicted the predominant site-of-collapse in the upper airway of OSA patients into four classes (“lateral wall”, “palate”, “tongue-base” related collapse or “multi-level” site-of-collapse) using the snore data recorded during hypopnoea events. An unbiased nested cross-validation method was deployed to select the high performing features for the classifiers. To estimate the performance of the classifiers, the data was divided into two sets. The “Learning Set” contained data from the first 45 patients and was used to develop the system. The second set (“Hidden Set”) contained data from the final 13 patients and was used to validate the final system. Results showed that the LDA classifier outperformed the GMM classifier, and the model achieved an overall accuracy on the Hidden Set of 85% for discriminating tongue and non-tongue collapse and an accuracy of 69% accuracy for all site-of-collapse classes.

Finally, an unsupervised machine learning algorithm was developed to form clusters using the features extracted from snore data during hypopnoea events and to analyse the correlation between the clusters generated and the site-of-collapse in the upper airway. If a correlation exists between clusters and the site-of-collapse, then this supports the role of the snore
signal as a correlate of the site-of-collapse. A $k$-means clustering analysis was developed using the most relevant features to show the tendency of the data to form clusters and to investigate the correlation between the site-of-collapse and the clusters generated. Two unsupervised feature selection algorithms using principal component analysis and a novel method combining silhouette analysis with the Laplacian score algorithm were developed. Internal cluster validation revealed that the data exhibits a good tendency in forming clusters, with the optimal cluster number of two and a mean silhouette value of 0.79. External cluster validation showed that the model achieved an accuracy of 68% in classifying the patients based on the predominant site-of-collapse (tongue/non-tongue).

In conclusion, an acoustic analysis of snoring can provide valuable information regarding the obstruction site in the upper airway. Therefore, snoring could potentially be used as an auxiliary method to identify the site-of-collapse and assist clinicians in deciding the most appropriate treatment in OSA patients. Further work is required to demonstrate that this method can be used to predict treatment response.
# Contents

Statement of originality ........................................ iii
Acknowledgement .............................................. iv
Abstract ................................................................... v
Abbreviations ........................................................ xii
List of Figures .......................................................... xiv
List of Tables ............................................................ xvi

## 1 Introduction

1.1 Motivation ......................................................... 1
1.2 Objectives and Outcomes ....................................... 2
1.3 Thesis Contributions ........................................... 3
1.4 Thesis Structure .................................................. 4
1.5 List of Publications ............................................. 6

## 2 Clinical Background

2.1 Sleep Apnoea ...................................................... 8
  2.1.1 Obstructive Sleep Apnoea .............................. 9
  2.1.2 Central Sleep Apnoea .................................. 9
  2.1.3 Mixed Sleep Apnoea .................................. 9
  2.1.4 Health Risk Associated with OSA ................. 11
  2.1.5 Prevalence of OSA .................................... 12
  2.1.6 Pathophysiology of OSA ............................ 12
  2.1.7 Diagnosis of OSA .................................... 13
  2.1.8 Upper Airway Physiology ......................... 16
2.2 Treatments for OSA ............................................ 18
  2.2.1 Continuous Positive Airway Pressure ............... 18
  2.2.2 Oral Appliance ....................................... 19
  2.2.3 Surgery .............................................. 21
  2.2.4 Life Style and Diet Treatment ....................... 21
2.3 Precision Medicine in OSA .................................. 21
2.4 Identification of Obstruction Sites ....................... 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.1 Fluoroscopy</td>
<td>23</td>
</tr>
<tr>
<td>2.4.2 Pressure Catheters</td>
<td>23</td>
</tr>
<tr>
<td>2.4.3 Computed Tomography Imaging</td>
<td>24</td>
</tr>
<tr>
<td>2.4.4 Magnetic Resonance Imaging</td>
<td>25</td>
</tr>
<tr>
<td>2.4.5 Drug-Induced Sleep Endoscopy</td>
<td>27</td>
</tr>
<tr>
<td>2.4.6 Airflow Shape Analysis</td>
<td>29</td>
</tr>
<tr>
<td>2.4.7 Summary of Identification of Obstruction Sites</td>
<td>31</td>
</tr>
<tr>
<td>2.5 Snoring Sound Characteristics</td>
<td>31</td>
</tr>
<tr>
<td>2.5.1 VOTE Classification</td>
<td>34</td>
</tr>
<tr>
<td>2.5.2 Summary of Snoring Sound Characteristics</td>
<td>35</td>
</tr>
<tr>
<td>2.6 Summary</td>
<td>36</td>
</tr>
<tr>
<td>3 Signal Processing and Machine Learning Methods</td>
<td>37</td>
</tr>
<tr>
<td>3.1 Feature Extraction</td>
<td>37</td>
</tr>
<tr>
<td>3.2 Machine Learning</td>
<td>42</td>
</tr>
<tr>
<td>3.2.1 Linear Discriminant Analysis Classifier</td>
<td>43</td>
</tr>
<tr>
<td>3.2.2 Gaussian Mixture Model Classifier</td>
<td>45</td>
</tr>
<tr>
<td>3.2.3 K-means Clustering</td>
<td>48</td>
</tr>
<tr>
<td>3.3 Nested Cross Validation</td>
<td>49</td>
</tr>
<tr>
<td>3.4 Feature Selection</td>
<td>50</td>
</tr>
<tr>
<td>3.4.1 Feature Selection using Nested Cross Validation</td>
<td>51</td>
</tr>
<tr>
<td>3.5 Summary</td>
<td>52</td>
</tr>
<tr>
<td>4 Data Base</td>
<td>53</td>
</tr>
<tr>
<td>4.1 Study Subjects</td>
<td>53</td>
</tr>
<tr>
<td>4.2 Data Acquisition</td>
<td>54</td>
</tr>
<tr>
<td>4.3 Alignment of Audio Signal with PSG</td>
<td>58</td>
</tr>
<tr>
<td>4.3.1 Finding the Delay in Audio Recording and PSG Signals</td>
<td>59</td>
</tr>
<tr>
<td>4.3.2 Synchronisation of PSG Signal and Audio</td>
<td>59</td>
</tr>
<tr>
<td>4.4 De-noising of Raw Data</td>
<td>61</td>
</tr>
<tr>
<td>4.4.1 Spectral Subtraction Method</td>
<td>61</td>
</tr>
<tr>
<td>4.4.2 Multi-band Spectral Subtraction Algorithm</td>
<td>63</td>
</tr>
<tr>
<td>4.4.3 Bandpass Filter Method</td>
<td>63</td>
</tr>
<tr>
<td>4.4.4 Comparison of De-noising Methods</td>
<td>64</td>
</tr>
<tr>
<td>4.5 Data Labelling</td>
<td>65</td>
</tr>
<tr>
<td>4.5.1 Labelling Site-of-Collapse</td>
<td>65</td>
</tr>
<tr>
<td>4.5.2 Labelling of Simple Snore and Non-Snore</td>
<td>71</td>
</tr>
<tr>
<td>4.6 Summary</td>
<td>72</td>
</tr>
<tr>
<td>5 Extraction of Snore Events From Sound Recordings</td>
<td>73</td>
</tr>
<tr>
<td>5.1 Existing Methods</td>
<td>74</td>
</tr>
<tr>
<td>5.2 Methodology</td>
<td>76</td>
</tr>
<tr>
<td>5.2.1 Study Subjects</td>
<td>76</td>
</tr>
<tr>
<td>5.2.2 Feature Extraction</td>
<td>76</td>
</tr>
<tr>
<td>5.2.3 Data Analysis</td>
<td>77</td>
</tr>
</tbody>
</table>
### 5.2.4 Data Partition
5.2.5 Feature Selection
5.2.6 Performance Evaluation

### 5.3 Results
5.3.1 Identification of Snore Events From Audio Recordings
5.3.2 Extraction of OSA Related Snoring From All Snore Events
5.3.3 Direct Method to Extract OSA Related Snore Events From Audio Recordings

### 5.4 Discussion

### 5.5 Summary

### 6 Automatic Classification of the Site of Airway Collapse in OSA patients
6.1 Existing Methods
6.2 Methodology
6.2.1 Study Subjects
6.2.2 Feature Extraction
6.2.3 Data Analysis
6.2.4 Data Partitioning
6.2.5 Feature Selection
6.2.6 Performance Evaluation

### 6.3 Results
6.3.1 Leave-One-Out Cross-Validation on Learning Set Using All Features
6.3.2 Leave-One-Out Nested Cross-Validation on Learning Set Using Feature Selection
6.3.3 Validation of the Best Model on the Hidden Set

### 6.4 Discussion

### 6.5 Summary

### 7 Cluster Analysis of OSA Related Snore Data
7.1 Existing Methods
7.2 Methodology
7.2.1 Study Subjects
7.2.2 Feature Extraction
7.2.3 Unsupervised Learning Approach
7.2.4 Feature Selection
7.2.5 Performance Evaluation

### 7.3 Results
7.3.1 Optimal Cluster Number
7.3.2 Feature Selection
7.3.3 Internal Cluster Validation
7.3.4 External Cluster Validation

### 7.4 Discussion

### 7.5 Summary
8 Conclusion and Future Work

8.1 Conclusion .......................................................... 131
8.2 Future Work Recommendations .............................. 133

Bibliography .......................... 136
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASM</td>
<td>American Academy of Sleep Medicine</td>
</tr>
<tr>
<td>AHI</td>
<td>Apnoea Hypopnoea Index</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>APAP</td>
<td>Autotitrating Positive Airway Pressure</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>BiPAP</td>
<td>Bi-level Positive Airway Pressure</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
</tr>
<tr>
<td>CPAP</td>
<td>Continuous Positive Airway Pressure</td>
</tr>
<tr>
<td>CRSWD</td>
<td>Circadian Rhythm Sleep-Wake Disorder</td>
</tr>
<tr>
<td>CSA</td>
<td>Central Sleep Apnoea</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>DISE</td>
<td>Drug Induced Sleep Endoscopy</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>EDF</td>
<td>European Data Format</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalogram</td>
</tr>
<tr>
<td>EMG</td>
<td>Electromyogram</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian Mixture Model</td>
</tr>
<tr>
<td>ICSD</td>
<td>International Classification of Sleep Disorders</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>LPC</td>
<td>Linear Predictive Coding</td>
</tr>
<tr>
<td>MAD</td>
<td>Mandibular Advancement Device</td>
</tr>
<tr>
<td>MFCC</td>
<td>Mel Frequency Cepstrum Coefficient</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>MSA</td>
<td>Mixed Sleep Apnoea</td>
</tr>
<tr>
<td>NREM</td>
<td>Non-Rapid Eye Movement</td>
</tr>
<tr>
<td>OA</td>
<td>Oral Appliance</td>
</tr>
<tr>
<td>OSA</td>
<td>Obstructive Sleep Apnea</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PPV</td>
<td>Positive Predictive Value</td>
</tr>
<tr>
<td>PSG</td>
<td>Polysomnography</td>
</tr>
<tr>
<td>REM</td>
<td>Rapid Eye Movement</td>
</tr>
<tr>
<td>SA</td>
<td>Sleep Apnoea</td>
</tr>
<tr>
<td>SBD</td>
<td>Sleep-related Breathing Disorders</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>TSD</td>
<td>Tongue Stabilizing Devices</td>
</tr>
<tr>
<td>UAR</td>
<td>Unweighted Average Recall</td>
</tr>
<tr>
<td>UPPP</td>
<td>Uvulopalatopharyngoplasty</td>
</tr>
<tr>
<td>ZCR</td>
<td>Zero Crossing Rate</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Normal breathing and obstructed breathing</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Numerous health risks related to OSA</td>
<td>11</td>
</tr>
<tr>
<td>2.3</td>
<td>Schematic representation of the typical physiological process during OSA and corresponding physiological processes</td>
<td>14</td>
</tr>
<tr>
<td>2.4</td>
<td>An illustration of polysomnography set-up and attachment of sensors</td>
<td>15</td>
</tr>
<tr>
<td>2.5</td>
<td>Anatomy and main muscles of the upper airway</td>
<td>17</td>
</tr>
<tr>
<td>2.6</td>
<td>Illustration of a CPAP setting</td>
<td>19</td>
</tr>
<tr>
<td>2.7</td>
<td>An example of oral appliance</td>
<td>20</td>
</tr>
<tr>
<td>2.8</td>
<td>An illustration of the placement of pressure catheter for the identification of airway obstruction</td>
<td>24</td>
</tr>
<tr>
<td>2.9</td>
<td>An example of CT and MRI scan of the upper airway</td>
<td>26</td>
</tr>
<tr>
<td>2.10</td>
<td>Drug-induced sleep endoscopy</td>
<td>27</td>
</tr>
<tr>
<td>2.11</td>
<td>Vibration of the vocal folds generates a varying airflow</td>
<td>32</td>
</tr>
<tr>
<td>2.12</td>
<td>Site-of-collapse positions of the VOTE classification in the upper airway</td>
<td>34</td>
</tr>
<tr>
<td>3.1</td>
<td>LDA classification for a sample dataset</td>
<td>44</td>
</tr>
<tr>
<td>3.2</td>
<td>GMM classification for a sample dataset</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>An example of k-means clustering</td>
<td>48</td>
</tr>
<tr>
<td>3.4</td>
<td>Data partitioning for a nested cross-validation approach</td>
<td>50</td>
</tr>
<tr>
<td>3.5</td>
<td>Block diagram representation of feature selection algorithm</td>
<td>52</td>
</tr>
<tr>
<td>4.1</td>
<td>Schematic diagram of PSG and audio acquisition system</td>
<td>55</td>
</tr>
<tr>
<td>4.2</td>
<td>Symbolic representation of various sensors and electrode placement in standard PSG recordings</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>Symbolic representation of thermistor and pressure transducer placement for the airflow recordings in PSG</td>
<td>57</td>
</tr>
<tr>
<td>4.4</td>
<td>Example of PSG recording screen using Compumedics Profusion Sleep Software</td>
<td>58</td>
</tr>
<tr>
<td>4.5</td>
<td>Loudness signal recorded as part of the PSG for the pilot test</td>
<td>60</td>
</tr>
<tr>
<td>4.6</td>
<td>Time difference between the PSG and audio signal throughout the pilot study recordings</td>
<td>61</td>
</tr>
<tr>
<td>4.7</td>
<td>A typical example of audio recording with noise (2-minute epoch) and results of noise reduction techniques.</td>
<td>64</td>
</tr>
<tr>
<td>4.8</td>
<td>Upper airway physiology</td>
<td>66</td>
</tr>
<tr>
<td>4.9</td>
<td>Identifying site-of-collapse using reference nasal pressure signals</td>
<td>67</td>
</tr>
</tbody>
</table>
# List of Tables

2.1 Comparison of obstructive sleep apnoea and central sleep apnoea . . . 10
2.2 Prevalence of OSA around the world. . . . . . . . . . . . . . . . . . . 13
2.3 Different criteria and updates provided by AASM for apnoea and hy-
popnoea. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

4.1 Participant characteristics details . . . . . . . . . . . . . . . . . . . 54
4.2 Reliability of labelling based on the labelling the data twice. . . . . 67
4.3 Participant characteristics details. . . . . . . . . . . . . . . . . . . . . 70

5.1 Confusion matrix for the performance assessment of a system. . . . . 79
5.2 Cross-validation results for snore and non-snore classification. . . . 80
5.3 Example of a feature selection process from the inner cross-validation. 81
5.4 Cross-validation results for the extraction of OSA related snoring. . . 82
5.5 Cross-validation results for the direct method to extract OSA related
snore events. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
5.6 Classification performance comparison for snore and non-snore events. 85

6.1 Participant characteristics details. . . . . . . . . . . . . . . . . . . . 90
6.2 Confusion matrix for the hypopnoea event based site-of-collapse clas-
sification using all features on the Learning Set (S1) . . . . . . . . . . 95
6.3 Confusion matrix for the predominant site-of-collapse using all features
on the Learning Set(S1) . . . . . . . . . . . . . . . . . . . . . . . . . . 95
6.4 Confusion matrix for the predominant site-of-collapse (non-Tongue/Tongue)
using all features on the Learning Set . . . . . . . . . . . . . . . . . . . 95
6.5 Example of a feature selection process from the inner cross-validation. 96
6.6 Confusion matrix for the hypopnoea event based classification using
feature selection on the Learning Set (S1). . . . . . . . . . . . . . . . . 97
6.7 Confusion matrix for the predominant site-of-collapse using feature
selection on the Learning Set(S1) . . . . . . . . . . . . . . . . . . . . . 97
6.8 Confusion matrix for the predominant site-of-collapse (non-Tongue/Tongue)
using feature selection on the Learning Set . . . . . . . . . . . . . . . . . 97
6.9 Most commonly selected features selected for the classifiers. . . . . 98
6.10 Most commonly selected features from the nested cross-validation pro-
cess. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 99
6.11 Confusion matrix for the hypopnoea event based site-of-collapse clas-
sification using selected features on the Hidden Set (S2) . . . . . . . . 99
6.12 Confusion matrix for the predominant site-of-collapse using selected features on the Hidden Set (S2) ................. 100
6.13 Confusion matrix for the predominant site-of-collapse (Tongue/non-Tongue) using selected features on the Hidden Set. ............... 100
6.14 Performance comparison of the state-of-the-art work achieved by models in published work ........................................... 102

7.1 Participant characteristics details. ........................................... 109
7.2 Number of features with different optimal cluster numbers. ............... 119
7.3 Features selected based on the number of data points in a cluster. ........ 120
7.4 Final features selected for the model. ..................................... 122
7.5 Confusion matrix for hypopnoea event based site-of-collapse (Tongue/non-Tongue) classification ........................................ 126
7.6 Confusion matrix for the predominant site-of-collapse (Tongue/non-Tongue) classification. .............................................. 126
7.7 Confusion matrix for hypopnoea event based site-of-collapse (Tongue/non-Tongue) classification ........................................ 127
7.8 Confusion matrix for the predominant site-of-collapse (Tongue/non-Tongue) classification. .............................................. 128
Chapter 1

Introduction

This thesis proposes a simple, non-contact method to automatically identify the site of the upper airway collapse in OSA patients using the acoustic properties of snoring during hypopnoeas in combination with machine learning algorithms. The information regarding the site-of-collapse could assist clinicians in choosing the most appropriate treatment for OSA.

This chapter discusses the introduction to the research and outlines the motivation, objectives and outcomes underpinning the current study. Next, the contributions of this work is explained. Finally, the structure of the thesis and the publications resulting from this work are listed.

1.1 Motivation

OSA is a highly prevalent and heterogeneous sleep-related breathing disorder characterised by repetitive, partial or complete obstruction of the upper airway [1–10]. OSA can lead to adverse consequences for health, quality of life and cognitive functions [11–15]. Conventional treatments for OSA include continuous positive airway pressure (CPAP) [16], oral appliance (OA) therapy [17], and surgery [18]; and the choice of therapy is guided by the severity of OSA and patient preference. A major challenge in sleep apnoea treatment is tailoring therapy to the individual patient in order to optimise outcomes. The first-line therapy, CPAP, is very effective but not accepted or tolerated by all patients; therefore, many tend to discontinue it after a short period [19]. Patients often prefer the alternative of OA therapy, but the treatment challenge is that OAs are only effective in some patients, and it is difficult to predict which patients will respond well to the therapy. OA is a custom made mouthpiece that is expensive to manufacture and fit to the patient, so the ability to predict treatment outcome has become an important goal for the field.

Several studies have been conducted to identify OA treatment outcome predictors using polysomnographic data, multi-sensor catheter parameters, drug-induced sleep endoscopy, and a remotely controlled mandibular positioner [20]. One indicator of the likelihood of success is identifying the location of the obstruction in the upper airway collapses during apnoea events. Researchers have shown that information
regarding the obstruction site could predict the treatment outcome and assist clinicians in choosing the most appropriate treatment. This is especially true for patients who have “tongue-based” airway collapse as they are far more likely to gain a large therapeutic benefit from OAs [21–25]. Conventional methods in determining the site-of-collapses are by using an endoscope [39–41] or a pressure catheter [42,43] during a drug-induced or natural sleep. Unfortunately, these methods are not well tolerated by the patients due to their invasive nature and the time/expense of the test (which limit its clinical applications). Also, studies have demonstrated inconsistencies regarding the obstruction site identified in terms of natural sleep and drug-induced sleep. Non-invasive imaging techniques such as the magnetic resonance imaging (MRI) [26] and the computed tomography (CT) [27] scan are also used to determine the obstruction site but these methods are not suitable for clinical practice due to its cost and the challenge of scanning during sleep.

Snoring is a clinical hallmark of OSA caused by the vibration of different tissues due to narrowing of the upper airway as the turbulent air passes through. Previous studies have shown that acoustic analysis of snoring has been successfully implemented in the diagnosis of OSA [28,29] and estimating OSA severity [30,31]. A few studies have been conducted in classifying the obstruction site based on snore data recorded simultaneously with drug-induced sleep endoscopy [32–34]. However, studies have demonstrated inconsistencies regarding the obstruction site identified in terms of natural and drug-induced sleep [35,36], and acoustic properties of snoring differ significantly from natural sleep [37,38]. Another drawback was that all the studies were based on analysing a single snore episode and thus were unable to determine the predominant site-of-collapse over an entire night. Therefore, the results have limited application in standard sleep studies. The challenge in identifying the site-of-collapse is to find a simple, non-invasive method without much inconvenience to the patients during natural sleep.

Motivated by these observations, this thesis explores the acoustic properties of snoring signal and their association with the site-of-collapse in the upper airway of OSA patients. This study innovatively applied a simple, non-contact method using audio recording during natural sleep to identify the site-of-collapse in the upper airway and then predicted the predominant site-of collapse for a patient. Therefore, the current method using snoring can benefit patients, as it avoids invasiveness and does not affect patients’ sleep quality. We believe the novelty of our work is that it is an important first step in establishing the feasibility of a practical non-invasive, low-cost diagnosis tool for improving the selection of appropriate therapy for OSA patients without any additional burden to the patients undergoing a sleep test.

1.2 Objectives and Outcomes

The overall goal of this thesis was to investigate if the acoustic properties of snoring sounds recorded during natural sleep could provide a simple, non-contact method to identify the site-of-collapse of the upper airway during hypopnoea events of OSA patients. This is based on the hypothesis that the characteristic of snore signal
varies depending on different excitation locations. The study was divided into three objectives.

The first objective was to generate a database of nocturnal audio recordings independently grouped into site-of-collapse classes using polysomnogram data from a sleep test. For this study, the probable site-of-collapse was determined by manual analysis of the shape of the airflow signal during hypopnoea, which has been shown to correlate well with the site-of-collapse. The second objective of this study was to develop an automatic method for labelling the OSA related snore events from the nocturnal audio recordings. The third objective was to identify the site-of-collapse in the upper airway by the acoustic analysis of snoring during hypopnoeas.

- Based on the aforementioned objectives, an automatic classifier was developed using machine learning algorithms to classify the snore signal into three classes ("lateral wall", "palate", "tongue-base" related collapse) and then identify the predominant site-of-collapse for each patient into one of the four classes ("lateral wall", "palate", "tongue-base" related collapse or multi-level" site-of-collapse). Supervised learning using classification algorithms were adopted to train the model using the labelled data and test to predict outcomes for new unseen data. This is based on the hypothesis that the acoustic properties of hypopnoea related snoring during natural sleep varies depending on different excitation locations.

- An unsupervised machine learning algorithm using cluster analysis was developed to form clusters using the features extracted from the snore data and to analyse the correlation between the clusters generated and the site-of-collapse in the upper airway. Clustering analysis was applied to show the tendency of the data to form groups according to their similarity without any labelling information. If a correlation exists between clusters and the site-of-collapse, then this supports the role of the snore signal as a correlate of the site-of-collapse.

- Feature selection algorithms were also developed based on the hypothesis that it results in building comprehensible learning models with good generalisation by selecting the high performing features.

1.3 Thesis Contributions

The principal contribution of this thesis is a simple, non-contact method for identifying the upper airway obstruction site in OSA patients using the snore signal recorded during hypopnoea events. This was achieved by developing efficient and systematic signal processing and machine learning algorithms for the identification of the site-of-collapse. All of the contributions will be discussed in more detail in the related chapters. The contributions in decreasing order of impact are as follows:

- Proposal of a novel approach for identifying the site-of-collapse in OSA patients using the snore signal with the help of machine learning classification algorithms (Chapter 6).
• Proposal of unsupervised machine learning using $k$-means clustering to form clusters using the features extracted from snore data during hypopnoeas and to investigate the correlation between the clusters generated and the site-of-collapse in the upper airway (Chapter 7).

• Adoption of an indirect method for labelling the snore during hypopnoea events based on the site-of-collapse using the data from a sleep study (airflow shape) without any additional burden to the patients undergoing a sleep test (Chapter 4).

• Creation of an audio database consisting of snore data based on the site-of-collapse. Also, the database consists of OSA related snore, simple snore and other sounds (Chapter 4).

• Identification of the most relevant features from the extracted features for the classification and clustering model by implementing a supervised feature selection algorithm (Chapter 5 and 6), as well as the development of a novel unsupervised feature selection algorithm (Chapter 7).

• Development of an automatic machine learning algorithm to extract OSA related snore from nocturnal audio recording (Chapter 5).

• Presentation of a review on OSA and the literature on common treatments for OSA, the significance of information regarding obstruction sites, common methods to identify obstruction sites, snore analysis and machine learning approaches used in OSA (Chapter 2).

1.4 Thesis Structure

The remainder of this thesis is organised as follows:

Chapter 2 presents necessary information regarding the clinical background of the study, upon which the thesis is built. This chapter reviews the literature relating to the clinical background of OSA. The clinical information regarding OSA, which consists of the health risks related to OSA, the prevalence of this disease, its pathophysiology, diagnosis, and upper airway physiology is given. This is followed by a description of the most common OSA treatments, such as CPAP, OA therapy, and surgery. In the next section, the details regarding the precision treatment for OSA are provided, followed by a discussion on the importance of the information regarding the obstruction sites and the various methods to assess upper airway obstruction. Next, snore sound characteristics and their application in OSA diagnosis are discussed.

Chapter 3 discusses the signal processing and machine learning approaches that are common across later chapters. This chapter first provides the details about the time and frequency domain features extracted from the nocturnal audio recordings. Fifty acoustic features are derived from the audio signal. In the next section, detailed information is provided regarding the machine learning methods adopted, including
two classification algorithms (Linear Discriminant Analysis (LDA) and Gaussian Mixture Model (GMM)), a cluster analysis (k-means) approach, and the motivation for selecting these approaches are given. This is followed by details of the nested cross-validation technique. In the final section, details of the feature selection algorithms are given.

Chapter 4 provides details regarding the database design process used to generate an audio database, which contains snore recordings based on sites-of-collapse, simple snore (other than snoring during hypopnoeas) and other sounds used for this study. This chapter begins with details on the study’s participants, followed by details of polysomnography (PSG) signals and the audio acquisition system. A pilot study was conducted to identify the timing differences between the audio recording and PSG data. A resampling method was adopted to synchronise the audio signal and the PSG signal. Details of the noise reduction technique to remove background noise from the audio recording using spectral subtraction is also discussed. In the next section, detailed information regarding the labelling of the site-of-collapse based on the airflow shape and the labelling of simple snore and other sounds are presented.

Chapter 5 describes the algorithm and methods used to extract the OSA related snore events from the audio recordings. This chapter provides detailed information on the signal processing and an automatic machine learning algorithm using an LDA classifier to extract OSA related snore events from the nocturnal audio recordings. Details on the feature selection algorithm, which uses a nested cross-validation technique are also provided. This is followed by an analysis of the results of the proposed algorithm and a discussion on the outcomes.

Chapter 6 presents the supervised machine learning algorithms developed for categorising the site-of-collapse using the snore signal recorded during hypopnoea events and then for predicting the predominant site-of-collapse for a patient. Details of the multi-class classifiers (LDA and GMM) and feature selection algorithms using nested cross-validation are given, followed by the results of the proposed model. Finally, an analysis of the results, and a discussion and comparison of the state-of-the-art research are given. This chapter emphasis the importance of acoustic analysis regarding snoring as a simple, non-contact method to identify the site-of-collapse.

Chapter 7 reports the details of an unsupervised machine learning algorithm using k-means clustering, which was developed to form clusters using the features extracted from the snore data during hypopnoeas and to analyse the correlation between the clusters generated and site-of-collapse in the upper airway. Details of the feature selection algorithm using principal component analysis and a novel method combining silhouette analysis with the Laplacian score algorithm are also provided. An analysis of the results based on the optimal cluster number, and the internal and external cluster validation are given, followed by a discussion and comparison of the supervised algorithm.

Chapter 8 concludes the thesis with recommendations for future research in this area.
1.5 List of Publications

**Journal Publications**


**Conference Publications**


Conference Abstracts


Other Publications

Chapter 2

Clinical Background

In this chapter, necessary information regarding the clinical background upon which the thesis is built is given. The main content of this chapter outlines the clinical information of OSA. The first section provides clinical information regarding OSA, which includes the health risks associated with OSA, the prevalence of the disease, its pathophysiology, diagnosis, and upper-airway physiology. In the following section, section 2.2, the most common treatments adopted for OSA are given, followed by the application of precision medicine principles in OSA treatment. Section 2.4 discusses the importance of identifying the obstruction sites and various techniques to identify these sites. In section 2.5, snoring sound characteristics and their application in OSA are discussed. The intention of this chapter is to provide a clinical understanding of the context of this thesis, which aims to provide the reader a clear idea about the potential importance and the application of this study. The chapter concludes with an explanation of the issues pertaining to the current method in identifying obstructions and discusses the main barrier to the widespread application of the method in precision OSA treatment.

2.1 Sleep Apnoea

The third edition of the International Classification of Sleep Disorders (ICSD-3) has categorised several sleep-related disorders into seven major classes for diagnostic and epidemiological purposes. This includes insomnia (the difficulty with which individuals fall asleep or find it hard to stay asleep), sleep-related breathing disorders (SBD; characterised by abnormal breathing patterns during sleep), central disorders of hypersomnolence (characterised by excessive daytime sleepiness), circadian rhythm sleep-wake disorders (CRSWDs; caused by de-synchronisation between internal sleep-wake rhythms and the light-darkness cycle), sleep-related movement disorders (characterised by abnormal movements that disturb sleep or its onset), parasomnia (abnormal events during sleep), and other sleep disorders.

SBD comprises a wide range of conditions characterised by abnormal breathing patterns during sleep. It is usually associated with irregular snoring, short-term upper airway resistance, and consecutive decrease or cessation of airflow produced
by partial or total obstruction of the upper airway [44]. SBD is broadly divided into OSA, central sleep apnoea (CSA), sleep-related hypoventilation, and sleep-related hypoxemia disorder based on ICSD-3 classification. Sleep apnoea (SA) is a common type of SBD characterised by repeated, involuntary cessation of breathing for a short duration during sleep. The Greek word “apnoea” means “without breath”. The cessation of breathing leads to a reduction in oxygen levels in the blood and can briefly awaken sleepers throughout the night. The cessation events may occur hundreds of times a night, and the patient may not be aware of any sleep disturbance the next day [45]. It can significantly affect the health, quality of life and cognitive functioning of the patient. SA can be classified as obstructive sleep apnoea, central sleep apnoea and mixed sleep apnoea. Currently, diagnosis and the severity of apnoea is carried out by measuring the apnoea-hypopnoea index (AHI), computed by taking the average number of apnoea (complete obstruction) and hypopnoea (partial obstruction) events per hour of sleep. An AHI of greater than five is defined as SA [46]. An AHI of five or less is generally not clinically significant and assumed normal.

2.1.1 Obstructive Sleep Apnoea

OSA is the most common type of sleep apnoea and is caused due to a blockage or collapse of the upper airways [47]. OSA is characterised by, repetitive partial or complete obstruction of the upper airway during sleep, as shown in Figure 2.1. OSA is recognised by the presence of thoracic and abdominal efforts for continued breathing while airflow completely stops. For OSA patients, breathing is obstructed by the collapse of the muscles that maintain the patency of the upper airway as it becomes less active during sleep. As there is insufficient oxygen reaching the lungs due to the airway blockage, the brain signals the body to wake up. This results in the reopening of the airway, causes the person to gasp for air and subsequently breathing resumes [48]. When the breathing resumes the blood oxygen level is restored to the normal level, and the person falls asleep again, setting off another cycle.

2.1.2 Central Sleep Apnoea

Central sleep apnoea (CSA) is characterised by a lack of respiratory effort during airflow cessations. CSA occurs when the brain fails to send a signal to the respiratory muscles to breathe. In contrast to OSA, there exists no issue with the respiratory organs in CSA. CSA arises when the brain temporarily stops signalling the muscle to breathe due to the damage in the part of the brain that controls breathing [49]. The airway is not obstructed during CSA, and there is nothing to restrict normal breathing. CSA is much less common than OSA.

2.1.3 Mixed Sleep Apnoea

Mixed sleep apnoea (MSA) is accompanied by a combination of OSA and CSA, where the symptoms of CSA are exhibited with no evidence of inspiratory effort in the first few seconds and are then followed by the signs of obstructive sleep apnoea [50]. In
Figure 2.1: Diagram showing normal breathing and obstructed breathing (source: biostrap.com/).

In this case, the breathing control system is more sensitive to the changes in oxygen or carbon dioxide, and, therefore, obstructed efforts to breathe are greater. And when the airway opens, ventilation is higher.

The main points of difference between OSA and CSA are shown in Table 2.1. The main difference between OSA and CSA arises due to the pathophysiologic mechanism that causes a respiratory disturbance [51]. First, the upper airway obstruction in OSA is usually related to the abnormal anatomy and/or collapse of the muscles that maintain the patency of the upper airway; whereas, MSA is caused by the dysfunction of the ventilator control in the central nervous system. Breathing effort indications such as chest wall movements can be observed in OSA events, but no sign of breathing effort is observed in CSA [51]. Although there are differences, there exists some overlap between the central and obstructive events during MSA. Additionally, some CSA patients exhibit snoring, which is a typical feature of OSA.

Table 2.1: Typical comparison of obstructive sleep apnoea and central sleep apnoea (reproduced from [51]).

<table>
<thead>
<tr>
<th>Obstructive Sleep Apnoea</th>
<th>Central Sleep Apnea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related to abnormal anatomy and/or collapse of the muscle</td>
<td>Caused by the dysfunction of the ventilator control in the central nervous</td>
</tr>
<tr>
<td>that maintain the patency of the upper airway</td>
<td>system</td>
</tr>
<tr>
<td>Breathing effort indications such as chest wall movement,</td>
<td>No sign of respiratory effort during apnoeas</td>
</tr>
<tr>
<td>oesophageal pressure, pulse transit time can be observed</td>
<td></td>
</tr>
<tr>
<td>Flow or nasal pressure tracing has flat tops in inspiration</td>
<td>Flow or nasal pressure has round tops in inspiration</td>
</tr>
<tr>
<td>Oxygen saturation curve is asymmetrical with slow decrease</td>
<td>Oxygen saturation curve is sinusoidal</td>
</tr>
<tr>
<td>and quick recovery</td>
<td></td>
</tr>
<tr>
<td>Variable apnoea cycle duration</td>
<td>Constant apnoea cycle duration</td>
</tr>
</tbody>
</table>
2.1.4 Health Risk Associated with OSA

The typical symptoms of OSA are excessive daytime sleepiness, snoring and presence of apnoea. The most common symptoms of OSA are abrupt awakenings accompanied by choking/gasping episodes, snoring, reduced total sleep amount, non-restorative sleep, sleep fragmentation/sleep maintenance insomnia, morning headaches, sexual problems, loss of libido, irritability, poor memory, lack of concentration and reduced cognitive potential [52]. Some of these symptoms are also present in mental health pathologies, such as depression and anxiety [53, 54]. Individuals often feel unrested, fatigued, and sleepy during the daytime, which can lead to impairments in vigilance, concentration, cognitive function, social interactions and quality of life. These symptoms may lead to increased rate of work-related and motor vehicle accidents [14,15].

OSA is associated with far-reaching health-related consequences. It has direct physiological effects that result in disturbed sleep, intrathoracic pressure oscillations, hypoxia, sympathetic activation and mechanical effects. The most common consequence of OSA is excessive daytime sleepiness which can adversely affect an individual’s quality of life and cognitive functioning [13–15]. OSA has been found to be heavily associated with hypertension and obesity [55–57], increased risk of cardiovascular disease and mortality [11, 12], poorer quality of life [15], and cognitive dysfunction [58]. Furthermore, studies have shown there exists a strong correlation between OSA and road traffic accidents [59]. Studies suggest that OSA has a complex and bidirectional relationship with obesity [57]. Additionally, the causal link between OSA and systemic hypertension has been clearly established [55]. Substantial evidence shows that an increased prevalence of hypertension is directly linked with the
existence and severity of OSA. Studies have also shown that OSA is associated with an increased risk of cardiovascular diseases such as cardiac arrhythmias, coronary artery disease, heart failure, transient ischemic attacks and stroke [11, 12, 59]. It is additionally associated with metabolic dysregulation, which affects glucose control and risk of diabetes [60, 61]. Therefore, early diagnosis and treatment of OSA are important for both patients and society to reduce its associated health costs. An outline of the numerous health risks related to OSA is given in Figure 2.2.

2.1.5 Prevalence of OSA

Studies have identified that OSA is a highly prevalent disorder, both in the general population and in population sub-groups with a specific disease. Based on the study population, and on how OSA has been defined, the prevalence of OSA differs significantly between the study groups. The difference over time could be due to the difference in various factors such as testing methodology, testing equipment type, the scoring criteria used, and the AHI threshold. Additionally, there is an increase in the prevalence of OSA, which may be affected by rising obesity worldwide.

Epidemiological studies have demonstrated that the prevalence of OSA is approximately 24% in male and 9% in females [2]. Table 2.2 provides details of the prevalence study for OSA worldwide. Different studies show that values for the prevalence of OSA in adults varies from 3% to 49%. On the other hand, approximately 80% - 90% of people with OSA remain undiagnosed [1,2]. OSA is more prevalent in men than in women. Age is an important risk factor in the prevalence of OSA, where the older an individual is, the greater their risk. For example, a study with a general population of aged 65 or older demonstrated that 31% of males and 19% of the females have OSA [8]. Another study estimated that 70% of the male and 56% of female participants in the 65–95 age range had OSA [10]. Furthermore, a recent study found that 83.8% of men and 60.8% of women 40 years or older had mild or severe OSA [9]. It should be noted that the studies discussed in Table 2.2 used different thresholds for the AHI of OSA, but it is usually defined between five and 10 (moderate to severe sleep-disordered breathing). OSA prevalence is higher in individuals with obesity [56, 57]. Individuals who are at greater risk of OSA are those who are overweight or obese, smokers, middle-aged or older (generally men), have large neck sizes, have hypertension, consume alcohol, those with Down’s Syndrome, women going through menopause, and those with a family history of SA [66]. OSA prevalence varies depending on several factors such as sex ratio, study population age, scoring criteria. The overall prevalence of OSA is likely to be higher given that the analysis focused only on adults aged 30–69 years.

2.1.6 Pathophysiology of OSA

The fundamental pathophysiology depends on several factors and may differ considerably between individuals based on factors such as sex, age, and whether or not they are obese [67]. The most important factors include anatomy of the upper airway, upper airway dilator muscle activity and reflex responsiveness, arousal during sleep, the
Table 2.2: Prevalence of OSA around the world.

<table>
<thead>
<tr>
<th>Location</th>
<th>Age Range (Years)</th>
<th>Scoring Criteria</th>
<th>OSA Prevalence Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA [2]</td>
<td>30-60</td>
<td>Chicago 1999</td>
<td>4 (M) : 2(F)</td>
</tr>
<tr>
<td>Asia [4]</td>
<td>30-60</td>
<td>Chicago 1999</td>
<td>5 (M) : 2.5 (F)</td>
</tr>
<tr>
<td>USA [6]</td>
<td>40-60</td>
<td>AASM 2012</td>
<td>25.5 (M) : 23.5 (F)</td>
</tr>
<tr>
<td>Spain [7]</td>
<td>30-70</td>
<td>AASM 2007</td>
<td>26.2 (M) : 28 (F)</td>
</tr>
<tr>
<td>USA [8]</td>
<td>≥ 65</td>
<td>Chicago 1999</td>
<td>31 (M) : 19 (F)</td>
</tr>
<tr>
<td>Switzerland [9]</td>
<td>40-85</td>
<td>AASM 2012</td>
<td>83.8 (M) : 61 (F)</td>
</tr>
<tr>
<td>Australia [65]</td>
<td>40-65</td>
<td>AASM 2012</td>
<td>25.5 (M) : 23 (F)</td>
</tr>
</tbody>
</table>

M = Male, F = Female. AHI ≥ 5 events per hour.

stability of the respiratory control system (loop gain), and lung volume [67]. From a purely anatomical perspective, a narrow upper airway is generally more susceptible to collapse than a larger one. Another important point in the pathogenesis of OSA relates to the capability of the upper airway dilator muscles to respond to the respiratory changes during sleep [68]. Arousal from sleep at the cessation during a hypopnoea or an apnoea has long been believed to be an important protective mechanism for airway reopening. Ventilatory control stability is believed to be an important contributor to OSA pathogenesis, and the stability of the respiratory control system can be described using loop gain. In the case of respiration, loop gain represents the gain or sensitivity of the negative feedback loop that controls ventilation. Mathematically, it is defined as the ratio of a corrective response (e.g., hyperpnea) to a disturbance (e.g., apnoea). Apart from this, the potential for state-related changes in lung volume is believed to be an important contributor to OSA pathogenesis [69].

A schematic representation of the typical physiological process during OSA and corresponding physiological processes is shown in Figure 2.3. The elements displayed outside the circles are either protective or restorative, while the factors indicated inside the circle are preservative. This flowchart represents the mechanisms caused by OSA or hypopnoea in a loop commencing by a drop of oxygen level, elevated CO₂ and augmented breathing effort, followed by arousals.

### 2.1.7 Diagnosis of OSA

The gold standard method for the diagnosis of sleep apnoea is carried out using polysomnography (PSG). PSG refers to the continuous monitoring and recording of various physiological data during sleep and is considered to be the standard method for the diagnosis of various sleep-related disorders such as OSA, CSA, narcolepsy, nocturnal seizures, rapid eye movement (REM) sleep behaviour disorder, and periodic
limb movement disorder [70]. It is a multisensory method characterised by the simultaneous recording of the electrical activity of the heart (electrocardiogram [ECG]), skeletal muscle (electromyogram [EMG]), brain (electroencephalogram [EEG]), eyes (electrooculogram [EOG]), respiratory effort (respiratory inductance plethysmography), airflow (thermistors, nasal air pressure sensors) and blood oxygen levels (pulse oximeter). Additional sensors including body position and auxiliary devices such as the sound level metre are also used [70, 71]. Sometimes video and audio monitoring are also conducted. These signals are used to determine the patients sleep state as well as to identify any sleep-related respiratory or movement disorders.

Generally, PSG is performed overnight in specialised sleep clinics (in-laboratory), and a patient stays in an unfamiliar environment (hospital or sleep clinic) with several sensors and wires attached to their body (i.e., face, skull, chest and limbs) [70–73]. An example of a simplified PSG set up is shown in Figure 2.4. For the in-laboratory study, all the study preparations (e.g., patient preparation, PSG hookup, calibration of measurements, biological calibration), data monitoring and data acquisition are monitored for the entire session by an expert sleep technician. A sleep technician monitors the patient throughout the night from the observation room with the help of closed-circuit television with audio and live streaming of the PSG data (to verify that any physiological signals were recorded correctly and to ensure that no signal was missed due to instrumentation problems or due to artefacts) and to provide assistance if required.

The American Academy of Sleep Medicine (AASM) aimed to standardise the clinical guidelines for PSG and manual for scoring sleep-related events [62, 64]. This manual provides recommendations for the sensors to use in PSG, the proper position-
Figure 2.4: An illustration of polysomnography set-up and attachment of sensors for monitoring sleep (source: /qa.virinchihospitals.com).

The AASM manual clarified the definition of apnoea, providing rules for measuring event duration, the extent of airflow reduction and the sensors to be used for apnoea detection by updating the AASM ‘Chicago Criteria’ [63] for respiratory event scoring. The definitions of hypopnoea and apnoea provided by the AASM have been updated over the years and are given in Table 2.3. The first publication by AASM for a standard SDB definition in 1999 indicated that obstructive apnoea or hypopnoea events are associated with a partial or total blockage of airflow in the upper airway leading to an interruption of regular breathing [63]. The 2007 AASM scoring manual recommended that a hypopnoea definition that requires a change in 30% or more in airflow be associated with a 3% oxygen desaturation or cortical arousal but allows for an alternative definition that requires association with a 4% oxygen desaturation without consideration of cortical arousals [62]. AASM updated the 2007 respiratory scoring manual in 2012 to further improve the standardisation of scoring respiratory events as well as the changes in the definition of a hypopnoea event [64]. The definition of hypopnoea was updated in AASM scoring Manual 2012 and recommended the hypopnoea definition requires a 30% or greater reduction in nasal pressure excursions for 10 seconds longer associated with ≥4% oxygen desaturation [64]. The Apnoea-Hypopnoea Index (AHI) was calculated as the average number of apnoea and hypopnoea events per hour computed over total sleep time. AHI was used in the OSA diagnosis (AHI ≥ 5) and categorising OSA severity (mild: 5 ≤ AHI < 15, moderate 15 ≤ AHI < 30, severe AHI ≥ 30).

Respiratory events (number of obstructive apnoeas, mixed apnoeas, central apnoeas and hypopnoeas), sleep stage (wakefulness, Stage 1, Stage 2, Stage 3, REM), sleep scoring (lights out and on time, total sleep time, total recording time, sleep
Table 2.3: Different criteria and updates provided by AASM for apnoea and hypopnoea.

<table>
<thead>
<tr>
<th>Year</th>
<th>Apnoea</th>
<th>Hypopnoea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999 Chicago</td>
<td>1. Drop in airflow ≥ 50% from baseline, or</td>
<td>1. Drop in airflow ≥ 50% from baseline1, or</td>
</tr>
<tr>
<td>Criteria</td>
<td>2. Arterial oxygen desaturation ≥ 3% or an arousal, and</td>
<td>2. Arterial oxygen desaturation ≥ 3% or an arousal, and</td>
</tr>
<tr>
<td></td>
<td>3. The event duration ≥ 10 seconds</td>
<td>3. The event duration ≥ 10 seconds</td>
</tr>
<tr>
<td>2007</td>
<td>1. Peak signal excursion* ≥ 90% of pre-event baseline, and</td>
<td>1. Nasal pressure peak signal excursion ≥ 30% of pre-event baseline, and</td>
</tr>
<tr>
<td>Recommended</td>
<td>2. The event duration ≥ 10 seconds, and</td>
<td>2. Duration of this drop ≥ 10 seconds, and</td>
</tr>
<tr>
<td></td>
<td>3. Event’s duration meeting amplitude reduction criteria ≥ 90%</td>
<td>3. Arterial oxygen desaturation from pre-event baseline ≥ 4%</td>
</tr>
<tr>
<td></td>
<td>* using oronasal thermal sensor,</td>
<td>4. Event duration meeting amplitude reduction criteria for hypopnoea ≥ 90%</td>
</tr>
<tr>
<td>2007</td>
<td>PAP (positive airway pressure)</td>
<td></td>
</tr>
<tr>
<td>Alternative</td>
<td>titration device or alternative apnoea sensors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. Nasal pressure signal drops ≥ 50% of baseline</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Duration of this drop ≥ 10 seconds, and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Arterial oxygen desaturation from pre-event baseline or an event</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Event duration meeting amplitude reduction criteria for hypopnoea</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥ 90%</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>1. Drop in peak thermal sensor excursion ≥ 90% of pre-event baseline</td>
<td>1. Nasal pressure signal excursion drop ≥ 30% of pre-event baseline, and</td>
</tr>
<tr>
<td></td>
<td>2. The event duration ≥ 10 seconds</td>
<td>2. Duration of this drop ≥ 10 seconds, and</td>
</tr>
<tr>
<td></td>
<td>3. If a portion of hypopnoea event, meets</td>
<td>3. Arterial oxygen desaturation or event with arousal ≥ 3%</td>
</tr>
<tr>
<td></td>
<td>criteria for apnoea, the entire event scored as apnoea</td>
<td></td>
</tr>
</tbody>
</table>

* Baseline is the mean amplitude of stable breathing or three largest breaths in the two minutes preceding the event onset for patients without stable breathing.

latency), arousal events, and movement events (number of periodic limb movements of sleep, the number of movements per hour divided by total sleep time) are visually annotated by an expert sleep technician using the AASM scoring rules as a part of the standard clinical diagnosis of OSA [64].

While the PSG provides a comprehensive picture of a patient’s sleep, it has many limitations. The main limitations comprise its cost, long waiting times due to limited resources, and the discomfort and poor-quality sleep associated with various sensors attachment. Home PSG was developed for the diagnosis of sleep apnoea, which can be set up in a familiar environment, but this still requires the attachment of multiple uncomfortable sensors to the body.

### 2.1.8 Upper Airway Physiology

The human upper airway is a unique structure that performs several functions such as vocalisation, swallowing, and respiration. The primary respiratory function of the upper airway is to regulate the airflow by permitting the air movement into and out of the lungs. The upper airway extends from the mouth into the trachea and consists of a collapsible portion that extends from the hard palate to the larynx [75]. The human
The upper airway comprises approximately 20 muscles (muscles regulating the position of the soft palate, tongue muscles, the hyoid apparatus, and the posterolateral pharyngeal walls) that work together to keep the airway open. The human pharynx is divided into four sections: 1) The nasopharynx, which begins from the nasal turbinates to the start of the soft palate; 2) the velopharynx or retropalatal oropharynx, which begins at the start of the soft palate to the tip of the uvula; 3) the oropharynx, which begins from the tip of the uvula to the tip of the epiglottis; and 4) the hypopharynx, which starts at the tip of the epiglottis and extends to the level of the vocal cords. The upper airway has a complex geometry consisting of various muscles and soft tissue but lacks rigid or bony support. Therefore, it is responsible for a major component of airway obstruction in humans during sleep. Soft-tissue structure affects the airway size, shape, and compliance, which in turn affect airflow during sleep [76]. The position and size of the tongue, palate, and tonsils are important considerations in OSA [67–69, 75, 76]. A narrow upper airway generally results in more collapse than a larger one. The narrowing of the upper airway occurs mainly due to the enlargement of various soft tissue parts in the vicinity of the airway, which results in increased airway surface area and thereby rises in the surface tension forces and thus collapsibility [67]. Obesity significantly contributes to airway compression through the increased area and volume of pharyngeal fat deposits [57,58]. Similarly, CT and MRI studies have given evidence for increased fat deposition in the lateral walls of the pharynx, tongue, soft palate, or uvula. Studies have shown that the upper airway cross-sectional area identified using imaging techniques such as CT and MRI is less in OSA patients compared with the control participants. The measurement of genioglossus (a tongue muscle) using multi-unit EMG intramuscular electrodes have demonstrated a reduced ability of the upper airway dilator muscles to maintain a patent airway during sleep in OSA patients [77]. These studies suggest that the upper airway structure is different in OSA patients compared with control participants.
and therefore, may be used to identify OSA patients.

2.2 Treatments for OSA

Conventional treatments for OSA include continuous positive airway pressure (CPAP), oral appliance (OA) therapy, and surgery, and the use of these varies depending on OSA severity and patient adherence. Adjunctive treatment such as diet and lifestyle, or behavioural measures and pharmacologic treatments are also adopted for some patients. OSA is a sleep-related anatomical problem, and most of the existing treatments for OSA aim to correct the anatomical problem.

2.2.1 Continuous Positive Airway Pressure

CPAP is the most common treatment for OSA patients and is considered the first-line treatment for moderate to severe OSA. The aim of CPAP is to provide mechanical stability in the upper airway with a column of pressurised air. The CPAP machine provides pressurised air through the airway at a constant pressure that prevents the airway from collapsing. CPAP was first introduced as a treatment for OSA by Dr. Colin Sullivan at the University of Sydney in 1981 [78]. Prior to this, the common treatment for OSA was a tracheotomy which was a highly invasive procedure with severe repercussions [79]. Dr. Sullivan developed a non-invasive ventilation system to avoid the collapse of the upper airway during sleep by providing continuous pressurised airflow through a mask. The treatment keeps the airway open and helps patients maintain normal oxygen levels.

The main components of a CPAP machine consist of an air pump, a mask, and a connecting tube. The air pump pressurises room air. The pressurised air is then transferred through the tube to nose/ nose and mouth into the upper airway [80]. Based on the machine type, air pressure can be fixed or variable. A CPAP generates fixed air pressure throughout the night. An autotitrating PAP (APAP) can provide varied air pressure based on sleep stage and body position, and can vary the airflow accordingly and supply the optimum pressure [81–83]. The bi-level PAP (BiPAP) provides two levels of pressure, with more pressure during inspiration and less during expiration [84, 85]. Figure 2.6 shows an illustration of a CPAP setting and the basic working of a CPAP preventing airway obstruction.

CPAP is considered the most effective and commonly used treatment for patients with moderate to severe OSA. Studies have already validated the performance of CPAP and demonstrated that it improves sleep quality, daytime sleepiness, health-related quality of life, oxygen saturation, and reduces blood pressure and cardiovascular risk. Studies have also shown that usage of CPAP on a regular basis improves treatment outcomes [86–89]. Moreover, treatment outcomes significantly depend on the duration of overnight use (greater benefit with increased nightly use, although the rate of improvement is relatively less for usage exceeding four hrs.) [90]. A study conducted to investigate the polysomnographic and anthropomorphic factors in predicting the treatment outcome of OSA showed that patients with moderate to severe
OSA responded well to the treatment compared with mild or moderate OSA patients [89]. Some studies have shown that patients prefer APAP over CPAP [89]. However, more studies are necessary to validate the adherence rate of APAP and to compare it with the results generated by the use of CPAP. A large number of studies have compared the adherence of CPAP and BiPAP, and results clearly indicate that BiPAP does not improve adherence [85].

Even though CPAP is considered first-line treatment and is highly efficacious in preventing upper airway collapse, low patient adherence and acceptance makes it less effective in the real world [20]. Factors that influence adherence to CPAP include disease and patient characteristics (severity, information regarding upper airway patency), treatment titration procedures, technological device factors and side effects (discomfort due to the mask, mask-related side effects at nose or face), and psychological and social factors (self-efficacy, outcome expectations and social support) [91]. Studies have reported that adherence to CPAP varies from 30% and 80% [92–94].

### 2.2.2 Oral Appliance

Oral appliance (OA) treatment for OSA has emerged as an alternative treatment to CPAP. OAs are currently widely used for the treatment of snoring, mild, and mild to moderate OSA, both as a primary therapy and as an alternative treatment for those who do not tolerate CPAP treatment [95, 96]. OAs are placed into the mouth at night before going to bed, and kept for the entire sleeping time, and taken out when awake. Oral appliances work by keeping tongue or mandible in an advanced position and help prevent the upper airway collapsing. OAs can be classified as 1) tongue retaining devices or tongue stabilising devices (by holding the tongue forward) and 2) mandibular advancement devices (MAD) (by protruding the mandible and attached tongue), with the latter being more commonly used.

OAs are a simple, reversible, quiet and cost-effective treatment compared with CPAP for selected patients with OSA. A study conducted using MRI showed that
OAs can enlarge the upper airway space, with the biggest impact in the velopharyngeal region [97]. OAs are considered to be as equally effective as CPAP in mild to moderate OSA and are less effective in severe OSA patients [98]. However, OA adherence is higher (76% to 95%) compared with CPAP (30% - 80%) [99]. There is strong evidence demonstrating that OAs improve OSA in many patients, including some with a more severe condition [100]. However, OAs are not efficacious for all, as around one-third of individuals will show negligible improvement [101]. Recent studies suggest that information regarding the site-of-collapse could assist clinicians in choosing better treatment using OA therapy [21–25, 97, 102–108]. More recently, studies have validated the performance of OAs and have achieved high accuracy in predicting the responders of OA therapy based on patients obstruction site information (tongue-based related collapse) [21, 22]. One study suggests that the combination of tongue exercises along with mandibular repositioning appliances (MRA) is a promising approach for patients with obstruction in velopharyngeal area [102]. Another study demonstrated that an increased antero-posterior diameter with anterior displacement of the tongue using tongue stabilising devices was effective [103]. Another study showed similar results of a greater increase in velopharyngeal and oropharyngeal CSA with tongue-base compared with mandible protrusion via nasopharyngoscopic investigation [104]. An increased velopharyngeal lateral diameter appears to be the main effect of MAS on the upper airway [97, 105, 106]. Studies have also demonstrated that the information regarding posteriorly-located tongue collapse in upper airway predicts oral appliance treatment success, with high accuracy [21–25]. Figure 2.7 shows an example of OA and the basic principle of OAs behind preventing airway obstruction. Another study demonstrated that tongue base collapse was found to increase the probability of a good oral appliance response [108]. Intuitively, patients with tongue-related obstruction are expected to exhibit a substantial improvement in collapsibility with OAs [107, 108].
2.2.3 Surgery

Surgery is an option in patients with obvious anatomical issues or who have failed or cannot adhere to other treatments [109–111]. Surgery is considered a last resort treatment in some patients who have failed to experience positive outcomes from other treatments. It aims to reduce anatomical upper airway obstruction in the nose, oropharynx, and hypopharynx. Surgery can be classified into the following categories: soft tissue (e.g. uvulopalatopharyngoplasty [UPPP]) [109], skeletal (e.g., maxillomandibular advancement [MMA]) [111], tracheostomy [112], and hypoglossal nerve stimulation [113]. The success rate for surgical treatments for OSA is highly unpredictable and has been found to be less effective compared with CPAP and OAs. Surgical success depends on accurate information regarding the obstruction site, identifying a suitable patient, the type of surgery performed, and the experience of the surgeon [109,110].

2.2.4 Life Style and Diet Treatment

Behavioural alterations such as losing weight, controlling diet, exercising, avoiding sleeping medications, alcohol and tobacco, or engaging in body position therapy (e.g., changing sleep position from supine to side or raising the head from the bed) can help some OSA patients reduce their AHI [114,115]. One study showed the improvement in pharyngeal patency with sleeping position is structure specific, with profound improvements seen in patients with epiglottic collapse, modest effects in those without tongue involvement and—unexpectedly—no effect in those with tongue-related obstruction [283]. Therefore, positional therapy can be considered as an alternative for CPAP treatment in these patients.

2.3 Precision Medicine in OSA

Even though different types of treatment are available for OSA, the key unresolved issue is the inability to predict the treatment outcome or to predict the most appropriate treatment for a patient. Recently, precision medicine approaches for OSA treatment are seeing a much wider acceptance [116]. Precision medicine is an approach to health care that recognises a group of patients with the same characteristics, namely unique subtypes due to individual differences (heterogeneity) [117]. Subtype membership appears to have implications for diagnosis, treatment, consequences and disease prevention.

The heterogeneous behaviour of OSA can be classified into smaller groups based on more homogeneous disorder subtypes such as symptoms, demographic characteristics, polysomnographic and physiologic metrics, or comorbidities. This can lead to a more specific diagnostic and treatment strategy, thereby improving treatment outcomes and successful clinical trials [118]. PSG information, novel PSG metrics derived from PSG information (i.e., sleep drive, sleep depth, arousal intensity and heart rate response to arousals, arousal threshold, and hypoxic burden) and other pathophysiological factors such as loop gain, information regarding the obstruction
site, arousal threshold and muscle compensation can assist in selecting the most appropriate therapy for a patient. One study demonstrated that anatomical phenotype based on the lack of complete concentric palatal collapse can assist in identifying airway neurostimulation treatment responders, enabling a new treatment option [119]. Another study incorporated demographic, anthropometric, AHI, and co-morbidity data in addition to information on symptoms were used to form clusters based on latent class analysis classifier and evaluated their relationship to CPAP treatment outcomes [120,121]. A cluster analysis was performed using OSA symptoms such as daytime sleepiness, and other difficulties in sleeping, and validated the CPAP treatment response based on the clusters generated [122]. It has been demonstrated that targeted therapy, (e.g., use of sedatives) is highly effective in patients with a low arousal threshold [123]. Zinchuk et al. demonstrated the poor performance of CPAP in patients with a low arousal threshold [124]. Sands et al. used pathophysiological information such as loop gain, collapsibility, muscle compensation and arousability to predict treatment responders using oxygen therapy [125]. Identifying the pharyngeal collapsibility and the site-of-collapse also has implications for responses to particular therapies, such as OA and surgery [23, 24]. Previous studies have suggested that identifying patients with tongue-based collapse could assist clinicians in determining better treatment using OA therapy [23, 24].

In this thesis, the main aim is to build a phenotype of OSA patients based on the site-of-collapse using simple, non-contact and inexpensive technology by exploiting snoring properties and machine learning techniques. This could assist in selecting the most appropriate treatment for OSA. In Chapter 6, an automatic classifier that predicts the predominant site-of-upper airway collapse for a patient was discussed. Chapter 7 introduces a clustering algorithm to form clusters using the features extracted from the snore data collected during hypopnoeas and analyse the correlation between the clusters generated and site-of-collapse in the upper airway. The aim was to subdivide the patients into more homogeneous subgroups with common features based on the obstruction site.

2.4 Identification of Obstruction Sites

The cause of obstruction in the upper airway is multifactorial. The factors contributing to the pathogenesis are anatomic narrowing, increased collapsibility of soft tissues, reflexes affecting the calibre of the upper airway and pharyngeal muscle function. Previous studies have successfully demonstrated that accurate information regarding obstruction sites could predict treatment outcomes and assist in choosing the most appropriate treatment for OSA [21–25]. Accurate information regarding the obstruction site is essential for effective surgical treatment [109–111].

As PSG studies do not provide much information regarding the obstruction sites, various techniques have been developed to understand the functioning and the behaviour of the upper airway and to identify the site-of-collapse. Once OSA is diagnosed, a further examination is typically conducted to identify the site or sites of obstruction. Although there are various techniques available, no standard diagnostic
A technique has been formulated to identify the exact site-of-obstruction.

### 2.4.1 Fluoroscopy

Fluoroscopy is a technique used to evaluate the anatomy of the upper airway by passing a continuous x-ray beam in OSA patients. The main advantages of this technique are that it is minimally invasive, allows real-time imaging, avoids anaesthesia and can be combined with PSG. Somnofluoroscopy has the ability to evaluate obstruction sites during episodes of apnoea and hypopnoea by combining fluoroscopy and PSG signals. However, the main disadvantages of fluoroscopy are limited testing time due to the use of radiation, superimposition of structures can affect the accuracy of the information, it represents the upper airway in two dimensions, and some patients require sedation.

A study conducted to identify the site-of-collapse using somnofluoroscopy showed that the success rate of UPPP is improved by selecting patients with oropharynx obstruction [126]. Lee et al. conducted a study with large population (≈1000) using sleep video fluoroscopy revealed that multi-level obstruction was observed in one-third of the population with severe OSA [127]. A study investigating the association of MAD and upper airway obstruction using sleep videofluoroscopy identified that MAD could widen retropalatal space and retrolingual space and thereby improve OSA [128]. Kim et al. showed that the evaluation of obstruction (performed by a less experienced sleep surgeon) using sleep videofluoroscopy achieved a higher inter-rater reliability [129].

### 2.4.2 Pressure Catheters

Pressure sensing catheters positioned in various locations of the upper airway can determine pressure differences during an apnoea event to identify sites-of-obstruction. Researchers have attempted to pinpoint a site of collapse using a wide variety of pressure catheters such as fluid-filled catheters, movable catheters, bias-flow catheters, and micro-pressure sensors. The catheter is inserted via the nose into the upper airway to record the pressure at different locations of the upper airway. An illustration of the placement of a pressure catheter for the identification of airway obstruction is shown in Figure 2.8. During an apnoea, pressure sensors located above the site of obstruction have no pressure and with increasing pressure swing in sensors located below the obstruction site. The pressure deflection pattern from different sensors during an apnoea can be utilised to identify the obstruction site. The main advantages of using a pressure catheter are that a full night recording can be obtained and that it can be combined with PSG. Certain conditions, however, are required to obtain accurate information. First, the catheter must be positioned in such a way that it does not disturb the sleep or breathing. Training is required before personnel can reliably position the catheter. Furthermore, the precise location of obstruction significantly depends on the number of pressure sensors used; however, sometimes the catheter prevents apnoea from occurring. Aside from these issues, the use of anaesthesia may alter the obstruction site.
Figure 2.8: An illustration of the placement of pressure catheter for the identification of airway obstruction (https://www.spiedigitallibrary.org/).

Studies have demonstrated that pressure catheters can identify the same number of obstructive and mixed apnoea events during sleep as the PSG [42]. A study conducted using the positioning of a pressure catheter at the nasopharynx, oropharynx, tongue base, hypopharynx and oesophagus identified that obstruction could occur at multiple locations in most participants [43]. In one study, a multisensory pressure catheter approach was adopted to locate the obstruction site and it demonstrated that UPPP could benefit patients with oropharynx collapse [130]. In another investigation, information regarding the predominant palatal obstruction detected by multi-channel pressure transducers, was found to assist in uvulopalatoplasty and achieve a reduction in OSA severity [131]. Osnes et al. conducted a study to evaluate the efficacy of UPPP using a pressure catheter, demonstrated that patients with trans-palatal apnoea show a higher successful treatment rate than patients with sub-palatal collapse [132]. Henke et al. conducted a study to determine the OA treatment response of patients based on obstruction site information identified using a pressure catheter showed that the mandibular advancement device had a positive response in OSA patients with pharyngeal closure in the hypopharynx or velopharynx [133]. Ng et al. investigated OA treatment response based on the obstruction site, OA was shown to exhibit a superior response in patients with predominantly sub-palatal collapse compared with those who had a palatal collapse [24].

2.4.3 Computed Tomography Imaging

A computerised tomography (CT) scan is a non-invasive imaging technique that utilises computers and a rotating X-ray machine to capture cross-sectional images of the body. CT scans can be used to provide a quantitative assessment of the upper airway. CT scan images can provide complete information regarding the cross-
sectional area of the upper airway during natural or drug-induced sleep. The main advantages of this method are that it is a non-invasive method that can scan the entire upper airway without using any sedation, it involves fast scanning times, and it allows for the visualisation of events outside the airway. An example of a CT scan of the upper airway is shown in Figure 2.9.

The majority of studies using a CT scan showed that there exist differences in upper airway structures and dimensions between OSA patients and control participants [134–137]. In general, the upper airway cross-sectional area has been found to be significantly smaller in OSA patients. One study observed a greater reduction in the diameter of the hypopharynx [138] and another study indicated that OSA patients have a lower retrolingual airway cross-sectional area, which is directly correlated with OSA severity [27]. Mayer et al. conducted a study using advanced CT method (cone beam computed tomography) identified that the oropharynx tended to have a more spherical structure, with a shorter lateral length for OSA patients with a higher BMI [136]. Several studies have been conducted to identify the treatment response of OA using CT scans. An experiment conducted to validate the OA response using CT scans demonstrated that the minimum cross-sectional area of the airway increased significantly with OA [139]. CT examinations have been conducted with MAD to evaluate its effectiveness [140]. These demonstrated that upper airway resistance decreases as a result of increased airway volume.

A study evaluated the effect of OA on upper airway morphology and its relationship to treatment response showed a reduced velopharynx and hypopharynx cross-sectional area [141]. However, this study was unable to identify a relationship between upper airway parameters and treatment outcomes. Tsuiki et al. indicated that OA can reduce the collapse at the velopharynx [142]. CT scanning obtained for those patients who remained awake showed a significant improvement in cross-sectional area via the use of OA treatment [143]. The treatment outcome evaluation of UPPP using CT scanning demonstrated that the upper-airway cross-sectional area doubled and led to surgical success [144]. Another study showed a similar result and demonstrated that UPPP leads to a significant increase in retropalatal space [145]. Patients who underwent UPPP for OSA showed that their responses contained a smaller airway using CT [109]. Additionally, a CT scan study showed that it could be used in the diagnosis of OSA with a high degree of accuracy [146]. The main disadvantages of CT scanning are that it is relatively expensive, it involves the inability to image the entire pharyngeal airway in a single plane, it has a short recording time, it does not allow for the recording of full-night data and there are health risks associated with the radiation exposure involved.

## 2.4.4 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a non-invasive imaging technique that can generate three-dimensional detailed anatomical images using radio waves and a powerful magnet. The main advantages of MRI compared with other imaging techniques such as CT scanning are its ability to provide more details using 3D images, a better choice of scanning plane and low risk of radiation exposure. MRI images have been used to
identify the anatomical problems in children with OSA. Fricke et al. identified that children with OSA had enlarged tonsils [147]. Arens et al. demonstrated that there is a significant structural difference in the airway of children with OSA [148]. In adults, OSA patients have more fat deposits in areas of collapse (mainly in the pharynx) [149]. Mortimore et al. showed that site-specific fat deposits could be observed in non-obese and obese OSA patients [150]. A significant increase in tongue volume was observed in OSA patients compared with control participants [26]. Additionally, significant velopharynx narrowing was observed in OSA patients [151].

Schwab et al. found there was significant larger tongue, lateral wall and soft palate volume in OSA patients [152]. Recently, real-time MRI was simultaneously recorded with several physiological signals and shown that this method could provide valuable information regarding the sites and nature of airway collapse during natural sleep. This method, however, cannot be applied to the general population [153]. Multiple levels of obstruction (mostly the combination of obstruction at the velum, oropharynx, and hypopharynx) were observed in OSA patients when MRI images were used to investigate anatomical changes in the upper airway [154]. An example of an MRI scan image of the upper airway is shown in Figure 2.9.

MRI has been used to evaluate the treatment response of CPAP and this method indicated that CPAP increased pharyngeal volume and responded well to the treatment [155]. Fleck et al. showed that patients with severe OSA benefited more when positive airway pressure treatment was used [156]. Ryan et al. indicated that UPPP achieved a high success rate based on information regarding the tongue volume and the airway ratio identified using CT scanning [157]. MRI has been used to visualise the effect of surgery on the soft palate and tongue [158]. The same study showed that MAS could be a promising alternative treatment to CPAP via the evaluation the efficacy of MAD. The performance of the MAD evaluation conducted using ultrafast MRI demonstrated that MAD could prevent pharyngeal collapse [159]. Similarly, MRI studies have shown that OA enlarges the upper airway, especially in the velopharyn-
geal region [97]. Lateral wall and tongue movement were observed in tongue retaining devices, and the results indicated a higher treatment response [160]. Additionally, a recent study indicated that remotely controlled MAD can be used in OSA patients who experience greater treatment effectiveness [161]. However, these techniques are unsuitable for the general population due to their cost, their inability to record full-night details, their lack of simultaneous PSG data, the influence of sedation on the upper airway and the challenges associated with scanning during sleep.

2.4.5 Drug-Induced Sleep Endoscopy

Drug-induced sleep endoscopy (DISE), also called sleep nasendoscopy, is a powerful tool for evaluating the anatomy of the upper airway in an OSA patient during sleep via direct visualisation. DISE can be used to assess anatomical findings during obstruction, assess the degree of obstruction, help identify the snoring mechanism, decide and plan the treatments for OSA, and assist in predicting the treatment outcome. DISE is conducted via passing a thin, flexible endoscope through one side of the nose during a drug-induced sleep. DISE can provide a reasonable snapshot of obstruction in the upper airway. Figure 2.10 shows a typical example of a DISE procedure setup, an evaluation of the upper airway and examples of DISE images of the upper airway.

Figure 2.10: Drug-induced sleep endoscopy. The left-hand panel (A) shows an example of the DISE procedure setup. The central panel (B) represents an illustration of DISE in the upper airway. The right-hand panel (C) image shows an example of DISE images of the upper airway (source: www.enttoday.org/).

DISE was used to classify the upper airway based on dynamic airflow, sound generation and the obstruction site [162]. This study also identified there was a significant reduction in the cross-sectional area of the velopharynx (resulting in palatal collapse) and at the level of the tongue [162]. The upper airway of a patient was divided into three groups based on the obstruction site: palatal, tongue base or a combination of the two [39]. In addition, another study classified the upper airway into the following: the nose, oropharynx, hypopharynx, and larynx (NOHL classification) regions based on the obstruction site identified using DISE [163]. Another important classification method of obstruction in the upper airway has been made
using a VOTE classifier, which differentiates upper airways into four levels (as per the acronym given): the velum (V); the oropharyngeal area, including the palatine tonsils (O); the tongue base (T), and the epiglottis (E) [40]. Furthermore, this study defined the severity of obstruction into three categories: no obstruction (no vibration), partial obstruction (vibration 50–75%) and complete obstruction (collapse, greater than 75%). Currently, the most commonly used classification scheme is based on the VOTE classification system under DISE. This classification is also used in choosing the most appropriate treatment and predicting treatment outcomes [40]. More recently another classification system using DISE of the upper airway was proposed [164]. This method classifies the anatomical structure of the airway into the palate, tonsils, lateral pharyngeal wall, tongue base and epiglottis, and can be utilised in surgical treatment. A study conducted with DISE and pressure catheters found that the absence of pharyngeal muscle contraction resulted in airway obstruction [41]. Truong et al. adopted a DISE in the diagnosis of OSA in paediatric patients [165]. Lan et al. analysed DISE and PSG results, along with patient demographic information to gain a better understanding of CPAP treatment responses [166].

DISE has demonstrated its potential in selecting appropriate treatment and predicting the treatment outcome using surgery [167]. DISE was used to evaluate the effectiveness of surgery by assessing the changes in upper airway pre- and post-surgery. This study classified the obstruction site as soft palate, circumferential, tonsillar-type, root-of-tongue obstruction, and a combination of palate and tongue [168]. Lan et al. observed a strong correlation between lateral wall collapse and OSA severity [169]. This study also demonstrated that surgery could be effective in patients with severe OSA. Similar results were achieved using maxillomandibular advancement procedure in severe OSA patients with lateral wall [170]. A DISE study showed that UPPP generate worse treatment outcomes for patients with palatal obstruction and tongue obstruction by comparing the airway pre- and post-treatment [110, 171]. Patients with palate collapse showed a positive outcome following single-level palatal surgery [172].

Kent et al. attempted to identify the obstruction site in OSA patients with an incomplete response to OA treatment [173]. Patients who experienced mild to moderate OSA showed substantial improvement in upper airway patency with MAD [174]. Beeck et al. conducted a study and observed that most patients showed increased airway cross-section with a positive response to MAD treatment [25]. DISE was used to predict treatment outcomes using MAD treatment for OSA and the use of this method revealed a positive response to the treatment [175]. Another recent study showed that patients with posteriorly-located tongue collapse, identified using the data from an upper airway endoscopy, exhibited a significant improvement in collapsibility with OA treatment and could predict the treatment responders with an accuracy of 82% [22]. The VOTE classification using DISE found that patients with tongue collapse experienced greater improvement in their sleep disordered breathing when treated with OAs [176]. Another study found that tongue-base collapse, identified using DISE, was found to increase the probability of successful treatment with OAs [24].

Although DISE is the most common way to assess the upper airway with a
moderate-to-substantial level of reliability, there are some major limitations associated with this approach. First, the most controversial aspect of DISE is the difference in obstruction patterns between natural and drug-induced sleep. Studies have demonstrated inconsistencies in terms of the obstruction site identified using natural sleep with DISE, as sedation leads to greater upper airway muscle relaxation [35,36]. Another study found a difference in surgical planning, which was conducted based on the difference in obstruction sites, and this study showed that there exists a notable difference between the hypopharyngeal and laryngeal structure [177]. Additionally, a study illustrated that information regarding the obstruction site identified using DISE could not provide a guarantee of treatment success [178]. Second, the depth of sedation introduces a different effect on airway muscle relaxation. The DISE protocol does recommend a set level of sedation, which results in different levels of sedation, as some people are not well-sedated at those recommended levels. Another study found that 18% of snorers do not snore and 45% of non-snorers snore during drug-induced sleep when compared with natural sleep [179]. One study indicated that inducing sedation using propofol may have a significant effect on the pattern of airway obstruction observed during DISE [180].

The third issue is based on observer variations. DISE requires an experienced human examiner for classification. Generally, inter-rater reliability shows a satisfactory level of agreement, with level of experience playing a key role in the accuracy of the DISE assessment. A study identified that junior ear, nose and throat (ENT) surgeons are prone to more errors [181]. Fourth, unlike PSG, DISE does not provide complete information regarding airway obstruction for an entire night. DISE only provides a snapshot of obstruction on the upper airway in a much shorter timescale. Therefore, DISE is unable to provide complete information regarding how an obstruction changes throughout the night and cannot determine the predominant site of collapse for a full night’s study. Finally, there is no standard grading system to classify obstruction sites and different studies define levels of collapse, such as VOTE and NOHL [40,163].

### 2.4.6 Airflow Shape Analysis

Recently, researchers demonstrated that the shape of nasal pressure waveforms can identify a site-of-airway collapse [182]. For this study, data was collected using a paediatric endoscope as well as simultaneous nasal flow and pharyngeal pressure recordings during natural sleep. The location of the collapse was classified into tongue-related, isolated palatal, and lateral walls. Epiglottic collapse was also considered, but due to its intermittent nature and the observation that it often occurred with the other classes of collapse, it was ignored. The classification was performed according to the degree of negative effort dependence (NED; reduction in peak airflow with increased respiratory effort) and the results were then compared with the endoscope analysis classification.

The study showed that the shape of inspiratory airflow could identify (with high accuracy) the pharyngeal site that resulted in collapse. The findings demonstrated that tongue-base-related obstruction can be determined with a sensitivity of 100%, a specificity of 92%, and an area under the curve of 0.96 (95% confidence interval (CI),
Furthermore, a small amount of NED (median, 19; interquartile range [IQR], 14%-25%) was found to be associated with tongue-related obstruction. The palatal collapse was identified in patients with a moderate NED (median, 45) and lateral wall collapse was also found to consist of moderate NED (median, 50). The epiglottis was associated with severe NED (median, 89) and abrupt discontinuities in inspiratory flow. Thirteen different inspiratory flow patterns corresponding to the three sites of collapse were identified and these are given in Figure 4.9. These results revealed that airflow signal with small NED associated with a flat flow limitation pattern was indicative of tongue-base related obstruction, whereas a scooped-out flow limitation pattern with large NED indicated that the tongue was not likely contributing to pharyngeal obstruction. The study also showed that 23% of patients selected for their study had multilevel obstructions.

Although the airflow shape method is considerably more convenient than methods requiring imaging or an endoscope, it has some limitations. Airflow in the above mentioned study was measured using a sealed mask and pneumotachograph simultaneously with a pressure catheter to evaluate pharyngeal pressure. Whereas these methods are not used in a standard clinical PSG procedure. Therefore, additional studies are required to find the correlation between airflow shape and NED. Another limitation was that low levels of CPAP were used to maintain a stable flow limitation. The third limitation arose due to how NED was defined and could be influenced by pharyngeal luminal pressure and, hence, can vary within a patient. Sometimes the presence of a pharyngeal catheter itself prevents apnoea from occurring and may influence the observation. Finally, the application of nasal anesthesia may have influenced the results of this study.

The above mentioned study was extended to evaluate the OA treatment response using a combination of airflow shape, age and BMI. The results indicated that palatal collapse was less likely to respond to OA therapy [21]. These results revealed that patients with palatal collapse were less likely to respond (<50% reduction in AHI) to OA therapy with a predictive accuracy of 74%. For the current study, the shape of the airflow signal during hypopnoea was used to manually identify the probable site-of-collapse. The details regarding the labelling method for the site-of-collapse are discussed in section 4.5.

Another study demonstrated that simple airflow-versus-time ("shape") features from individual breaths on an overnight sleep study can be used to quantify the severity of airflow obstruction without oesophageal catheterisation [284]. Individuals with suspected/diagnosed OSA underwent overnight PSG with gold standard airflow measures (oronasal pneumotach) and ventilatory drive (calibrated intraoesophageal diaphragm electromyogram). A multi-variable linear regression model was developed for predicting obstruction severity using the flow shape features. This model estimated obstruction severity breath-by-breath ($R^2=0.58$ versus gold-standard, $p<0.00001$; mean absolute error 22%) and the median obstruction severity across individual patients ($R^2=0.69$, $p<0.00001$; error 10%). However, this study made no attempt to identify the site-of-collapse using the flow shape.
2.4.7 Summary of Identification of Obstruction Sites

Even though various methods are available to identify an obstruction in the upper airway, the challenge in identifying the site of collapse is to find a simple, non-invasive method with minimal impact on patients during natural sleep. Building on these methods, a simple, non-contact method using snore recording during natural sleep was adopted to identify the site of collapse in the upper airway in the current study. Chapter 6 discusses an automatic classifier that predicts the predominant site of upper-airway collapse for a patient. In Chapter 7, a clustering algorithm was introduced to form clusters using the features extracted from the snore data during hypopnoea and then analyse the correlation between the clusters generated and the site-of-collapse.

2.5 Snoring Sound Characteristics

Snoring is a familiar phenomenon in the general population. Snoring is a breathing noise that occurs mostly during inspiration and sometimes occurs during expiration in the respiratory cycle. Snoring is caused by the vibration of different tissues such as the soft palate, pharyngeal lateral wall, tongue, epiglottis in the upper airway, or by the collapsible walls of the airway as turbulent air passes through. Epidemiological studies have shown that snoring is a highly prevalent breathing disorder, with the most accepted prevalence estimate being 40% in men and 20% in women, even though variability is extremely large [183, 184].

How snoring is generated is considered to be similar to speech production due to physiological similarities, particularly the fact that both are generated in the vocal tract [185]. Figure 2.11 shows a representation of the snore generation mechanism. An acoustical analysis of the speech and snore signal has shown that fundamental frequencies, formant frequencies and harmonics can be observed in both phenomena [186]. The upper airway acts as a variable acoustic filter that generates the snoring sound, which is similar to the function of the vocal tract in speech production. Depending on the shape and physical dimensions of the upper airway, different snoring sounds with diverse acoustic properties, (or formant frequencies) are produced. Therefore, snoring carries information related to the site and degree of obstruction of the upper airway. But the main difference is that the speech signal is generated by vocal cords while snoring is generated by the vibrations of pharyngeal structures. Endoscopic evaluations have shown that vibration location occurs at the soft palate, tonsils, tongue-base and epiglottis, which varies with the physical characteristics (such as shape and dimension) of the upper airway [187].

Snoring is common in the general population and does not necessarily indicate a breathing disorder requiring treatment. In contrast, irregular snoring is a clinical hallmark of OSA [44]. Snoring is then interrupted by a long silent period during which there is no breathing. This can typically last for 20–30 seconds and is followed by a loud snort and gasp as the patient tries to breathe [188]. Therefore, snoring can be regarded as acoustic physiological signals that can be measured in anyone,
Figure 2.11: Vibration of the vocal folds generates a varying airflow (A) that produces a spectrum of equally-spaced frequency peaks with a fundamental frequency (F0). This source signal is transferred to the vocal tract. The tract behaves like a variable acoustic filter (B). Resonance peaks R1 and R2 add gain to specific frequencies of the harmonic spectrum. The input signal and the vocal tract, together with the radiation properties of the mouth, face and external field, produce a sound output (C). The resonances R1 and R2 can be determined approximately from the peaks in the envelope of the sound spectrum. These peaks are called the formants (F1 and F2). Reproduced from [186].

and, furthermore, can conveniently be measured using a non-contact microphone. As discussed in section 2.1.7, the gold-standard method for diagnosis of sleep apnoea is carried out using PSG, where various sensors are attached to the body in an unfamiliar environment [70]. Unattended ambulatory PSG can be used in a familiar environment but, still, this requires multiple uncomfortable sensors attached to the body [74]. Snoring can be more simply and conveniently recorded using a low-cost microphone (the price of the microphone is significantly less compared to developing a sleep study centre and setting the PSG equipment for the sleep study) without any contact with the patient. Acoustic analyses of snore sounds have been conducted to extract snore signals from audio recordings, diagnosis of OSA, identifying the severity of OSA and to find any association with airway obstruction sites.

The main limitation of snore data analysis is that no standards are available for the measurement of snoring, and different sensors do not measure snore events
in the same manner. Although several studies have been conducted in the field of OSA using snore recordings [189–204], those studies don’t consider the effect of using different types of microphone and its placement. One study conducted to investigate the impact of the microphone position on the frequency analysis of snoring sounds. The used microphones had a flat frequency response and a similar frequency range. This study showed that the microphone position influences the frequency analysis of snoring sounds, with wider frequency range for air-coupled microphones compared to body contact microphones [281]. Another study found that distance of a microphone is more important than the sleep position [282].

Several studies have demonstrated that the acoustic characteristics of snoring are different in patients with OSA when compared to simple snorers [189–191]. Previous studies based on snore sounds have shown that algorithms can achieve reasonably high accuracy in characterising abnormal upper airway activity in OSA patients [192–196].

Several automatic snore signal extraction algorithms from audio recordings during sleep have been proposed and have achieved a high degree of accuracy. Most algorithms utilise the low-level descriptors of the snore signal (e.g., energy, Mel-frequency cepstral coefficients, spectral roll-off and entropy), exploiting various supervised (linear regression, hidden markov model,) and unsupervised (K-means clustering, Fuzzy c-means clustering) machine learning algorithms. Akhter et al. used a machine learning model to label snore sounds during REM or NREM sleep in OSA patients [192]. Studies have been conducted to identify the correlation between certain acoustic features of snoring and OSA [193–195]. Various studies have been performed in the diagnosis of OSA [28, 29, 197–199] and OSA severity classification [30, 31, 199–202] using snore signals with various classification and clustering models. Another study showed that an analysis of snore sounds in combination with gender information could provide superior performance in OSA diagnosis [203]. An OSA identification algorithm using tracheal breath sound analysis during wakefulness in PSG studies was developed [204]. In most of the above-mentioned studies, the audio recording was acquired with a microphone positioned near the patient’s bed [188–193, 195–197]. In some studies, tracheal sounds were used for the diagnosis of sleep apnoea which requires a specialised microphone to be attached to the trachea [194,198,204]. The main advantage of tracheal sound signals is that they are rich in information (i.e., they are louder and cover a wide frequency range) compared with a non-contact microphone, but tracheal-sound sensors are attached to the suprasternal notch skin using double-faced tape with an adhesive bandage over the sensor, which can be uncomfortable for patients. A few studies utilised audio recordings using the built-in sensors in mobile phones, but the performance of this method depends significantly on the phone distance, orientation, position and quality of the mobile audio sensor [199,202].

Although acoustic features extracted from respiratory sounds have been successfully employed to study the relationship between snoring and OSA, a few studies have been conducted to determine the relationship between snoring and obstruction sites. A study was conducted to examine the association between the fundamental frequency of snore sounds and obstruction site in the upper airway (soft palate, tonsils/tongue-base, combined type and the larynx), which was identified using intraluminal pressure [32]. The study showed that there exists a positive correlation
between snoring intensity and amplitude fluctuation in intraesophageal pressure. Another study utilising the fundamental and formant frequency of snoring, which was generated from different locations (determined by DISE), was conducted [33]. The results showed that the fundamental frequency and 2nd formant frequency of palatal related snore was significantly lower and could be used to distinguish between palate and non-palate snoring. Agrawal et al. compared obstruction sites, which were identified using nasendoscopy with the spectral characteristics of snoring. The results revealed that the median peak frequency feature of a snore could be used to differentiate the site-of-collapse (palate, tongue-base and epiglottic collapse) [205]. In a more recent study, maximal snoring sound intensity was found to be significantly correlated with the site of collapse and it was revealed that the mean snoring intensity of the tongue-related snore was significantly lower compared with the sound intensity of other obstruction sites [206].

### 2.5.1 VOTE Classification

Another important research project used a classifier to identify the four VOTE obstruction areas with the help of DISE [34, 207–210]. VOTE classifier differentiates the upper airways into four levels (velum (V), oropharyngeal area (O), tongue (T), and epiglottis (E)), as shown in Figure 2.12. VOTE classification has been explored by researchers in a few studies utilising different classification algorithms and signal processing techniques to categorise the snore signal.

![Figure 2.12: Site-of-collapse positions of the VOTE classification in the upper airway. Reproduced from [34].](image)

The audio signal was recorded along with the DISE videos using a microphone and was then synchronously stored in the same file. Based on the recordings, the locations of sound generation were classified by an ENT expert based on the VOTE classification. Snore segments containing a clearly identifiable single source of snoring...
sound were included. Snoring events with an unclear source of vibration or several vibration locations were excluded. Finally, the snore events labelled using VOTE classification were extracted from the audio recording. Various low and high-level descriptive audio features were then extracted for analysis of the snore signal based on the obstruction sites. Furthermore, feature selection methods were also adopted to locate the most relevant features for the classification models [34, 207].

Machine learning algorithms such as the Random Forest classifier, Support Vector Machine classifier, and Convolutional Neural Network were developed in the aforementioned studies to classify the snoring signal based on the obstruction sites, and reasonable accuracy (classification accuracy in the range of 55-78%, while the inter-rater reliability of DISE classifications is up to 86% when performed by human experts) was achieved [34, 207–210]. More detailed results will be discussed in Chapter 6 (Table 6.14). The main limitation of these studies was that snore data were recorded simultaneously with DISE or a pressure catheter, and the studies demonstrated that acoustic characteristics of drug-induced snore were significantly different from natural snoring, which limits the clinical application of the outcomes [179, 205, 211, 212]. More details of this are given in section 2.4.5. Another drawback was that all such studies were not designed to analyse the multi-level site of collapse and the investigation was based on analysing a single snore episode. Studies conducted to identify pharyngeal collapse using endoscopy, a pressure catheter and image techniques have shown that obstruction can occur at more than one site [35–38, 42, 127, 154, 179, 213–218]. Additionally, the investigation in these studies was based on analysing a single snore episode and, thus, was inadequate towards providing complete information regarding the obstruction site or towards identifying the predominant site of collapse. Ideally, therefore, airway collapsibility should be measured during a full night of natural sleep to ensure complete information regarding the site of collapse is obtained.

2.5.2 Summary of Snoring Sound Characteristics

As respiratory sounds can be recorded conveniently with a non-contact microphone without the need for any sensors attached to the body, acoustic features extracted from respiratory sounds have been successfully implemented in studies to determine the relationship between snoring and OSA. Previous studies have shown that an acoustic analysis of snoring has been successfully implemented in the diagnosis of OSA, in estimating the severity of OSA, and in determining the variation of OSA at different sleep stages. However, a limited number of studies have been conducted to examine the association between various acoustic features of snore episodes and the site-of-collapse. One important research project used a classifier to identify the four VOTE obstruction areas using an acoustic analysis of snoring episodes with the help of DISE. However, this method contains major limitations, such as the clinical relevance of obstruction sites identified during drug-induced sleep, and is not designed to analyse a multi-level site-of-collapse and a full night of sleep. The current study proposes the solution to this problem lies in using a complete set of snore signals during hypopnoea events collected during natural sleep and determining the predominant site-of-collapse for a patient. The probable site-of-collapse was determined by manual analysis of the
airflow signal shape during hypopnoea, which has been shown to correlate well with
the site-of-collapse. To the best of my knowledge, no studies have been conducted to
identify an obstruction site using a full-night sleep study during natural sleep.

2.6 Summary

This chapter surveyed the literature relating to the clinical background of OSA used
in this study. OSA is a highly prevalent and heterogeneous sleep-related breathing
disorder characterised by repetitive partial or complete obstruction of the upper air-
way. OSA can lead to adverse consequences for health, quality of life and cognitive
functioning. The conventional treatments for OSA are CPAP, OAs and surgery. A
major challenge in OSA treatment is choosing an appropriate therapy that the pa-
tient can tolerate well and that can generate positive outcomes. One of the important
indicators of the likelihood of success in treatment is the information regarding the
obstruction site in the upper-airway collapse during apnoea events. Studies have suc-
cessfully demonstrated that information regarding obstruction can be used to predict
treatment outcomes, especially for patients who have tongue-based airway collapse.
These patients are far more likely to gain significant therapeutic benefits from OAs.

Conventional methods to determine the site of collapse include the use of an endo-
scope or pressure catheter, which is generally performed during drug-induced sleep.
Unfortunately, these methods are not well-tolerated by patients due to their invasive
nature, the time/expense of the test, and the clinical relevance of the obstruction
sites identified during drug-induced sleep limiting their clinical application. Various
non-invasive imaging techniques such as CT and MRI scanning can be used as an
auxiliary tool to identify the site of collapse. These approaches, however, have their
own unique limitations, such as a short scan time, expense, and challenges in scanning
during sleep. Recently, researchers have demonstrated that the shape of nasal pres-
sure waveforms can identify the site of airway collapse. Additionally, most methods
are used during drug-induced sleep for short periods of time and, therefore, cannot
provide complete information regarding the dynamics of the obstruction site during
a full-night sleep.

As snoring data can be more conveniently recorded using simple and non-invasive
techniques, acoustic features extracted from snoring have been successfully imple-
mented in the diagnosis of OSA and in identifying OSA severity. A few studies have
been conducted to determine any correlation between various acoustic features of
snore episodes and the site-of-collapse. However, all such studies were conducted
using very few snore episodes recorded during drug-induced sleep, which results in
snoring with significantly different acoustic characteristics from natural snoring and,
therefore, does not provide complete information regarding the obstruction for full-
night sleep.

Ideally, therefore, airway collapsibility should be measured during a full night of
natural sleep using a simple, non-contact method. Additional information regarding
obstruction sites may assist clinicians in their determinations of the most appropriate
treatment for OSA.
Chapter 3

Signal Processing and Machine Learning Methods

In the previous chapter, the clinical background related to the study was discussed. The physiology of the upper airway, the most common treatments for OSA, methods to identify the site-of-collapse and its relationship with treatment outcome and the main drawbacks of current tools in identifying the site-of-collapse were discussed.

This chapter discuss the signal processing and machine learning approaches that are common across later chapters. Signal processing methods that apply to a single chapter will be discussed in the relevant chapter. This chapter begins with the details of features extracted from the nocturnal audio recording. Section 3.2 provides the details of machine learning approaches for this study. Two types of machine learning algorithms were adopted using a classification (Linear Discriminant Analysis and Gaussian mixture model classifiers) and cluster analysis (k-means clustering) approach to categorise snore signal based on the site-of-collapse. Section 3.3 discusses the nested cross-validation technique to identify the high performing features to build the model. Finally, the detailed information regarding the features extracted for the study are discussed in Section 3.4.

3.1 Feature Extraction

As the snore and speech generation share many common similarities in anatomical structure and generation mechanism, it is highly likely that acoustic features that have been successfully used in the speech and snore analysis will also be useful in identifying the obstruction site using snore data. Based on this hypothesis, ten low-level descriptive feature groups were extracted comprising of 50 features derived from each hypopnoea event which have been successfully implemented in various speech
and snore signal analysis for this study [32,34,192–194,203–206,215,244,245,248,249]. The features were chosen as they captured information about upper airway physiology. Extraction was done by using a frame size of 10 secs and a frame step of 5 secs (50% overlap), then taking the average across one hypopnoea event. For this study, I assumed that the snore during hypopnoea exhibits homogeneous behaviour, and our window length was chosen to be the minimum length of a hypopnoea event which is 10 seconds [64]. Features were extracted from the time and frequency domains [235].

**Time domain features**

**Energy:** Energy of an audio signal is considered as one of the most straight forward features. It provides a convenient representation of the variations in the amplitude of the signal and can be computed as

\[
E_n = \sum_{n=-\infty}^{\infty} [x(n)w(m - n)]^2, \tag{3.1}
\]

where \(x(n)\) is the discrete time audio signal, \(n\) is the time index and \(w(m)\) is the rectangle window and \(m\) is the window size (For the current study the width of the window was 10 seconds). Energy of the nocturnal audio recording was used in various studies to identify snoring episodes from the sounds recorded by a microphone [29,32,240,244,247], OSA diagnosis [248] and OSA severity classification [192,241].

**Zero-crossing rate (ZCR):** ZCR is a simple measure of the frequency content of a signal, which measures the number of times the signal changes sign during a given time frame. Mathematically, it is the number of times the signal passes through the zero value, divided by the duration of the frame.

\[
Z_n = \sum_{n=-\infty}^{\infty} |sgn[x(n)] - sgn[x(n - 1)]|w(m - n), \tag{3.2}
\]

where \(sgn[x(n)]\) is the signum function given by

\[
sgn[x(n)] = \begin{cases} 
1, & x \geq 0 \\
-1, & x < 0 
\end{cases} \tag{3.3}
\]
and \( w(m) \) is the rectangle window as defined in Energy. ZCR feature is commonly used for silence and snore sound episodes identification with a reasonable accuracy [29, 240, 241, 247].

**Entropy:** Entropy can be considered as a measure of abrupt changes in the energy level of an audio signal. Entropy \( H \) of the sequence \( x(n) \) is computed as

\[
H = - \sum_{n=1}^{N} E_n \log_2(E_n),
\]

(3.4)

where \( N \) is the number of frames and \( E_n \) is the ratio of the energy in a frame to the total energy of the signal. One recent study showed that entropy could be used in the site-of-collapse classification model using the snore data recorded along with DISE [34].

**Frequency domain features**

**Spectral Entropy:** The spectral entropy represents the spectral power distribution of a signal and is given by

\[
H_N = \frac{\sum_{n=1}^{N} P(n) \log_2 P(n)}{\log_2 N},
\]

(3.5)

where \( P(n) \) is the probability distribution given by \( P(n) = \frac{|X(n)|^2}{\sum_{n} |X(n)|^2} \), \( |X(n)|^2 \) is the power spectrum and \( X(n) \) is the discrete Fourier transform of the signal \( x(n) \). Spectral entropy has been used in snore signal analysis to extract snore signal from nocturnal audio recording [29, 245] and site-of-collapse classification [34].

**Formant Frequency:** Formant frequency has been successfully implemented in various acoustic analysis of snore signal applications [28–34, 185, 186, 189, 192–196, 203–206, 227, 236, 264]. Formants are peaks in the spectrum which have a high concentration of acoustic energy around a particular frequency. In speech processing, formant frequencies are defined as an acoustic resonance of the human vocal tract measured using the amplitude peak in the sound spectrum. Linear predictive coding (LPC) provides a simple way of determining the formant peaks automatically, providing signal intensity variation across its frequency range. For this study, the first three formant frequencies were extracted using an 18th-order LPC [236]. Formant frequencies can
be estimated from the all-pole model using

\[ H_z = \frac{1}{1 - \sum_{n=1}^{18} \alpha_n z^{-n}} \]  

(3.6)

where \( \alpha_n(n = 1, 2, ..., 18) \) are the LPC parameters.

Similar to the vocal tract in speech production, the upper airway acts as a variable acoustic filter in the generation of snoring sounds. Depending on the shape and physical dimensions of the upper airway, different snoring sounds with diverse acoustic properties or formant frequencies are produced. Since OSA patients have weaker dilator muscle tone and narrower UAs than non-OSA patients, the formant frequencies could differ based on the excitation location. The main reason for selecting formant features is based on the hypothesis that the formant frequencies might carry important information related to the upper airway obstruction site since they correspond to various acoustic features of the upper airway. Also, previous studies in OSA diagnosis and OSA severity using snore analysis have consistently shown that fundamental frequency and formant frequency play an important role [28–34, 185, 186, 189, 192–196, 203–206, 227, 236, 264].

**Mel Frequency Cepstrum Coefficients (MFCCs) and its first derivative:**

Mel-Frequency Cepstrum is a representation of the power spectrum obtained by transforming the audio signal through a series of steps that mimics the human cochlea using the linear cosine transform of the low power spectrum on a non-linear mel frequency scale. Conversion to mel frequency \( m \) from normal frequency \( f \) is given by

\[ m = 2595 \log_{10}(\frac{f}{700} + 1) \]  

(3.7)

Mel-frequency cepstral coefficients (MFCCs) have previously been demonstrated to be efficient features for speech analysis, diagnosing OSA using snore and various types of snore related studies [183–191, 196, 197, 197–212]. We have extracted thirteen cepstral coefficients \( (c_1, ..., c_{13}) \) and the corresponding 1st derivative using a uniformly triangular Mel filter bank \( W_k \). MFCCs are calculated by taking discrete cosine transform (DCT) of log energies using

\[ c_n = \sum_{k=1}^{N} \log_e(E(k)) \cos[n(k - \frac{1}{2}) \frac{\pi}{M}], \text{ for } n = 1, ..., K, \]  

(3.8)

where \( E(k) = \sum_n W_k |X(n)|^2 \) and \( |X(n)|^2 \) is the energy of the signal calculated from
the Power Spectrum defined in Spectral Entropy.
MFCC approximates the response of the human system more closely than any other system using the logarithmic frequency bands, and it is a perfect representation for sounds when the source characteristics are stable and consistent. MFCC features have been successfully used in speech [224,225] and snore analysis (OSA diagnosis, severity classification, and site-of-collapse classification) [32–34,185–191,196,197,197–212], and therefore, adopted for this study.

**Spectral Chroma:** Spectral chroma features give the aggregate spectral information of the entire spectrum projected onto 12 bins representing the 12 distinct semitones/tonal. Chroma features can be obtained by converting the spectrogram of the signal into a logarithmic axis (measured in pitches) from a linear frequency axis (measured in Hertz). From the spectrogram in a logarithmic scale, a time-chroma representation of the signal by combining pitch bands that correspond to the same chroma was acquired [237]. Spectral chroma is considered an important set of features in speech and music analysis, and additionally, researchers have found that this feature can be used in snore analysis [238,239].

**Spectral Centroids:** Spectral centroid represents the centre of mass of the signal spectrum. Higher spectral centroid corresponds to more signal energy concentrated in higher frequency. The spectral centroid is given by

\[
C(n) = \frac{\sum_{n=0}^{N-1} n|X_m(k)|}{\sum_{n=0}^{N-1} |X_m(k)|},
\]  

(3.9)

where \(X_m(k)\) is the Discrete Fourier Transform of signal sample from the \(m^{th}\) frame of \(x_m(n)\) and \(N\) is the length of windowed signal.

**Spectral Flux:** It measures the rate at which the spectrum is changing between two consecutive frames and is calculated by the squared difference between the normalised spectral magnitude of the two consecutive short-term window frame and is given as

\[
F_r = \sum_{k=1}^{N/2} (|X_r(k)| - |X_{r-1}(k)|)
\]  

(3.10)

**Spectral Roll-off:** It measures the “skewness” of the spectral shape. It represents
the frequency below which 85% of total spectral energy is concentrated and it takes higher values for signal with a right-skewed spectral energy. It is defined as

$$\sum_{k=1}^{K} |X(k)| \leq 0.85 \sum_{k=1}^{N/2} |X(k)|,$$

(3.11)

where $K$ is the largest bin that fulfils the above condition. Spectral centroids, flux, roll-off have been used in snore signal analysis to identify snore signal from audio recording [245] and site-of-collapse classification [34].

**Fundamental frequency and harmonic frequency:** Fundamental frequency gives the lowest frequency of the vibrating object and can be estimated based on spectrum shifting on a logarithmic frequency and taking the sub-harmonic to harmonic ratio. Harmonic frequencies are the multiples of the fundamental frequency. Fundamental frequency and harmonic frequency have been successfully implemented in snore related studies to extract snore events from audio recordings [189, 244], OSA diagnosis [28], OSA severity classification [30, 192] and site-of-collapse classification [32, 34].

### 3.2 Machine Learning

Machine learning creates algorithms that can learn from a state of data and make predictions on new data. Recently many machine learning algorithms have been implemented in the field of medicine for diagnosing and predicting the treatment outcomes of disease using various physiological signals. A wide spectrum of machine learning algorithms have been successfully implemented in the OSA research field for the diagnosis of OSA, evaluating the OSA severity, patient phenotyping, identifying the pathophysiology of OSA, choosing the most appropriate treatment and predicting the treatment outcome [34, 37, 107, 108, 121, 122, 156, 188, 192–212]. Recently, researchers started exploiting acoustic properties of the respiratory signal to build machine learning algorithms related to OSA. Typically, machine learning techniques can be classified into two broad categories: supervised learning and unsupervised learning [219].

For this study, I have adopted techniques from both categories of machine learning algorithms. A classification (supervised learning) and a cluster analysis (unsupervised learning) approach were adopted to categorise snore signal based on the site-of-collapse. A classification algorithm was adopted to train the model using the
labelled data and test to predict outcomes for new unseen data. Clustering analysis was applied to show the tendency of the data to form groups according to their similarity without any labelling information. While signal processing (wavelets [207]) and advanced non-linear machine learning methods (e.g., tensor, convolutional neural network (CNN)-based models [209], long short term memory (LSTM) [220], end-to-end, transfer learning) are currently popular in the field of machine learning, and have not considered these methods in this thesis. The main reason for not using advanced non-linear machine learning methods is that the database generated for this study is of modest size (58 patients, 1807 hypopnoea events), which is unlikely to be enough data to adequately train the high number of parameters that non-linear advanced machine learning algorithms require [285]. Another reason to use traditional linear classifiers is to provide more interpretable results than deep neural networks. Also, the linear machine learning methods used were better suited to the modest dataset due to the decreased number of parameters that needed to be calculated compared to non-linear methods. As shown in Chapters 5 and 6, the classification model achieved comparable performance to advanced learning methods which have been reported by other studies. Apart from that, the computation complexity of advance learning techniques are high and can result in long training times. The linear methods adopted are the Linear Discriminant Analysis Classifier and Gaussian Mixture Model Classifier. For the clustering analysis, k-means clustering was used to build the machine learning model.

### 3.2.1 Linear Discriminant Analysis Classifier

The Linear Discriminant Analysis (LDA) algorithm is one of the simplest and most widely used classification algorithms. The LDA classifier projects the data from a high dimensional space into a lower dimension by maximising the separability between different classes of events with linear boundaries. It is basically computed by calculating the variance of data within and between classes [221]. The LDA models each class by a Gaussian distribution. It assumes the data has a Gaussian distribution, that features are statistically independent, and that the data has the same covariance for all classes. Classification is achieved through a simple probabilistic decision-making algorithm using the maximum likelihood principle which models the conditional distribution of the data. LDA has been frequently applied in biomedical signal processing tasks and is the best linear Bayesian classifier. It produces the best estimate of posterior probabilities based on the assumption that the data shows a
Gaussian nature with perfectly known statistics [222], and common covariance matrix across classes. An example of LDA classification for a sample dataset with three classes separated by linear decision boundaries is shown in Figure 3.1. The LDA classification is achieved through a simple probabilistic decision process which models the class conditional distribution of the data $P(X = x \mid y = k)$ for each class $k$. An estimate of the predictions can be done with Bayes’ rule:

$$P(Y = k \mid X = x) = \frac{f_k(x)\pi_k}{\sum_{l=1}^{K} f_l(x)\pi_l},$$  \hspace{1cm} (3.12)

where $f_k(x)$ is the class-conditional density of $X$ in class $Y = k$ and it is modelled as a multivariate normal distribution, $\pi_k$ is prior probability of class $k$ and is estimated by the fraction of training samples of class $k$, and $\sum_{k=1}^{K} \pi_k = 1$.

LDA is modelled as a multivariate Gaussian distribution with a common covariance matrix ($\Sigma$) shared with every class and is given as

$$f_k(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}exp\left\{-\frac{1}{2}(x - \mu^*_k)\Sigma^{-1}(x - \mu^*_k)\right\}$$  \hspace{1cm} (3.13)
where \(d\) is the number of features, \(\Sigma\) is the covariance matrix and \(\mu^n\) are the class mean vectors.

For a two class problem (\(k\) and \(l\)), the decision surface can be found by the log-ratio of the class posteriors (\(\log_e \frac{P(Y=k|X=x)}{P(Y=l|X=x)}\)). This provides an equation which is linear in \(x\) and is given as

\[
\log_e \frac{f_k(x)}{f_l(x)} + \log_e \frac{\pi_k}{\pi_l} = \log_e \frac{\pi_k}{\pi_l} - \frac{1}{2}(x + \mu_k)^T \Sigma^{-1}(\mu_k - \mu_l) + x^T \Sigma^{-1}(\mu_k - \mu_l) \quad (3.14)
\]

This log-odds function provides a linear decision boundary between classes \(k\) and \(l\) in \(x\), and also provides a hyper-plane in \(p\) dimensions separating the classes, with \(p\) as the dimension of the feature space. The training data is used to estimate the parameters of the Gaussian distributions

1. \(\pi_k = \frac{N_k}{N}\) \quad (3.15)

   with \(N_k\) is the number of class \(k\) observations.

2. \(\mu_k = \sum_{y_i = k} \frac{x_i}{N_k}\) \quad (3.16)

3. \(\Sigma = \sum_{k=1}^{K} \sum_{y_i = k} \frac{(x_i - \mu_k)(x_i - \mu_k)^T}{N - K}\) \quad (3.17)

The LDA rule assigns the data to class 2, in the case that

\[
x^T \Sigma^{-1}(\mu_2 - \mu_1) > \frac{1}{2} \mu_2^T \Sigma^{-1} \mu_2 - \frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1 + \log(\frac{N_1}{N}) - \log(\frac{N_2}{N}) \quad (3.18)
\]

Otherwise, the output is classified as class 1

### 3.2.2 Gaussian Mixture Model Classifier

A Gaussian mixture model (GMM) is a probability density distribution, represented as a weighted combination of different Gaussian probability density functions (pdf). GMM can be used to model data from one of the various groups where the groups might be different from each other, but data points within the same group can be well-modelled by a Gaussian distribution [223]. The parameters for GMM models are evaluated from the training set using the iterative Expectation-Maximisation (EM)
algorithm from a well-trained prior model. For a GMM classifier model, the aim of the model training is to evaluate the parameters of the GMM, such that the Gaussian mixture density can best match the distribution of the training set. Even though the basic building block of GMM and LDA are similar, the methods differ in many aspects. A major difference of the GMM is that it models the data as a linear combination of several generative Gaussian models while LDA uses a single probabilistic distribution to describe the whole data.

Figure 3.2: GMM classification results for a sample dataset with three classes. Dots represent the trained data and crosses represent the test data. The four plots represent classification of GMMs with spherical, diagonal, full, and tied covariance matrices (source: https://scikit-learn.org/).

The GMM has become one of the most popular approaches in speech and audio analysis over the past several years [224,225]. The motivation of applying the GMM classifier in classifying the site-of-collapse problem using snore signal, is that the GMM has been demonstrated as an exceptionally powerful tool in speech and audio classification approaches especially in speaker verification and identification systems. Also, the generation mechanism of speech and snore bears a strong similarity.
A GMM density can be represented as a weighted combination of different Gaussian density functions. For an $N$-class classification, there will be GMMs $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ for each class. The mixture density for the $n^{th}$ model associated with $D$-dimensional vector $(x)$ can be represented as

$$p(x|\lambda_n) = \sum_{i=1}^{M} w^n_i p^n_i(x), \quad (3.19)$$

where $M$ is the number of mixture components; $w^n_i$ are the weights of the mixture which satisfy the condition $\sum_{i=1}^{M} w^n_i = 1$ and $w^n_i \geq 0$. The mixture density $p^n_i(x)$ is a weighted linear combination of component uni-model Gaussian density functions and is given as

$$p^n_i(x) = \frac{1}{(2\pi)^{D/2}|C^n_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu^n_i)'(C^n_i)^{-1}(x - \mu^n_i) \right\} \quad (3.20)$$

where $\mu^n_i$ is the mean vector and $C^n_i$ is the covariance matrix. The complete Gaussian mixture density is parametrised by the mixture weights, mean vectors and covariance matrices from all component densities and is represented as

$$\lambda_n = \{w^n_i, \mu^n_i, C^n_i\}, \quad i = 1, \ldots, M. \quad (3.21)$$

Usually, a covariance matrix GMM classifier can be represented with spherical, diagonal, full, or tied covariance matrices. Figure 3.2 shows an example of GMM classification for a sample dataset with three classes using different covariance matrix. For a GMM classifier, the aim of the model training is to evaluate the parameters of the GMM, $\lambda_n$ so that the Gaussian mixture density can best match the distribution of the training feature vectors. The parameters of the GMM can be estimated using the expectation-maximisation (EM) algorithm. For the given data, EM algorithm maximises the likelihood generated by each GMM, $p(x|\lambda_n)$ by an iterative procedure. Evaluation is done on a test set, in which the objective is to find the class model which has the maximum a posteriori probability for a given observation sequence. Usually GMM model calculation is made with a variable number of densities for optimisation purpose by selecting the best trade-off between the best results and the lowest density numbers in the mixture.
3.2.3  *K*-means Clustering

In unsupervised machine learning, the model draws inferences from the natural structure present within a set of data points without labels or without prior knowledge of the output values. It allows the model to work on its own to build representations of the input data and thereby discovering undetected patterns and information, and the model can be used for decision making and predicting the future inputs response [226]. Cluster analysis is one of the simplest and most commonly used unsupervised learning methods. Clustering automatically splits the dataset into different disjointed groups (clusters) by identifying the homogeneous subgroup within the data based on their similarities. The splits are made such that the data points in the same subgroup (cluster) are very similar, while data points in different clusters are very different. Figure 3.3 shows a sample dataset with three classes, clustered using a 3-means clustering algorithm.

![Figure 3.3: A sample dataset with three classes, clustered using a 3-means clustering algorithm. Plot demonstrate that the data can be cluster into three groups with three different centroids.](image)

*K*-means clustering is a simple, fast and an efficient data clustering algorithm that works iteratively to allocate each data point to one of $k$ subgroups (to the nearest cluster) based on the similarity of features provided. *K*-means is considered one of the most used clustering algorithms due to its simplicity. Moreover, it is fast, robust and uncomplicated to understand and yields the best outcomes when datasets are well distinctive from each other. The algorithm consists of two steps. The first step...
randomly selects $k$ centres, where the value $k$ is fixed in advance. The next phase of the first step assigns each data point to the nearest cluster centre (also called centroid). A distance function is used for computing the distance between each data object in a cluster and the centroid. When all the data points are assigned to one of the clusters, the first step is completed, and the initial grouping is done. In the next step, the means (centroids) for observations assigned to each cluster are recalculated. This iterative process repeats until the positions of the centroids converges. The objective of $k$-means clustering algorithm is to minimise total intra-cluster variance, in this case the squared distance function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_{i}^{(j)} - c_{j} \|^2$$

(3.22)

where $k$ is the number of clusters, $n$ is the number of cases, $c_{j}$ is the centroid of cluster $j$ and $\| x_{i}^{(j)} - c_{j} \|^2$ is the distance function. The algorithm clusters the data into $k$ groups for a predefined $k$. It consists of the following steps.

(i) Randomly select $k$ points as the cluster centroids

(ii) Assign observations to their closest cluster centroid based on the distance function

(iii) Revise cluster centroids as mean of assigned observations

(iv) Repeat steps (ii) and (iii) until convergence

### 3.3 Nested Cross Validation

Nested cross-validation is the most popular way to independently select the best parameters to train an optimal prediction model and get an estimate of an unbiased performance [252]. Nested cross validation comprises of double cross-validation loops, and it is performed to obtain a performance estimation on the training set to find the optimal hyper-parameters for the model [253]. Training data from the outer loop is split into $L$-folds to create an inner fold, as shown in Figure 3.4. The outer loop of cross-validation is used to provide the performance estimate, and the inner-loop used to select and tune the hyper-parameters of the model. Once the optimal hyper-parameters are chosen, the classification performance is evaluated on the test data. The algorithm consists of the following steps for the inner loop.
(i) Select the hyper-parameter.

(ii) Divide the training set into \( L \) ‘folds’

(iii) Train the model using \( L - 1 \) folds using the hyper-parameter and test on the remaining fold.

(iv) Repeat step iii \( L \) times until every fold serves as test set

(v) Repeat steps i - iv for every possible hyper-parameter

(vi) Evaluate the performance and select the best hyper-parameter.

Figure 3.4: Data partitioning for a nested cross-validation approach.

3.4 Feature Selection

Feature selection is a form of dimensionality reduction technique that provides an effective way to remove irrelevant (features that have no positive effect on the performance of the model), and redundant data (features that add no relevant information as they are very highly correlated with other features) in a machine learning algorithm [254]. Feature selection algorithms can be broadly classified as supervised or unsupervised, based on the type of learning method adopted. Supervised feature selection identifies the relevance of the feature by evaluating the correlation of feature with the class [255]. For this study, I have tried different feature selection methods for the classification model, including stepwise forward feature selection, stepwise backward elimination and PCA. However, results indicate that stepwise forward feature
selection outperforms the other two, and I have considered reporting only the step-wise forward feature selection method in this thesis. For unsupervised learning, data variance and separability of the features are evaluated to determine the relevance of the features.

3.4.1 Feature Selection using Nested Cross Validation

For supervised feature selection, a nested cross-validation is implemented to select the high performing features for the classification model. \( L \)-fold cross-validation was applied to the training set for a particular split in the inner loop and selected the \((L-1)\)-fold data as the training set and remaining data as the testing set. Ideally, a feature selection method would search through all the possible combinations of features and find the optimal feature set that produces the best machine learning model performance. However, this method is computationally expensive and is often impracticable. In addition, different machine learning algorithms may produce different optimal feature subsets. For this study, a step-wise selection algorithm is deployed by adding one feature at a time to build a machine learning model and the performance is determined. Feature sets were determined using a recursive feature addition process. The process starts with a single feature and finds the feature which gives the highest average accuracy (calculated from each fold) measured using the inner loop cross-validation. The next step was to update the feature set by taking the most accurate feature from the previous step and append all other features, one at a time to find the most accurate combination. This step was repeated until the inner loop cross-validation test set accuracy reached the first maximum and the feature set associated with the maximum was selected. The block diagram for the feature selection algorithm is shown in Figure 3.5. This approach is adopted for selecting the high performing features in section 5.2.5 and 6.2.5.
3.5 Summary

This chapter provides the details regarding the common signal processing and machine learning approaches used in the next chapters. Details of time and frequency domain features extracted from the nocturnal audio were discussed in section 3.1. Fifty acoustic features that have been successfully used in the speech and snore analysis, were derived from the audio signal. Section 3.2 provides the detailed information regarding two classification algorithms (LDA and GMM), a cluster analysis ($K$-means) approach, and the motivation for selecting these approaches. Section 3.3 describes the nested cross-validation technique. In the final section, the details of the feature selection algorithm and supervised feature selection method using nested cross-validation technique were given.
Chapter 4

Data Base

The previous chapter discussed the signal processing and machine learning approaches that are common across many of the chapters. The features extracted, machine learning approaches, nested cross-validation and supervised feature selection using nested cross-validation were discussed.

This chapter presents the database design process. An audio database was generated, which contains snore recordings based on various sites-of-collapse in OSA patients to tune, develop and test the various automatic detection algorithm. This chapter begins with the details of the study participants used for the study. Section 4.2 gives the details of PSG signals and audio acquisition system. The audio and PSG data were not synchronised at the time of signal acquisition, and this resulted in a delay of up to 5 minutes between the recorded audio signal and PSG data due to a drift in the sampling clocks. A pilot study was designed to estimate the time delay parameters and then used these parameters to correct the alignment of the PSG and audio signals. This is discussed in section 4.3. Section 4.4 describes the signal processing techniques and three different de-noising technique to remove noise from the raw signal. Section 4.5 provides the information regarding the labelling of the audio signal with the help of airflow signal shape, and details of labelling of simple snore and non-snore data.

4.1 Study Subjects

PSG and audio data of the patients who attended a full night sleep study using PSG for OSA diagnosis at the Sleep Investigation Unit, Royal North Shore Hospital Sydney, Australia, between July 2018 - January 2020 were collected for this
study. All the collected data received a diagnosis of OSA (AHI ≥ 5) after necessary clinical evaluations based on PSG assessment. Ethical approval for the study was obtained from the Human Research Ethics Committees at the Northern Sydney Local Health District (RESP/18/184) with site-specific approval for the Royal North Shore Hospital (AU/1/1CD7312). The study was retrospective in nature, and all patient identifiers were removed from the database. Our study corpus consists of 58 patients (aged between 25 and 91) who received a diagnosis of OSA (AHI between 7 and 90). Demographic details of the patients are given in Table 4.1.

Table 4.1: Participant characteristics (Data presented as the mean ± SD (Range)).

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>57.2 ± 14.8 (25 - 91)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, M:F(n)</td>
<td>37:21</td>
</tr>
<tr>
<td>Body mass index (kg m⁻²)</td>
<td>32.9 ± 7.9 (23 - 53)</td>
</tr>
<tr>
<td>AHI (events/hr)</td>
<td>30.1 ± 19.2 (7 - 90)</td>
</tr>
</tbody>
</table>

### 4.2 Data Acquisition

The patient’s respiratory sound signals were recorded concurrently with the routine PSG recordings using a computerised data acquisition system. Figure 4.1 shows the PSG and audio acquisition setup in the sleep study room. The sleep investigation unit consists of six sleep study room, and each sleep room consists of PSG recording system along with an installed night vision camera and audio recording system for surveillance. A microphone was attached to the ceiling about 1.5m above the patient’s bed to record the audio signal during sleep. The camera was mounted on the ceiling in the corner of the room and directed towards the bed to get a field of view covering the patient bed area and surroundings. A flexible Gooseneck Mini Shotgun Condenser Microphone (Electret-Condenser microphone; Line + gradient polar pattern; frequency range of 20Hz to 16kHz) was installed on the ceiling of the sleep study room to record respiratory sound signals. The respiratory sound signals were then passed through a mixer, amplifier (gain of 200) and a band-pass filter (cut-off frequency of DC-10kHz). The amplified audio was then sampled at a frequency of 32kHz with a resolution of 16-bits per sample using an Analogue to Digital converter (ADC). The acquired audio signals were stored as mp3 format along with PSG signal.
All monitoring and recording equipment for the sleep study were located in the observation/control room near to the sleep study room. All study preparations (e.g., patient’s preparation, PSG hookup, calibration of measurements, biological calibration), data monitoring and data acquisition were monitored by an expert sleep technician for the entire session. A sleep technician monitored the patient throughout the night from the observation room with the help of closed-circuit television with audio and live streaming of PSG data to verify that all the physiological signals were recorded correctly and to ensure that no signal was missed due to instrumentation problems or artefacts and provide assistance if required.

Figure 4.1: Schematic diagram of PSG and audio acquisition system. Sleep room consists of a PSG recording system along with an installed night vision camera and audio recording system for surveillance. A normal microphone was installed on the ceiling about 1.5m above the patient’s bed. All the recorded signals were stored in a computer using Profusion Sleep Software. All the activities and signal recordings were monitored by a sleep technician in the nearby observation room.

The diagnosis of OSA was carried out by evaluating the patient’s sleep for an entire night in the sleep laboratory with the help of PSG. The PSG instrument records various physiological signal simultaneously by attaching multiple sensors and electrodes to the body (see section 2.1.7). For the current study, PSG recording was done with
clinical PSG equipment (Somte PSG). PSG records 20 channels requiring wires attached to the patient using the AASM guidelines for standardised signal recording methodology for PSG. The PSG equipment simultaneously records physiological signals such as EEG (C4-A1, C3-A2), left and right EOG, ECG, two-channel EMG, left and right leg movements, airflow, nasal pressure, respiratory effort using abdominal and thoracic belt movements, blood oxygen saturation, and body positions to monitor sleep and evaluation of OSA. A symbolic representation of the various sensors and electrode placement in normal PSG recordings is shown in Figure 4.2. Auxiliary devices such as sound level meter (kept near the bed) were also used to capture all sounds intensity in the testing room. All the physiological data collected were then transferred to a computer for monitoring and recording. Compumedics Profusion Sleep Software was used to record the data, display multiple channel recording, manual scoring of the data, reporting and exporting of the recorded data. All the PSG signals were stored as European Data Format (EDF) [227] file and imported into Compumedics Profusion software for annotation.

![Figure 4.2: Symbolic representation of various sensors and electrode placement in standard PSG recordings (source: http://bozwell.co.uk/).](image)

Two channels were used to record the airflow signal with the help of nasal pres-
sure transducer and a thermistor, as a part of PSG. Figure 4.3 shows the thermistor and pressure transducer placement for the airflow recordings in PSG. The nasal flow was recorded to measure the nasal pressure changes during inspiration and expiration of breathing [228]. The airflow signal was recorded with a thermistor (SleepSense Thermistor), placed between nose and mouth without any contact with the body by detecting the variation in temperature (cool air flows during inspiration and warm air flows during expiration). This signal can only be used to detect the presence and absence of the airflow and is not reliable to identify hypopnoea events by measuring the airflow limitation [229]. However, the thermistor can be used in identifying apnoea.

Nasal pressure transducers can be used as an alternative method to determine the inspiratory pressure using a supraglottic or an esophageal catheter. The PSG system consists of a nasal cannula, connected to a pressure-sensitive transducer (PRO-BREATHE nasal cannula), which measures the pressure difference. Nasal pressure monitoring is more sensitive and reliable compared to thermal devices in accurately detecting airflow limitation and hypopnoeas [229]. Accordingly, the airflow signal recorded using a nasal pressure transducer was collected and analysed as we were interested in the airflow during hypopnoea events to label the signal manually.

Figure 4.3: Symbolic representation of thermistor and pressure transducer placement for the airflow recordings in PSG (source: www.ternimed.de).

The respiratory events (apnoea and hypopnoea) and the sleep stage (wakefulness, Stage 1, Stage 2, Stage 3, REM) were manually annotated by an expert sleep technician using the AASM (V2 2012) scoring rules as a part of the standard clinical
diagnosis of OSA [64]. More details were described in section 2.1.7. All the labelling was done by visual comparison of various physiological signal with the help of Compumedics Profusion Sleep Software. An example of PSG recording screen using Compumedics Profusion Sleep Software is shown in Figure 4.4.

![Example of PSG recording screen using Compumedics Profusion Sleep Software](https://www.compumedics.com.au/).

For the current study, PSG signals (airflow from the pressure transducer and loudness signal) were exported in EDF format and audio signal in mp3 format. All the annotation files were exported in Extensible Markup Language (XML) from the Compumedics Profusion software.

### 4.3 Alignment of Audio Signal with PSG

The sleep investigation unit consists of six sleep study rooms, and each room required calibration. Initial testing demonstrated that there was a time delay between PSG signal and the audio data, and synchronisation of the two signals was done before applying any further signal processing was applied as we were interested in the audio signal during hypopnoea events. For this, a semi-automatic synchronisation process was adopted by manually identifying the delay between the signals and then automatically compensating for the delay.
The delay between the PSG recordings and the audio signal was identified by conducting a pilot test in all the sleep study rooms without any patients. The PSG signals included a loudness signal using a sound level meter (TECPEL-330 Digital sound level meter) that was placed near the bed to capture all the sounds in the room. For the pilot test, the loudness signal and audio signal were recorded for a duration of 3 hours. A high-intensity sound stimulus was artificially made in the sleep study room at documented times such that it could be clearly seen as a spike in audio and loudness signal.

4.3.1 Finding the Delay in Audio Recording and PSG Signals

Once the recording was completed, the time difference between the PSG signal (loudness signal) and the audio recording was identified by visual inspection of the two signals. Figure 4.5 shows an example of a typical loudness signal recorded for the pilot test. This was achieved manually by finding the time during the sound stimulus and corresponding high amplitude in the loudness signal, then detecting the time difference between the two signals. We noticed that there was no lag between the signals during activation, and then the PSG signal started to lag during the audio recording, which resulted in the time delay between audio and PSG signal as shown in Figure 4.6. The difference increases with the recording time in all the room recordings at an average rate of 30s per hour. By looking at noisy events in actual sleep studies, this average rate of drift was confirmed in sleep studies. The increase in the time delay indicates that the delay was introduced because of unsynchronised sampling clocks resulting in clock drift.

4.3.2 Synchronisation of PSG Signal and Audio

With the correction factor determined, the audio signal was synchronised with the PSG signal using a resampling process. Resampling or sample rate conversion is the process of converting the sample rate of a given signal into a different sample rate with minimal loss of data [230].

Resampling for the audio signal was done with the help of resampling function in signal processing toolbox of MATLAB. For the pilot study, the resampling factor was determined using the average delay per hour. The resampling factor was determined as \( \frac{3600-30}{3600} = 0.9916 \). With the known resampling factor, the audio signal was resampled and verified with the duration for PSG signal. To verify the effectiveness of the resampling rate, the delay between PSG and resampled audio signal was identified.
Figure 4.5: Loudness signal recorded as part of the PSG for the pilot test. The peaks in the signal represent the artificially made high-intensity sound stimulus. The lower figure represents a high-intensity sound stimulus event. For the given event the PSG signal lags the audio signal by 47s.

by visual inspection method, as mentioned in the previous section. Results found that the resampled audio signal have the same duration with the PSG data and the audio signal was synchronised with the loudness signal with less than 1s error over the duration of the recording.

Based on the pilot study results, the audio recordings of all the patients were resampled. To cross-verify the synchronisation process, the timing of a sample of obvious data points in the loudness signal from the sleep studies were compared with the timing of audio recording. In all cases, the difference was less than 1 second, and for our purpose, this was sufficiently well synchronised.
4.4 De-noising of Raw Data

The audio signal was recorded using a general-purpose microphone that was not designed to record special sounds such as snoring and breathing. Additionally, the microphone was installed on the ceiling so that it did not interfere with the patient. Because of its remoteness from the patient, it also recorded various unwanted noises which included noise from an air conditioner. The background noise also contained hiss and hum. As the raw audio signal contained significant background noise, enhancement of raw signal was performed to improve the signal to noise ratio (SNR) [231]. We deployed three methods of noise reduction techniques, including spectral subtraction [232], multi-band spectral subtraction algorithm [233] and a band-pass filter method and identified the most effective method for the current application.

4.4.1 Spectral Subtraction Method

Spectral subtraction method is one of the simple, effective and the most popular noise reduction algorithms. Higher SNR can be achieved by subtracting an estimate of the noise spectrum from the noisy audio spectrum with an assumption that noise is stationary in nature. The spectrum of noise was estimated during a time period
with no signal from the patient, and only the background noise. This was achieved by estimating the noise from the first 1 second of the audio recording and then subtracting the noise spectrum from the whole recording. The inverse discrete Fourier transform (IDFT) of the estimated signal spectrum using the phase of the noisy signal is calculated to obtain the enhanced time-domain signal \[232\]. For a given audio signal with noise, the combined signal \(s(m)\) is given by

\[
s(m) = x(m) + n(m),
\]

where \(x(m)\) and \(n(m)\) are original audio signal without any noise and the noisy signal respectively, and \(m\) is the discrete time index. Frequency domain representation of the signal is given by

\[
S(f) = X(f) + N(f),
\]

where \(S(f)\), \(X(f)\) and \(N(f)\) represents the Fourier transforms of the noisy audio signal \(s(m)\), original signal without noise \(x(m)\) and noise \(n(m)\). The equation for Spectral Subtraction to estimate noise free signal can be expressed as

\[
|\hat{X}(f)| = |S(f)| - \alpha |\hat{N}(f)|,
\]

where \(|\hat{X}(f)|\) is an estimate of the magnitude spectrum of the original signal \(|X(f)|\), \(\alpha\) is a subtraction parameter that controls the amount of noise power spectrum subtracted from the noisy signal (\(\alpha = 1\) for full noise subtraction, \(\alpha > 1\) for over-subtraction) and \(|\hat{N}(f)|\) is the average of noise spectrum. The averaged noise spectrum is calculated from the noise only region without any signal of the audio recording (for this study, it was set to the first 1 sec of the audio recordings) and is given by

\[
|\hat{N}(f)| = \frac{1}{K} \sum_{i=1}^{K-1} |N_i(f)|,
\]

where \(|N_i(f)|\) is the spectrum of \(i^{th}\) frame, assuming there are \(K\) frames in the noise signal duration. The time domain noise free signal \(\hat{x}(\hat{m})\) is obtained by taking IDFT of estimated signal \(|\hat{X}(f)|\) and is given by

\[
\hat{x}(\hat{m}) = \sum_{k=0}^{N-1} |X(\hat{k})|e^{j\theta_Y(k)}e^{-j\frac{2\pi km}{N}},
\]

where \(\theta_Y(k)\) represents the phase of noise signal frequency \(Y(k)\).
4.4.2 Multi-band Spectral Subtraction Algorithm

Multi-band spectral subtraction algorithm is an improved version of the simple spectral subtraction method and can identify non-stationary noise signal. In this approach, the complete audio spectrum is divided into $K$ non-overlapping uniformly spaced frequency bands and independently performed the spectral subtraction in each band [233]. Noise spectra for each band is estimated and updated by adaptively smoothing the noisy signal power. The estimated noise free signal of $i^{th}$ band can be calculated as

$$|X_i(\hat{f})|^2 = \begin{cases} |S(\hat{f})|^2 - \alpha_i |\hat{N}(\hat{f})|^2, & \text{if } |X_i(\hat{f})|^2 > \beta |S(\hat{f})|^2 \\ \beta |S(\hat{f})|^2, & \text{otherwise} \end{cases}$$

(4.6)

where $k_i < \hat{f} < k_{i+1}$ and $k_i, k_{i+1}$ are the beginning and ending frequency bins of the $i^{th}$ sub band, $\alpha_i$ is the band specific over-subtraction factor ($1 \leq \alpha_i \leq 5$), and $\beta$ is the spectral floor parameter ($0 \leq \beta < 1$).

Compared to simple spectral subtraction method, this method provides an additional band subtraction factor ($\delta_i$) which gives an extra degree of control over the noise subtraction level within each band and the values are set to

$$\delta_i = \begin{cases} 0.8, & f_i \leq 1kHz \\ 1.3, & 1kHz < f_i \leq \frac{f_s}{2} - 2kHz \\ 1, & f_i > \frac{f_s}{2} - 2kHz \end{cases}$$

(4.7)

where $f_i$ is the ending (upper) frequency of $i^{th}$ frequency band and $f_s$ is the sampling frequency.

4.4.3 Bandpass Filter Method

A bandpass filter was used to attenuate the noise relative to the signal to improve the SNR. For this study, a $516^{th}$-order Finite Impulse Response (FIR) bandpass filter was implemented with a lower cut-off frequency of $60Hz$ and an upper cut-off frequency of $500Hz$, as the spectral energy of snoring is concentrated at lower frequencies [186].
4.4.4 Comparison of De-noising Methods

Figure 4.7 represents the results of various noise reduction techniques in a sample audio recording of 2 minutes duration and achieved percentage root-mean-square difference (PRD) of 60%, 55% and 53% for spectral subtraction method, multi-band spectral subtraction algorithm and bandpass filter respectively. PRD is given by,

$$100 \times \sqrt{\frac{\sum_{m=1}^{M} (\hat{x}_m - s_m)^2}{\sum_{m=1}^{M} x_m^2}},$$

where $\hat{x}$ is the de-noised version of the original signal $s$.

The visual analysis also indicates that the spectral subtraction method is more efficient in removing noise with minimal loss of information. PRD results indicate that spectral subtraction was the most effective method in removing the background noise.
noise from the raw signal. All the noise reduction method were applied on multiple
data, and the results show a similar trend as given above. Therefore, the spectral
subtraction method was adopted as the standard method of reducing the noise for
all the patient’s data in this study. This noise corrected audio signal was used for all
further processing.

4.5 Data Labelling

4.5.1 Labelling Site-of-Collapse

The conventional method for identifying the site-of-collapse involves the use of a
pressure catheter or DISE. This study has adopted an indirect process for labelling
the site-of-collapse that bypassed the conventional, invasive method. Genta et al.
proposed a method that used the shape of the airflow signal during hypopnoea to
manually identify the probable site-of-collapse and have adopted it for this study [182].
Their study revealed that the shape of the inspiratory airflow signal (flattening or
scooping of airflow contour), which is directly associated with NED could successfully
identify the site-of-collapse with high accuracy. The level of obstructions was classified
as a) pharyngeal lateral wall collapse, b) palatal collapse, and c) tongue-base related
collapse [182].

Figure 4.8 shows the anatomy of the upper airway. Epiglottic collapse was also
considered, but due to its intermittent nature, it was ignored. Thirteen different
representative examples of inspiratory flow shapes corresponding to three sites of
collapse classes (lateral wall, palate and tongue-base) were given and are reproduced
in Figure 4.9. We used this inspiratory flow shapes as the basis for manual labelling
of the site-of-collapse. More details of airflow shape analysis are described in section
2.4.6.

Sleep stages and respiratory event annotations were collected along with the PSG
signal (airflow signal and loudness signal) and the audio recordings. Apnoea events
were not considered for this study as they were scored when there is a reduction
in the airflow by 90% or more and hence typically have minimal associated snoring
[64]. Audio signal segments containing snore during hypopnoea events were manually
labelled and extracted for further processing. In the first part, each hypopnoea event
was labelled as one of three collapse regions (lateral wall, palate and tongue-base
related collapse) by visually comparing the shape (mainly flattening or scooping) of
inspiratory airflow signal using the reference pattern shown in Figure 4.9. Labelling
Figure 4.8: Upper airway physiology. The location of the nasopharynx is behind the nose; the velopharynx is behind the soft palate; the oropharynx is between the tips of the soft palate and epiglottis; and the hypopharynx is below the epiglottis. Reproduced from [280].

was only done if a close visual resemblance to one of three reference airflow shape patterns was found. Representative examples of an airflow shape indicating different types of collapse during an hypopnoea event and the corresponding audio signal are shown in Figure 4.10. Airflow signal with an inconsistent shape or an unclear shape during hypopnoea events were excluded, as shown in Figure 4.11. Labelling was done with the help of an experienced sleep physician. To validate the reliability of labelling, it was performed twice for the first twenty patients, and the labelling achieved a Cohen’s Kappa value of 0.82 (95% CI, 80%-85%) for intra-rater reliability. The detailed results are given in Table 4.2.

For the first 20 patients, we were able to label 661 hypopnoea events (65%) with snoring from a total of 1020 events. As the inspiratory airflow shape for lateral wall and palate-based collapse are similar in terms of scooping, this makes it hard to distinguish and may have resulted in labelling errors. We believe this resulted in 50 labelling errors in Table 4.2 between the lateral wall and palate. Tongue-base collapse was associated with flat inspiratory flow shape and can be easily differentiated from the other two types of collapse classes.
Figure 4.9: Identifying site-of-collapse using reference nasal pressure signals. Thirteen different representative examples of inspiratory flow patterns (darker tracing represents the ensemble average inspiratory flow). Reproduced from [182].

<table>
<thead>
<tr>
<th>Manual Labelling (1\textsuperscript{st} time)</th>
<th>Lateral</th>
<th>Palate</th>
<th>Tongue</th>
<th>Unclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lateral</td>
<td>191</td>
<td>31</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Palate</td>
<td>19</td>
<td>120</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Tongue</td>
<td>3</td>
<td>8</td>
<td>256</td>
<td>10</td>
</tr>
<tr>
<td>Unclassified</td>
<td>12</td>
<td>16</td>
<td>12</td>
<td>309</td>
</tr>
</tbody>
</table>

Table 4.2: Reliability of labelling based on the labelling the data twice.

In the next phase, the audio signal associated with each hypopnoea event was extracted. Data from all 58 patients were scored. This resulted in a respiratory sound signal database containing 1807 hypopnoea events (lateral wall collapse: 524, palate collapse: 357, tongue-base related collapse: 926). In the final phase, the predominant site-of-collapse for each patient was identified. Previous studies revealed that most of the OSA patients have a multi-level obstruction and that site-of-collapse changes dynamically over the full night’s sleep. Accordingly, a complete picture of
Figure 4.10: Characteristics example of the airflow signal during an hypopnoea event indicating a collapse and the corresponding audio signals. A, Represents a lateral wall collapse which consists of a more scooped inspiratory pattern. B, Palate collapse consists of less scooped flow pattern. C, Tongue-base collapse comprise of flat inspiratory flow pattern and can be easily differentiated from the other two. All the labelling was done using the reference pattern given in Figure 4.9. Examples A-C were taken from three different subjects during an hypopnoea event.
Figure 4.11: Characteristics example of the airflow signal during an hypopnoea event with an inconsistent or unclear shape. A, Represents airflow which does not have any resemblance to the reference patterns. B, Represents airflow which seems to be inconsistent within the hypopnoea event. Both of these were labelled as unclassified data. Examples A and B were taken from two different subjects during an hypopnoea event.
the obstruction pattern over a whole night was obtained rather than taking a few snore episodes.

A study conducted by Xu et al. in identifying predominant obstructive site using snore data collected simultaneously with DISE, defined the predominant site as the site where the percentage of obstructive events of total obstructive events $\geq 70\%$ for randomly selected ten snore episode for each patient [234]. For the current study, a similar rule implemented by Xu et al. [234] to determine the predominant site was adopted. If more than 60% of the total events were the same event type then the predominant site-of-collapse for a patient was set to the majority type, otherwise we define as a "multi-level collapse". This process resulted in 11 (19%) patients classified as the lateral wall, 10 (17%) with the palate, 32 (55%) with tongue-based and 5 (9%) with multi-level collapse. Detailed results of labelling and other demographic details are given in Table 4.3. Our manual labelling using airflow shape revealed that 33 out of the 58 patients (57%) had more than one site-of collapse. The tongue-based collapse plus either the lateral wall or the palate was involved in 23 patients, the lateral wall and the palate were involved in seven patients, and the combination of three site-of-collapse was observed in three patients.

<table>
<thead>
<tr>
<th>Predominant site of collapse based on airflow shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects characteristics</td>
</tr>
<tr>
<td>Number of subjects</td>
</tr>
<tr>
<td>Sex, M:F(n)</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>BMI (kgm$^{-2}$)</td>
</tr>
<tr>
<td>AHI(events/hr)</td>
</tr>
<tr>
<td>Hypopneas events (n)</td>
</tr>
<tr>
<td>Duration (minutes)</td>
</tr>
</tbody>
</table>

Data presented as the mean $\pm$ standard deviation unless specified. The last two rows represent the details of labelled data into three sites-of-collapse using the airflow shape. Predominant site-of-collapse for a patient was identified using majority rule on labelled hypopnoea events (more than 60% of the total events were of the same type).
4.5.2 Labelling of Simple Snore and Non-Snore

For the extraction of OSA related snoring signal from the audio recordings, an automatic classifier was implemented to identify snoring during the hypopnoea events. For the audio signal classification into non-snore, OSA related snore and simple snore (other than snoring during hypopnoeas), the audio recording was manually labelled. Manual labelling of audio recording was based on visual and auditory inspection of the audio recordings. Assistance was provided by a sleep expert to identify the snore events as there was no clear definition for a “snoring event”. OSA related snoring was defined as the snoring during an hypopnoea event for the current study, and this was manually labelled and extracted with the help of PSG annotations. Simple snoring was then defined as an episode of snoring that did not occur during an apnoea event. Simple snoring was labelled and extracted if a clear acoustic perception of snoring was apparent from the audio recording. Non-snore episodes were defined as episodes of other sounds such as normal breathing, noise from body movement, coughing, talking and other environmental noises. A typical example of labelling simple-snore and OSA related snore for an audio recording is shown in Figure 4.12. By this process, we have a database from 58 patients with 2666 hypopnoea events with snoring (average of 20 minutes per patient, ranging from 7 minutes to 50 minutes), simple snore (average of 1hr 45 minutes for each patient, ranging from 30 minutes to 3hrs. 30 minutes) and other sounds (average of 3hrs. 30 minutes for each patient ranging from 2hrs. 30 minutes to 4 hours 45 minutes).
4.6 Summary

This chapter provides the details regarding the database design process to generate an audio database which contains snore recordings based on various site-of-collapse. The database also consists of manually labelled audio recording of simple snore and other audio. A pilot study was conducted to identify the time delay details between the audio recording and PSG data, as a small delay between the recorded audio signal and PSG data was noticed. The delay was due to sampling error, and a resampling method was adopted to synchronise the audio and PSG signal which successfully resynchronised the signals. Noise reduction was also implemented to remove background noise from audio recording, and identified that spectral subtraction was the most effective method in removing the background noise. Section 4.5 describes the labelling of the site-of-collapse based on the airflow shape and the labelling of simple snore and other sounds.
Chapter 5

Extraction of Snore Events From Sound Recordings

In Chapter 4, the details regarding the database design process were given. In the current chapter, the development of an automatic algorithm is presented to classify the full night audio recording of an OSA patient as OSA related snore, simple snore and other sounds. Two algorithms were developed to extract OSA related snore using an LDA classifier based on the hypothesis that OSA related snoring can assist in identifying the site-of-upper airway collapse.

In this chapter, an automatic machine learning algorithm using an LDA classifier was deployed to extract OSA related snore events from nocturnal audio recordings. The acoustic properties of OSA related snoring could potentially be used as a non-invasive method to identify obstruction sites in OSA patients. This chapter (section 5.1) begins with the details of the existing methods that utilise nocturnal audio recording in the diagnosis of OSA and OSA severity classification. In section 5.2, the methodology adopted for this study is described, which includes study participants, the feature extraction method, data analysis, the partitioning of data for the model development, the feature selection algorithm and the performance evaluation. In the next section, the performance evaluation of the model developed based on three experiments is discussed. The results of the extraction of snore events from the audio recording are given in section 5.3.1. In the next experiment, OSA-related snore events are identified from all of the snore events, and the detailed results are given in section 5.3.2. In the final experiment, a direct method is implemented using a multi-class LDA classifier to classify nocturnal audio recordings, and the results are given in section 5.3.3. Finally, section 5.4 provides a discussion based on the results.
5.1 Existing Methods

The gold-standard method for the diagnosis of OSA is PSG. The disadvantages of the overnight PSG in an attended centre include the expense of the test, long waiting times for patients due to limited resources, and patients may not sleep well during the study due to the attachment of uncomfortable sensors. More details were provided in section 2.1.7. As snoring data can be more conveniently recorded using simple and non-invasive techniques, acoustic features extracted from snoring have been successfully implemented in the diagnosis of OSA and to identify OSA severity.

Several studies have been conducted in developing an automatic snore signal extraction algorithms from nocturnal audio recording with a high accuracy. Ambient microphone data was used in one study to classify audio signals as a snoring episode or a non-snoring episode using Hidden Markov models and spectral-based features [240]. However, this study comprises of data from 6 patients only and achieved an accuracy of 82-89% using spectral-based features of snore. Another study developed a snore-event extraction algorithm based on linear regression that stemmed from sub-band energy distributions [241]. This study achieved an accuracy of 90% when the model was trained using data from simple snorers and OSA patients. It achieved an accuracy of 87% for detecting snore episode when trained with OSA patient data.

Azarbarzin et al. adopted an unsupervised machine learning technique using a fuzzy C-means clustering algorithm to label the full night audio recording as a snore or no-snore episode, and the audio was recorded using a tracheal microphone and a non-contact microphone [242]. The results showed that the model achieved an overall accuracy of 98% and 93% using tracheal sound recording and microphone recording, respectively to extract snore events from audio recordings. Similarly, a $k$-means clustering algorithm was developed to label audio signals recorded during a full night sleep study into snore and non-snore episodes based on the clusters generated [243]. An artificial neural network (ANN) classifier was developed to extract snore sounds from a continuous sound recording using low-level descriptors of the audio signal and achieved an accuracy of 86-89% [244]. Nonaka et al. achieved a high accuracy of 96% for extracting snore events from the audio recordings using the AIM technique [245]. The data set for training consisted of OSA patients and simple snorers. A classification algorithm was also developed using a recurrent neural network to categorise audio recordings during sleep into snore and non-snore episodes and achieved an accuracy of 95% [246].

Studies have been conducted to identify any correlation between certain acoustic
features of snoring and OSA [193–195]. A study using the intra-snore-pitch-jump property of snoring found that a diagnostic tool for OSA can be used [247]. The snore signal was adopted as a screening tool for the diagnosis of OSA by extracting low-level acoustic features with a Bayes classifier [248]. An OSA severity identification algorithm was developed using an inter-snore property (apnoea phase ratio, running variance, inter-event silence) and a Bayes classifier, and the audio was recorded using a non-contact microphone [249]. Alencar et al. observed the correlation between number of irregular snores with AHI, which was identified using Hurst analysis of the snore sound [250]. A two-layer neural network was developed to automatically detect snore from a full-night audio recording using a tracheal microphone and to classify snorers based on OSA severity[251].

Researchers have previously used a two-step process to identify apnoea events for OSA diagnosis and to identify its severity, as shown in Figure 5.1. The first step is a snore detector, which extracts the snore signal from the audio recording. Snore events are then further processed to determine if they constitute an apnoea or hypopnoea event-related snore or simple snore. Using the apnoea or hypopnoea event information, OSA diagnosis and OSA severity classifications are made. In all of these studies, the performance evaluation was based on comparing the model with AHI, which was evaluated using PSG data.

![Figure 5.1: Block diagram of the conventional two-step process to extract OSA related snore.](image)

For the current study, the aim was to extract snoring during hypopnoea events, which could be beneficial in the identification of the site-of-collapse without using...
invasive technology. For the extraction of the OSA-related snore, two methods were adopted. The first method adopted the conventional two-step process by identifying OSA-related snoring from the snore detection algorithm, as shown in Figure 5.1. In the second method, an automatic system using machine learning algorithm was implemented to directly classify the audio signal recorded using a full-night sleep study into OSA-related snore (snoring during a hypopnoea event), simple snore (snoring episode not associated with an apnoea event) and other sounds (normal breathing, noise from body movement, coughing, talking and other environmental noises), as shown in Figure 5.2. This classification was carried out using an LDA classifier and the most relevant features for the model were identified using a nested cross-validation technique.

5.2 Methodology

5.2.1 Study Subjects

The data from 58 patients were used for this study described in section 4.1. All the data acquisition methods, preprocessing methods and labelling procedure were discussed in Chapter 4. For the current study, manually labelled audio recordings of all patients were used, consisting of hypopnoea events related snoring, simple snore and other sounds. More details were described in section 4.5.

5.2.2 Feature Extraction

Fifty identical features were derived from the audio recording, consisting of time and spectral features. Time-domain features consisted of (1) Energy, (2) Entropy, (3) ZCR. The frequency-domain feature consisted of (1) First three formant frequencies, (2) Thirteen MFCC and its first derivative, (3) Twelve spectral chroma features, (4) Spectral Entropy, Flux, Centroids and roll-off (5) Fundamental Frequency and Harmonic Frequency. More details of the features extracted were discussed in section 3.1. For the snore data during hypopnoeas, extraction was done by using a frame size of 10 secs and a frame step of 5 secs (50% overlap) then taking the average across one hypopnoea event. For the simple snoring and other sounds, feature extraction was done using the window of width 10s without overlapping. Based on feature extraction, features from approximately 50,000 audio events (∼30,000 other sound events; 17,000, simple snore event; ∼3000 OSA related snore events) were extracted.
5.2.3 Data Analysis

A supervised machine learning algorithm was developed to extract OSA related snore from the nocturnal audio recordings. The classification was carried out with an LDA classifier using the most relevant features identified using a nested cross-validation technique. The performance estimation was done with a nested leave-one-patient-out cross-validation. For the first method (conventional two-step process), all the snore episode were extracted from the audio signal and then the events were classified as OSA related snore and simple snore. In the second method, the automatic classifier directly classifies the audio signal into three classes (OSA related snore, simple snore and other sounds). More details of LDA classifier were discussed in section 3.2.1.

5.2.4 Data Partition

To estimate the performance of the model, supervised machine learning methods required the splitting of the data into different subset for training, validating and evaluating the final performance of the classifier. For the current study, a nested-leave-one-patient-out cross-validation was implemented to find the most relevant parameters for the classifier, develop the model, and evaluate the classifier’s final performance. The data was partitioned into two disjoint levels comprised of four subsets (S1-S4), as shown in Figure 5.3. For the level 1 partitioning, the database of 58 patients was split into two subsets called the Training Set (S1) consisting of 57 patients and the remaining one patient data set to as Test Set (S2). The outer loop was used to estimate the performance of the model using all features and selected feature sets. For the inner loop, the training set (S1) from the level 1 partitioning was again split into a training set (S3) and test set (S4) to select the high performing feature combination.

5.2.5 Feature Selection

A feature selection algorithm was developed to identify the best model with the high performing features for the current test subject using a nested cross-validation method. More details of nested cross-validation were described in section 3.3. Ten-fold cross-validation was done to S1 in the inner loop, and then select the nine-fold data as the training set (S3) and remaining data as the testing set (S4). Feature
sets were determined using a recursive feature addition process. Detailed feature selection process were given in section 3.4. The block diagram for the feature selection algorithm is shown in Figure 3.5. By this process, 58 different relevant feature sets were identified for each patient.

### 5.2.6 Performance Evaluation

For the first method (conventional two-step process), the classification algorithm labels each event as snore and non-snore, and in the next phase, the snore data is categorised as OSA related snore and simple snore. In the second method, an automated machine learning model labels each event as one of the three classes (OSA related snore, simple snore and other sounds) for a patient. The performance was evaluated for the label assignment of two cases by comparing the automatic and manually labelled data. Based on the comparison, accuracy, sensitivity and specificity were calculated. Accuracy was calculated by determining the proportion of correct predictions and is given as

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (5.1)$$
The other measures were calculated as follows:

\[
Sensitivity = \frac{TP}{TP + FN} \tag{5.2}
\]

\[
Specificity = \frac{TN}{FP + TN} \tag{5.3}
\]

\[
PositivePredictiveValue = \frac{TP}{TP + FP} \tag{5.4}
\]

\[
NegativePredictiveValue = \frac{TN}{TN + FN} \tag{5.5}
\]

where TP represents True Positive (an outcome where the model correctly predicts the positive class), TN represents True Negative (an outcome where the model correctly predicts the negative class), FP represents False Positive (an outcome where the model incorrectly predicts the positive class), and FN represents False Negative (an outcome where the model incorrectly predicts the negative class) with the confusion matrix shown in Table 5.1.

Table 5.1: Confusion matrix for the performance assessment of a system.

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>P</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted</td>
<td>P</td>
<td>TP</td>
</tr>
<tr>
<td>Class</td>
<td>N</td>
<td>FN</td>
</tr>
</tbody>
</table>

95% Confidence Interval (CI) for the accuracy is also provided for results to give an indication of precision of the results [279] and can be computed as

\[
interval = 1.96 \times \sqrt{\frac{accuracy \times (1 - accuracy))}{N}} \tag{5.6}
\]

5.3 Results

The LDA classifier was deployed to extract OSA related snore events from a full night nocturnal audio recording. A nested-leave-one-patient-out cross-validation method was adopted to identify the best model with the most relevant features for each test subjects. Two algorithms were developed to extract OSA related snore from the
audio recording. To evaluate the performance of OSA related snoring extraction, three experiments were performed using the LDA classifier, as follows:

(i) Identification of snore events from audio recordings

(ii) Extraction of OSA related snoring from all snore events

(iii) Direct method to extract OSA related snore events from the audio recording

5.3.1 Identification of Snore Events From Audio Recordings

In the first phase, the LDA classifier was deployed to classify the full night audio recording as snore events and other sounds. Snore events consisted of all OSA related snore events and all other snoring episodes. The classification performance was evaluated using all of the features extracted and the most relevant features identified using the nested cross-validation method. Classification results showed that the LDA classifier achieved an overall accuracy of 82% (95% CI, 80%-84%) for classifying snore events and other sounds using all of the available features. The detailed results are shown in Table 5.2.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>All features</th>
<th>Using feature selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>82 (80 - 84)</td>
<td>87 (86.6 – 87.3)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>86 (85 - 87)</td>
<td>90 (89.5 – 90.5)</td>
</tr>
<tr>
<td>Specificity</td>
<td>81 (80 - 81.5)</td>
<td>85.5 (85 - 86)</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>66 (65 – 67)</td>
<td>75 (74 – 75.5)</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>93 (92.5 - 93.5)</td>
<td>95 (94.5 - 95.5)</td>
</tr>
</tbody>
</table>

Next, a nested leave-one-patient cross-validation process was developed to evaluate the performance of the model using the best combination of features for each patient. Table 5.3 shows an example of a feature selection process from one set of data into training and test sets in the inner cross-validation loop (S3 and S4). The inner cross-validation accuracy increased until it reached the first maximum. This occurred when 12 features were selected; hence, the first 12 features were selected for this particular split of data to build the model.
Table 5.3: Example of a feature selection process from the inner cross-validation.

<table>
<thead>
<tr>
<th>Step</th>
<th>Feature</th>
<th>Inner cross validation accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1\textsuperscript{st} Coefficient of First derivative MFCC</td>
<td>81.9</td>
</tr>
<tr>
<td>2</td>
<td>3\textsuperscript{rd} Coefficient of First derivative MFCC</td>
<td>83.2</td>
</tr>
<tr>
<td>3</td>
<td>2\textsuperscript{nd} Coefficient of First derivative MFCC</td>
<td>83.9</td>
</tr>
<tr>
<td>4</td>
<td>Spectral Centroid</td>
<td>85.1</td>
</tr>
<tr>
<td>5</td>
<td>Spectral Entropy</td>
<td>86.0</td>
</tr>
<tr>
<td>6</td>
<td>10\textsuperscript{th} Coefficient of MFCC</td>
<td>86.7</td>
</tr>
<tr>
<td>7</td>
<td>12\textsuperscript{th} Coefficient of MFCC</td>
<td>87.4</td>
</tr>
<tr>
<td>8</td>
<td>5\textsuperscript{th} Coefficient of Chroma feature</td>
<td>87.9</td>
</tr>
<tr>
<td>9</td>
<td>Energy</td>
<td>88.1</td>
</tr>
<tr>
<td>10</td>
<td>4\textsuperscript{th} Coefficient of Chroma feature</td>
<td>88.6</td>
</tr>
<tr>
<td>11</td>
<td>2\textsuperscript{nd} Formant Frequency</td>
<td>88.9</td>
</tr>
<tr>
<td>12</td>
<td>3\textsuperscript{rd} Coefficient of Chroma feature</td>
<td>89.2</td>
</tr>
<tr>
<td>13</td>
<td>7\textsuperscript{th} Coefficient of MFCC</td>
<td>88.7</td>
</tr>
</tbody>
</table>

The classification results achieved an overall accuracy of 87% (95% CI, 86.6% - 87.3%) when classifying snore events and other sounds using the selected features from the nested cross-validation technique. Results from the feature selection outperformed the results using all of the features (see Table 5.3). Evaluated across the different cross-validation data splits, there was an average of 19 features selected, which is less than 40% of total the features extracted. The most frequently selected 19 features were the six first derivative MFCC coefficients, three MFCC coefficients, 2\textsuperscript{nd} and 3\textsuperscript{rd} formant frequency, fundamental frequency, four spectral chroma features, spectral entropy, and energy. The main reason for the low PPV may be due to the low number of snore events compared with the other sound events, which can introduce a small bias towards the other sound events.

5.3.2 Extraction of OSA Related Snoring From All Snore Events

Once the snore event had been identified, the next phase was to extract the OSA related snoring from the snore events. For this phase, the LDA classifier was deployed to classify snore events as OSA related snore and simple snore using all the features extracted and the most relevant features using the nested cross-validation
Table 5.4: Cross-validation results for the extraction of OSA related snoring.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>All features Value (% with 95% CI)</th>
<th>Using feature selection Value (% with 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>67 (66 - 69)</td>
<td>72 (71 – 74)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>69 (67 - 71)</td>
<td>73 (71 – 74.5)</td>
</tr>
<tr>
<td>Specificity</td>
<td>65 (63.5 - 67.5)</td>
<td>72 (70 - 74)</td>
</tr>
<tr>
<td>Positive Predictive Value</td>
<td>69.5 (68 – 71)</td>
<td>77 (75.5 – 78)</td>
</tr>
<tr>
<td>Negative Predictive Value</td>
<td>65 (63 - 67)</td>
<td>67 (66 - 69)</td>
</tr>
</tbody>
</table>

method. For this experiment, 50 OSA related snore events and simple snore events were selected for each patient to obtain balanced data for the classification model.

The classification results showed that the LDA classifier achieved an overall accuracy of 67% (95% CI, 66%-69%) when classifying snore events into OSA snore events and simple snore using all of the features and with an accuracy of 72% (95% CI, 71%-74%) using the selected features from the nested cross-validation technique. The detailed results are shown in Table 5.4. Evaluated across the different cross-validation data splits, there was an average of eight features selected, which is less than 20% of the total features extracted. The most commonly selected eight features were energy, ZCR, two MFCC coefficients, one first derivative of MFCC coefficient, 2nd formant frequency, one spectral chroma feature and spectral entropy. Comparing the performance of the feature selection using all of the features illustrated that feature selection resulted in a higher-performing model.

5.3.3 Direct Method to Extract OSA Related Snore Events From Audio Recordings

In this experiment, a multi-class LDA classifier was developed to classify the nocturnal audio recording into OSA related snore, simple snore and other sounds. This method is capable of directly extracting OSA related snore from audio recordings without using the conventional two-step process. The classification was performed using all of the features extracted, and the most relevant features were extracted using the nested cross-validation method. The classification results showed that the LDA classifier achieved an overall accuracy of 59% (95% CI, 57%-61%) when classifying the snore events into OSA snore events and simple snore using all of the features,
Table 5.5: Cross-validation results for the direct method to extract OSA related snore events.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>All features (% with 95% CI)</th>
<th>Using feature selection (% with 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSA Related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snore</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>59 (57 - 61)</td>
<td>63 (61 - 63)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>60 (58 - 62)</td>
<td>65 (63 - 67)</td>
</tr>
<tr>
<td>PPV</td>
<td>66 (64 - 67.5)</td>
<td>69 (68 - 70)</td>
</tr>
<tr>
<td>Simple Snore</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>60 (59 - 61)</td>
<td>64 (62 - 67)</td>
</tr>
<tr>
<td>PPV</td>
<td>49 (47 - 50)</td>
<td>52 (49 - 53)</td>
</tr>
<tr>
<td>Other Sound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>61 (59 - 63)</td>
<td>63 (61 - 65)</td>
</tr>
<tr>
<td>PPV</td>
<td>67 (65 - 67)</td>
<td>71 (69 - 73)</td>
</tr>
</tbody>
</table>

and with an accuracy of 63% (95% CI, 61%-65%) using the selected features from the nested cross-validation technique. The detailed results are shown in Table 5.5. The results demonstrated that the feature selection method using nested cross-validation performed better than using all of the features. Evaluated across the different cross-validation data splits, there was an average of 18 features selected, which is less than 40% of the total features extracted. The most commonly selected 18 features were energy, entropy, ZCR, spectral entropy, four MFCC coefficients, five spectral chroma features, 2nd and 3rd formant frequencies and three first derivative MFCC coefficients.

5.4 Discussion

An automatic classification algorithm using an LDA classifier was developed to extract OSA related snore during hypopnoeas from a full night audio recording. Two methods were adopted to label OSA related snore. The first method adopted a conventional two-step process by identifying OSA related snoring from the snore detection algorithm, while the second method directly extracted OSA related snore using a multi-class LDA classifier. The results indicate that snore events can be extracted from audio recording with an accuracy of 87% using selected features and can achieve an accuracy of 72% when identifying OSA related snore from the snore events. The direct method adopted to extract OSA related snore events achieved an accuracy of
63\% when using the feature selection algorithm. The results showed that the model achieved considerable accuracy when directly extracting the OSA related snore. Overall, both methods resulted in a similar performance with a slightly higher accuracy for the direct method compared with the conventional method. It can also be noted here that the feature selection algorithm resulted in a higher performance across all of the experiments conducted. The feature selection algorithm indicates that most of the features consisted of frequency-related features across all experiments (18 out of 19 features in the snore detection algorithm, six out of eight features when identifying OSA related snore, and 15 out of 18 features for the direct method to extract OSA related snore). As MFCC features made up nine of the 18 selected features in the snore detection algorithm, three out of eight features when identifying OSA related snore, and seven out of 18 features for the direct method to extract OSA related snore, our results indicate that MFCC features provided the most discriminating information. This result is verified through the observation that MFCCs have successfully been used by other researchers in OSA diagnosis and severity classification using audio analysis.

Table 5.6 shows the classification performance comparison for snore and non-snore events. Our current method showed that snore events could be extracted from OSA patients’ nocturnal sound recordings with an accuracy of 87\%. The result indicate that my proposed method can achieve robust results compared to the other published methods in extracting snore events from audio recordings. Although this study consists of more study participants, the microphone placement was far from patient compared to other studies. The microphone was not intended to record special sounds such as the patient’s breath. Also, advanced non-linear machine learning methods (e.g., tensor, CNN-based models, LSTM, end-to-end, transfer learning) were not considered for this study. These reasons might result in a slightly lower snore detection rate compared to other studies.

Although various studies have been conducted to diagnose OSA and its severity classification, all of these studies were based on estimating the AHI range. To the author’s knowledge, no studies have demonstrated the performance evaluation of the system when identifying OSA-related snore events. For the current study, the main objective was to extract OSA-related snore during hypopnoea events from nocturnal sound recordings. This was based on the hypothesis that OSA-related snoring can assist in identifying obstructions sites. Classification based on AHI was beyond the scope of this study. The method proposed in the current study resulted in the achievement of high accuracy when extracting OSA-related snore and gave a clear indication
Table 5.6: Classification performance comparison for snore and non-snore events.

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Subjects</th>
<th>Microphone placement (Distance from head: cm)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[240]</td>
<td>Hidden Markov Model</td>
<td>OSA patients (6)</td>
<td>20</td>
<td>82-89</td>
</tr>
<tr>
<td>[241]</td>
<td>Linear Regression</td>
<td>SS and OSA patients (30)</td>
<td>15</td>
<td>87</td>
</tr>
<tr>
<td>[242]</td>
<td>fuzzy C-means clustering</td>
<td>OSA patients (30)</td>
<td>20-30</td>
<td>93</td>
</tr>
<tr>
<td>[244]</td>
<td>ANN classifier</td>
<td>OSA patients (34)</td>
<td>40-50</td>
<td>86-89</td>
</tr>
<tr>
<td>[245]</td>
<td>AIM technique</td>
<td>SS and OSA patients (40)</td>
<td>50</td>
<td>96</td>
</tr>
<tr>
<td>[246]</td>
<td>Adaptive energy threshold</td>
<td>SS and OSA patients (40)</td>
<td>70</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Our proposed method</td>
<td>OSA patients (58)</td>
<td>150</td>
<td>87</td>
</tr>
</tbody>
</table>

that OSA-related snore events can be extracted from nocturnal sound recordings. Although this study indicates that OSA-related snore events can be extracted from nocturnal sound recordings, hypopnoea related snore events were manually extracted based on the site-of-collapse (using the airflow shape analysis as discussed in Section 4.5) was used to develop machine learning models in the coming Chapters (6 & 7).

There are several limitations to the current study. First, the audio was recorded using a simple microphone, which was not designed to record breathing sounds, and it was positioned at approximately 1.5m from the patient’s bed. This resulted in poorer sound quality than could have been achieved with a more specialised setup. Furthermore, the audio signal was selected without considering the effect of body position. Therefore, both limitations may have affected the signal quality and audio frequencies of the recording. I expect that a better performance could be achieved with a more specialised setup.
5.5 Summary

In this chapter, the development of an automatic classification algorithm using an LDA classifier to extract OSA-related snore events from a nocturnal audio recording was presented. The acoustic properties of OSA-related snoring could potentially be used as a non-invasive method to identify obstruction sites in OSA patients. Two methods were adopted to extract OSA-related snore events. The first method comprised a two-step process in which OSA-related snore events were identified from the snore events, where the snore events were extracted in the first stage from the audio recording. In the second method, the nocturnal audio signal was directly classified as OSA-related snore, simple snore and other sounds. An unbiased process using nested leave-one-patient-out cross-validation was used to choose the most relevant features and to evaluate the performance of the model. The results revealed that the snore events were extracted from audio recordings with an accuracy of 87% using the selected features and achieved an accuracy of 72% when identifying OSA-related snore from the snore events. The direct method to extract OSA-related snore events achieved an accuracy of 64% using the feature selection algorithm.

The use of this method achieved high accuracy when extracting OSA-related snore and gave a clear indication that OSA-related snore events can be extracted from nocturnal sound recordings.
Chapter 6

Automatic Classification of the Site of Airway Collapse in OSA patients

In Chapter 5, an automatic machine learning algorithm was developed to classify audio signal recorded during sleep as OSA related snore, simple-snore and other sounds. In this chapter, a novel approach for identifying the site-of-collapse in OSA patients is discussed. Two different automated classifiers using machine learning algorithms were developed to predict the predominant site-of-collapse in upper airway into four classes (“lateral wall”, “palate”, “tongue-base” related collapse or “multi-level” site-of-collapse) using the snore data recorded during hypopnoea events. The data used for this study is taken from manually labelled hypopnoea related snore events, as discussed in section 4.5.

This chapter aims to apply a supervised machine learning technique to the audio data collected in a full night sleep study to identify the obstruction site in the upper airway. The information regarding the site-of-collapse could assist clinicians in choosing a personalised or structure-specific treatment for OSA. The data for this study was taken from the audio database generated, as discussed in chapter 4. Section 6.1 begins with the details of the existing method to identify the site-of-obstruction using snore data. In the following section, the details of the method adopted for this study are discussed. This section consists of the details regarding the study subject, feature extraction, data analysis, data partitioning, feature selection procedure, and performance estimation procedure used for this study. Section 6.3 evaluates the performance results based on three experiments. In section 6.3.1, the performance evaluation of the Learning set using all the features extracted, followed by the performance evaluation of the Learning set using the nested cross-validation method to
identify the most relevant feature set is presented. In section 6.3.3, the performance results of the Hidden set using the best model identified through the first two experiments are discussed. Finally, the discussion and comparison to state-of-the-art research are given in Section 6.4.

6.1 Existing Methods

Conventional treatments for OSA include CPAP, OA therapy, and surgery. Research has shown that information regarding the obstruction site could predict the treatment outcome and assist clinicians in choosing the most appropriate treatment [20–25]. The most common methods in identifying the site-of-collapses are by using an endoscope or a pressure catheter. More details regarding the treatments for OSA and methods for identifying obstruction sites were described in section 2.2 and 2.4, respectively. An ideal test in identifying the site-of-collapse would be a simple, non-invasive and reproducible method with minimal interference for the patients during natural sleep. Previous studies have shown that acoustic analysis of snoring has been successfully implemented in the diagnosis of OSA [28,29,197–199], and estimating the OSA severity [30,31,199–202]. However, a limited number of studies have been conducted in examining the association between various acoustic features of snore episode and the site-of-collapse.

A study had been conducted to examine the association between the fundamental frequency of snore sounds and obstruction site in the upper airway (soft palate, tonsils/tongue-base, combined type and the larynx) identified using intraluminal pressure [32]. This study observed that there exists a positive correlation between snoring intensity amplitude fluctuation in intrareophageal pressure with the average fundamental frequency for the four site-of-collapse classes. The average fundamental frequency was 102.8 Hz for soft palate, 331.7 Hz for tonsils/tongue-base, 115.7 Hz for combined type, and 250 Hz for the larynx. A study utilising the fundamental and formant frequency of snore generated from different locations determined by DISE was developed [33]. The results showed that fundamental frequency and 2nd formant frequency of palate related sore was significantly lower (80 Hz and 1.8 kHz, respectively), and could be used to distinguish between palate and non-palate snore. Another study was introduced to compare the obstruction sites identified using nasendoscopy and spectral characteristics of snoring [205]. The results indicate that the median peak frequency feature of the snore can be used to differentiate the site-of-collapse. The median peak frequency was 137Hz for the palate, 1.2 KHz for
tongue-base, and 490 Hz for epiglottic collapse. A recent study found that maximal snoring sound intensity was significantly correlated with site-of-collapse and showed that mean snoring intensity of tongue related snore was lower compared to other obstruction sites [206].

Another important research project used a classifier to identify the four VOTE obstruction areas using acoustic analysis of snoring episodes with the help of DISE [34, 207–210]. However, this method has major limitations such as inconsistency of obstruction site identified with natural sleep and drug-induced sleep, and was not designed to analyse the multi-level site-of-collapse and full night study. More details were described in section 2.5.1.

Figure 6.1: Block diagram of the proposed model. The model consists of two stages: 1) Developing phase, where the model is developed with Learning Set using different classifiers (GMM and LDA) and feature selection algorithms, then identifying the best model. 2) Testing phase, where the performance of the final model is estimated using the Hidden Set.

Building on these studies, our novel approach used the entire set of snoring during
hypopnoea events collected during natural sleep to predict the predominant site-of-collapse in the upper airway. For the current chapter, two classification algorithms were developed that categorise the obstruction site in the upper airway and identify the best approach by comparing the results. The block diagram of the proposed system for predicting the predominant site-of-collapse in OSA patients is shown in Figure 6.1. The proposed system consists of several steps to build the database (discussed in Chapter 4) and provide prediction results based on a model implemented. The model consists of developing and testing phase using the data taken from the database.

6.2 Methodology

6.2.1 Study Subjects

For this study, the audio recordings of 58 patients with 1807 hypopnoea events were considered. Labelling information regarding hypopnoea events and predominant site-of-collapse for all patients were also selected from the database. Details of the data used for this study is given in Table 6.1 and the detailed characteristics of the patients and labelling were given in Table 4.3.

<table>
<thead>
<tr>
<th>Subject characteristics</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>58 (Lateral: 11, Palate: 10, Tongue: 32, Multilevel: 5)</td>
</tr>
<tr>
<td>(Predominant site-of-collapse)</td>
<td></td>
</tr>
<tr>
<td>Hypopnoea events</td>
<td>1807 (Lateral: 524, Palate: 357, Tongue: 926)</td>
</tr>
<tr>
<td>Duration (minutes)</td>
<td>849 (Lateral: 259, Palate: 176, Tongue: 414)</td>
</tr>
</tbody>
</table>

6.2.2 Feature Extraction

Fifty identical features were derived from each hypopnoea event, consisting of time and frequency features. The time-domain features: (1) Energy, (2) Entropy, and (3) ZCR. The frequency-domain feature consisted of (1) First three formant frequencies, (2) Thirteen MFCC and its first derivatives, (3) Twelve spectral chroma features, (4) Spectral Entropy, (5) Spectral Flux, (6) Spectral Centroids, (7) Spectral roll-off, and
6.2.3 Data Analysis

A supervised machine learning algorithm was developed to classify the snoring signal during hypopnoea events into three sites-of-collapse (“lateral wall”, “palate”, “tongue-base” related collapse) and then predicting the predominant site of upper airway collapse for a patient into four classes (“lateral wall”, “palate”, “tongue-base” related collapse or “multi-level” site-of-collapse) based on the classification results. For this, automated multi-class classifiers (GMM and LDA) were developed to build the model and test the model on new unseen data. An unbiased nested cross-validation method was developed to select the high performing feature set. Detailed information regarding the LDA and GMM classifiers were given in section 3.2.1 and 3.2.2, respectively.

6.2.4 Data Partitioning

To estimate the performance of the model, supervised machine learning methods required the splitting of the data into different subsets for training, validating and evaluating the final performance of the classifier. For this study, a nested-leave-one-patient-out cross-validation was implemented to find the optimal parameters for the classifier, to develop the model and to evaluate the final performance of the classifier. The data was partitioned into three disjoint levels comprised of six subsets (S1-S6), as shown in Figure 6.2. For level 1 partitioning, the database of 58 patients was split into two subsets called the Learning Set and the Hidden Set. The Learning Set (S1) consisted of the first 45 patients (77%) data (9 lateral wall, 7 palate, 25 tongue-base and 4 multi-level collapse patient), was used to develop the classifier model using the signal processing and machine learning techniques. The Hidden Set (S2) comprised of the remaining 13 patients (23%) data (2 lateral wall, 3 palate, 7 tongue-base and 1 multi-level collapse patient) to validate the final model performance, providing a prospective generalisation performance of the system.

For the next level of partitioning, a nested cross-validation method was applied to the data from S1 to build and evaluate the performance of the model using the high performing feature set. The Learning Set performance was determined using a leave-one-patient-out cross-validation process from the outer loop using training sets (S3) and test set (S4). The outer loop was used to estimate the performance of the
model using all-features and selected feature set. For the inner loop, the training set (S3) from level 2 partitioning was again split into a training set (S5) and test set (S6) to select the high performing feature combination. Training set and testing set in the inner loop consisted of randomly selected 40 patients data and 4 patients data, respectively.

### 6.2.5 Feature Selection

In this study, a nested-leave-one-patient-out cross-validation method was developed using the Learning Set (S1) to identify the best model with the high performing features for the current test subject. The inner loop of cross-validation was implemented with 11-fold cross-validation. More details of nested cross-validation and feature selection process were given in section 3.3 and 3.4, respectively. The block diagram for the feature selection algorithm is shown in Figure 3.5. By this process, 45 different high performing feature sets were identified for each patient in Learning Set (S1). To estimate the performance of the final generalised model in the Hidden Set (S2), a high performing feature set was identified by selecting the most commonly selected features from the nested cross-validation method.
6.2.6 Performance Evaluation

The automated machine learning model labelled each hypopnoea event for a patient into one of the three sites-of-collapse. Based on the hypopnoea labelling, the predominant site-of-collapse for each patient was determined by identifying the most common type of collapse when more than 60% of the total events were of the same type. Although the main aim of this study was to identify the predominant site-of-collapse, the classification results of categorising the site-of-collapse are also provided as it is of interest to sleep physicians to know the sites-of-collapse for individual hypopnoeas. The performance was evaluated for the label assignment of two cases by comparing the automatic and manually labelled data using 1) the labelling of hypopnoea events, and 2) labelling the predominant site-of-collapse for each patient. Accuracy was calculated for the two cases by determining the proportion of correct predictions and is given as

\[
\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total number of predictions made}}
\]  

(6.1)

95% CI is also provided and can be computed using the equation 5.6. Also, the unweighted average recall (UAR) was computed to provide a clear information regarding the classification error, especially in an imbalanced dataset (the number of observations is not the same for all the classes) [34]. UAR can be evaluated by taking the average of the recall for each class and is given by

\[
\text{UAR} = \frac{\sum_{\text{class}=1}^{N} N_{\text{correct class}}/N_{\text{all class}}}{N} \times 100\%
\]  

(6.2)

where \(N_{\text{correct class}}\) and \(N_{\text{all class}}\) are the number of correctly recognised segments, and all segments in one certain \(\text{class}\) respectively, \(N\) is the total number of classes.

6.3 Results

Two multi-class classifiers (GMM and LDA) were deployed for categorising the patients based on the predominant site-of-collapse by snore signal analysis and identified the best performing system using the high performing feature set. It is crucial to find the performance difference between training and test data because a high difference in performance may indicate over-training. Therefore, training data (Learning Set) was used to provide an unbiased evaluation of a model while tuning model hyperparameters and evaluated the performance of the final model on test data (Hidden
set). For evaluating the performance of the Learning set and to arrive at the best final system, three experiments were performed using the two classifiers. The first two experiments were done to evaluate the Learning Set performance by using all the features extracted and using the high performing features selected through the nested cross-validation technique. In the final experiment, the performance of the Hidden set using the best classifier (identified based on the first two experiments performance), was evaluated. The steps were

(i) Leave-one-out cross-validation on Learning Set using all features

(ii) Leave-one-out nested cross-validation on Learning Set using feature selection

(iii) Validation of the best model on the Hidden Set

6.3.1 Leave-One-Out Cross-Validation on Learning Set Using All Features

In the first phase, the Learning set (S1) performance was evaluated using all of the features extracted from the snore data and compared the performance of LDA and GMM classifiers. For this experiment, leave-one-patient-out cross-validation was opted to assess the model performance of the Learning Set (S1), where the model was trained using the data from 44 patients (S3) and tested on the remaining one patient (S4) and the process iterated across all the patients.

The performance evaluation showed that GMM classifier achieved an overall accuracy of 48% with UAR of 44%, while LDA classifier achieved an accuracy of 54% with UAR of 46%. Performance estimated based on predominant site-of-collapse showed that GMM classifier achieved an overall accuracy of 49% (95% CI, 39%-61%) for classifying patients into one of the four classes (Lateral, Palate, Tongue and Multi-level), whereas LDA classifier achieved an accuracy of 56% (25/45) (95% CI, 42%-71%). Clinical studies have been successfully demonstrated that the information regarding tongue-base collapse in the upper airway predicts oral appliance treatment success and gain a large therapeutic benefit from OAs. Therefore, the performance of the system for classifying tongue related and non-tongue related collapse was also evaluated. The results showed an overall accuracy of 71% (32/45) (95% CI, 56%-84%) and an accuracy of 66% (30/45) (95% CI, 51%-79%) for LDA and GMM classifier. The detailed results are shown in Tables 6.2 - 6.4.
Table 6.2: Confusion matrix for the hypopnoea event based site-of-collapse classification using all features on the Learning Set (S1). S1 consists of 45 patients data with 1477 (Lateral wall: 447, Palate: 270, Tongue-base: 760) labelled hypopnoea events.

<table>
<thead>
<tr>
<th>LDA Classifier</th>
<th>GMM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Automatic Classification</strong></td>
<td><strong>Automatic Classification</strong></td>
</tr>
<tr>
<td>Lateral</td>
<td>Palate</td>
</tr>
<tr>
<td>Manual: Lateral</td>
<td>124</td>
</tr>
<tr>
<td>Manual: Palate</td>
<td>56</td>
</tr>
<tr>
<td>Manual: Tongue</td>
<td>132</td>
</tr>
</tbody>
</table>

Table 6.3: Confusion matrix for the predominant site-of-collapse using all features on the Learning Set (S1). S1 consists of first 45 patients data: 9 lateral wall (L), 7 palate (P), 25 tongue-base (T) and 4 multi-level (M) collapse patient.

<table>
<thead>
<tr>
<th>LDA Classifier</th>
<th>GMM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Automatic Classification</strong></td>
<td><strong>Automatic Classification</strong></td>
</tr>
<tr>
<td>Lateral</td>
<td>Palate</td>
</tr>
<tr>
<td>Manual: L</td>
<td>2</td>
</tr>
<tr>
<td>Manual: P</td>
<td>1</td>
</tr>
<tr>
<td>Manual: T</td>
<td>0</td>
</tr>
<tr>
<td>Manual: M</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6.4: Confusion matrix for the predominant site-of-collapse (non-Tongue/Tongue) using all features on the Learning Set: 20 Non-tongue and 25 tongue-base collapse patient.

<table>
<thead>
<tr>
<th>LDA Classifier</th>
<th>GMM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Automatic Classification</strong></td>
<td><strong>Automatic Classification</strong></td>
</tr>
<tr>
<td>Non-tongue</td>
<td>Tongue</td>
</tr>
<tr>
<td>Manual: Non-tongue</td>
<td>12</td>
</tr>
<tr>
<td>Manual: Tongue</td>
<td>5</td>
</tr>
</tbody>
</table>
6.3.2 Leave-One-Out Nested Cross-Validation on Learning Set Using Feature Selection

In this experiment, a nested leave-one-patient cross-validation process to evaluate the performance of the model using the best combination of features for each patient was developed. Table 6.5 shows an example of a feature selection process from one split of data into training and test sets in the inner cross-validation loop (S5 and S6). The inner cross-validation accuracy increased until it reached the first maximum after selecting eight features, and the first eight features were selected for this particular split of data to build the model.

Table 6.5: Example of a feature selection process from the inner cross-validation.

<table>
<thead>
<tr>
<th>Step</th>
<th>Feature</th>
<th>Inner cross validation accuracy ( %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1\textsuperscript{st} Coefficient of MFCC</td>
<td>59.4</td>
</tr>
<tr>
<td>2</td>
<td>1\textsuperscript{st} Formant Frequency</td>
<td>61.6</td>
</tr>
<tr>
<td>3</td>
<td>6\textsuperscript{th} Coefficient of Chroma feature</td>
<td>63.0</td>
</tr>
<tr>
<td>4</td>
<td>4\textsuperscript{th} Coefficient of MFCC</td>
<td>65.7</td>
</tr>
<tr>
<td>5</td>
<td>2\textsuperscript{nd} Coefficient of First derivative MFCC</td>
<td>67.1</td>
</tr>
<tr>
<td>6</td>
<td>12\textsuperscript{th} Coefficient of MFCC</td>
<td>68.0</td>
</tr>
<tr>
<td>7</td>
<td>Energy</td>
<td>68.4</td>
</tr>
<tr>
<td>8</td>
<td>4\textsuperscript{th} Coefficient of Chroma feature</td>
<td>68.8</td>
</tr>
<tr>
<td>9</td>
<td>10\textsuperscript{th} Coefficient of MFCC</td>
<td>68.1</td>
</tr>
</tbody>
</table>

Classification performance of categorising site-of-collapse based on hypopnoea events showed that the GMM classifier achieved an accuracy of 53% with UAR of 49% and an accuracy of 63% with UAR of 57% for LDA classifier. Performance estimation for the predominant site-of-collapse for all the classes achieved an overall accuracy of 56% (23/45) (95% CI, 42%-71%) and an accuracy of 67% (30/45) (95% CI, 54%-81%) with GMM and LDA classifier respectively. Tongue and non-tongue classification (predominant site-of-collapse) resulted in an overall accuracy of 71% (32/45) (95% CI, 55%-83%) for GMM classifier and an accuracy of 82% (37/45) (95% CI, 69%-91%) for LDA classifier. The detailed results are shown in Tables 6.6 - 6.8.

The results were achieved with an average of 7.1 features for all patients (less than 15% of total features) for LDA classifier and with an average of 11.2 features classifier (less than 25% of total features) for GMM. The high performing features from the
Table 6.6: Confusion matrix for the hypopnoea event based classification using feature selection on the Learning Set (S1). S1 consists of 45 patients data with 1477 (Lateral wall: 447, Palate: 270, Tongue-base: 760) labelled hypopnoea events.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>LDA Classifier</th>
<th>GMM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Automatic Classification</td>
<td>Automatic Classification</td>
</tr>
<tr>
<td></td>
<td>Lateral</td>
<td>Palate</td>
</tr>
<tr>
<td>Lateral</td>
<td>224</td>
<td>39</td>
</tr>
<tr>
<td>Palate</td>
<td>48</td>
<td>88</td>
</tr>
<tr>
<td>Tongue</td>
<td>90</td>
<td>65</td>
</tr>
</tbody>
</table>

Table 6.7: Confusion matrix for the predominant site-of-collapse using feature selection on the Learning Set (S1). S1 consists of first 45 patients data: 9 lateral wall (L), 7 palate (P), 25 tongue-base (T) and 4 multi-level (M) collapse patient.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>LDA Classifier</th>
<th>GMM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Automatic Classification</td>
<td>Automatic Classification</td>
</tr>
<tr>
<td></td>
<td>Lateral</td>
<td>Palate</td>
</tr>
<tr>
<td>L</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6.8: Confusion matrix for the predominant site-of-collapse (non-Tongue/Tongue) using feature selection on the Learning Set: 20 Non-tongue and 25 tongue-base collapse patient.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>LDA Classifier</th>
<th>GMM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Automatic Classification</td>
<td>Automatic Classification</td>
</tr>
<tr>
<td></td>
<td>Non-tongue</td>
<td>Tongue</td>
</tr>
<tr>
<td>Non-tongue</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Tongue</td>
<td>3</td>
<td>22</td>
</tr>
</tbody>
</table>

The model depend on the average feature number by identifying the most commonly selected feature selected in each split of testing. The most commonly selected features selected for the classifiers are shown in Table 6.9. The most commonly selected seven features selected for the LDA classifier were three MFCC coefficients, the two first...
Table 6.9: Most commonly selected features selected for the classifiers.

<table>
<thead>
<tr>
<th>Most Repeated Features</th>
<th>LDA Classifier</th>
<th>GMM Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 MFCC Coefficients</td>
<td>2 MFCC Coefficients</td>
<td></td>
</tr>
<tr>
<td>2 First derivatives of MFCC Coefficients</td>
<td>2 First derivatives of MFCC Coefficients</td>
<td>Fundamental Frequency</td>
</tr>
<tr>
<td>Fundamental Frequency</td>
<td>Fundamental Frequency</td>
<td></td>
</tr>
<tr>
<td>1 Spectral Chroma Feature</td>
<td>4 Spectral Chroma Feature</td>
<td>Energy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd Formant Frequency</td>
</tr>
</tbody>
</table>

derivatives of MFCC coefficients, fundamental frequency and the one chroma feature were selected. For the GMM model, the high performing feature set consists of 11 features: two MFCC coefficients, the two first derivatives of MFCC coefficients, energy, fundamental frequency, the 2nd formant frequency and the four chroma features.

Comparing the performance of LDA and GMM classifier, the LDA classifier outperformed the GMM classifier results in all the confusion matrices. The computation complexity was much higher in GMM classifier compared to LDA classifier, as it required complex methods to find the optimum number of Gaussian mixtures. Also, choosing the form of the covariance matrix (spherical, diagonal, full) requires a systematic approach to find the best-suited matrix for the model. Although a full covariance matrix was expected to perform best in general, it can result in over-fitting on small datasets and provide poor generalisation performance on new data. Based on these observations, LDA model using the high performing features was selected as the best system for the classification of obstruction sites using snore data and evaluated its performance on the Hidden Set.

6.3.3 Validation of the Best Model on the Hidden Set

Based on the learning set results, the best performing system was the LDA with the features selected. In the final experiment, the most frequently selected seven features were determined from the nested cross-validation process to build the best performing model using LDA classifier and prospectively evaluated the performance with the Hidden Set data. Table 6.10 shows the selected features for the final model. For the final model, the system was retrained using the selected features in Table 6.10, using all the data in the Learning Set and validated the performance of the
model in the Hidden Set.

Table 6.10: Most commonly selected features from the nested cross-validation process.

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4th Coefficient of MFCC</td>
</tr>
<tr>
<td>2</td>
<td>4th Coefficient of Chroma feature</td>
</tr>
<tr>
<td>3</td>
<td>1st Coefficient of First derivative MFCC</td>
</tr>
<tr>
<td>4</td>
<td>9th Coefficient of MFCC</td>
</tr>
<tr>
<td>5</td>
<td>3rd Coefficient of First derivative MFCC</td>
</tr>
<tr>
<td>6</td>
<td>Fundamental Frequency</td>
</tr>
<tr>
<td>7</td>
<td>10th Coefficient of MFCC</td>
</tr>
</tbody>
</table>

Performance evaluation showed that the model achieved an overall accuracy of 67% with UAR of 66% site-of-collapse based on hypopnoea events. Also, the model achieved an overall accuracy of 85% (11/13) (95% CI, 59%-98%) for tongue/non-tongue classification and an accuracy of 69% (9/13) (95% CI, 48%-90%) for the predominant site-of-collapse with all the sites, which is in-line with the Learning Set cross-validation performance. Table 6.11 - 6.13 show the detailed results. The Hidden set performance clearly indicates that the feature selection method (selecting the features up-to the first maximum inner cross-validation accuracy and identifying the most repeated features) performs well on unseen data.

Table 6.11: Confusion matrix for the hypopnoea event based site-of-collapse classification using selected features on the Hidden Set (S2). S2 consists of 13 patients data with 330 (Lateral wall: 77, Palate: 87, Tongue-base: 166)labelled hypopnoea events.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>Automatic Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lateral</td>
</tr>
<tr>
<td>Lateral</td>
<td>53</td>
</tr>
<tr>
<td>Palate</td>
<td>19</td>
</tr>
<tr>
<td>Tongue</td>
<td>16</td>
</tr>
</tbody>
</table>
Table 6.12: Confusion matrix for the predominant site-of-collapse using selected features on the Hidden Set (S2). S2 comprise of 13 patients data: 2 lateral wall, 3 palate, 7 tongue-base (T) and 1 multi-level collapse patient.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>Automatic Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lateral</td>
</tr>
<tr>
<td>Lateral</td>
<td>1</td>
</tr>
<tr>
<td>Palate</td>
<td>1</td>
</tr>
<tr>
<td>Tongue</td>
<td>0</td>
</tr>
<tr>
<td>Multi-level</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6.13: Confusion matrix for the predominant site-of-collapse (Tongue/non-Tongue) using selected features on the Hidden Set.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>Automatic Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Tongue</td>
</tr>
<tr>
<td>Non-Tongue</td>
<td>4</td>
</tr>
<tr>
<td>Tongue</td>
<td>0</td>
</tr>
</tbody>
</table>

6.4 Discussion

Motivated by the need for a simple, non-contact and inexpensive technology to identify the site-of-collapse in the upper airway, supervised machine learning models were developed using the snore signal during natural sleep to identify the site-of-collapse. In particular, the proposed method was based on an automated classification algorithm to determine the site-of-collapse during hypopnoea events and thereby predict the predominant site-of-collapse for a patient. The results demonstrate that the acoustic properties of hypopnoea related snoring sound during natural sleep varies in association with site-of-collapse, therefore potentially enabling snoring sounds to be used as an auxiliary method to identify the predominant site-of-collapse in OSA patients. The information regarding the predominant site-of-collapse could be utilised as an important factor in selecting the appropriate treatment for OSA patients.

Comparing the results from the first two experiments, the feature selection method using the nested cross-validation process has resulted in a higher-performing model compared to the method using all of the features. Analysis of the confusion matrix in Tables (6.2, 6.3, 6.6 and 6.7) provides some insight into the sources of error. As the number of tongue-base hypopnoea events contributes more than half of the total
events, there seems to be a small bias toward tongue-base collapse in the confusion matrix for hypopnoea event classification (Tables 6.2 and 6.6) and in the confusion matrix for the predominant sites-of-collapse (Tables 6.3 and 6.7), and this might be the reason for the misclassification of other classes into tongue-base. The scooping pattern for the palate and lateral wall collapse was similar, making them hard to differentiate, and this contributes to the misclassification of palate labelled patients as lateral wall and vice versa in Tables (6.2, 6.3, 6.6 and 6.7).

Feature selection for both the classifiers indicated that the best-suited feature set mainly consists of frequency-related features as shown in Tables (6.9 and 6.10). MFCCs tend to play an important role in obstruction site classification, which have been successfully used in speech [224,225] and snore analysis [183–191,196,197,197–212]. MFCC approximates the response of the human system more closely than any other system using the logarithmic frequency bands and it is a perfect representation for sounds when the source characteristics are stable and consistent. In this study, MFCC has shown an important role in obstruction site classification, which confirms the hypothesis that the characteristic of snore signal varies depend on different excitation locations.

A few studies have been conducted in examining the relationship between snoring and the site-of-collapse. Some of these studies were focused on finding the correlation between certain acoustic features of snoring and site-of-collapse. A comprehensive study was conducted using various features and classifiers to compare the performance of classifying obstruction site based on VOTE recognition. This study demonstrated that VOTE classification can achieve an UAR of 78% (on test data) implemented by a random forest classifier with 374 features selected from 4000 features using the RelieF algorithm [34]. A support vector machine learning classifier was proposed using Wavelet features extracted from snore segments to discriminate VOTE achieved UAR of 71% processed by a subject independent 2-fold cross-validation [256]. Another study utilised low-level texture image descriptors such as local binary patterns and histogram of oriented gradients as for audio-based snore sound classification into VOTE was developed and achieved an UAR of 38% and 72% for development and test split respectively with 828 snore events [208]. A deep CNN using 2246 deep spectrum features extracted from snore events, was implemented and achieved an UAR of 57% and 66% for development and test sets respectively [209]. An SVM classifier was developed using low-level descriptors for VOTE classification and resulted in UAR of 56% [210]. Even though the proposed method developed cannot be directly compared to these studies as there exists a large number of basic differences in terms of data
Table 6.14: Performance comparison of the state-of-the-art work achieved by models (on the test set) in published work. All the studies except our proposed method are based on VOTE classification.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Method</th>
<th>Accuracy (UAR, %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[34]</td>
<td>Random forest classifier and RelieF algorithm</td>
<td>78</td>
</tr>
<tr>
<td>[256]</td>
<td>SVM classifier with wavelet transform features</td>
<td>71</td>
</tr>
<tr>
<td>[208]</td>
<td>SVM classifier using low level image texture features</td>
<td>72</td>
</tr>
<tr>
<td>[209]</td>
<td>CNN</td>
<td>66</td>
</tr>
<tr>
<td>[210]</td>
<td>SVM classifier with low level descriptors</td>
<td>56</td>
</tr>
<tr>
<td><strong>Our</strong></td>
<td>Proposed model</td>
<td><strong>67</strong></td>
</tr>
</tbody>
</table>

collection, data extraction, manual labelling, and levels of obstructions, my method achieved a similar result with UAR of 67% using a simple classification algorithm and with very few features selected to build the model. The comparison results (test data) based on UAR for various studies discussed, and the current system is given in Table 6.14.

A major drawback of the studies discussed above was that snore data were recorded simultaneously with DISE or pressure catheter, and studies have demonstrated that acoustic characteristics of drug-induced snore are significantly different from natural snoring [179, 205, 211, 212], which limits the clinical application of previous studies. More details were discussed in section 2.4.5. Another limitation was that all these studies were based on analysing a single snore segment and thus were unable to determine the predominant site-of-collapse across an entire night of sleep. Moreover, these studies were not designed to analyse the possibility of collapse in multiple sites for a patient. Studies conducted in identifying pharyngeal collapse using endoscopy, pressure catheter and image techniques have shown that obstruction can occur at more than one site [35–38, 42, 127, 154, 179, 213–218]. A study conducted using DISE showed that 68% (out of 1249) patients have multi-level sites-of-collapse which was mainly the combination of palate and tongue base collapse (26% of total patients) [257]. A study based on MRI images showed that all patients with severe OSA had multiple sites-of-collapse, whereas 20% of mild OSA patients also have multi-level site-of-collapse [257]. Another study using DISE and airflow revealed that 23% of patients selected for their study had multi-level obstructions [182]. Corso et al. found that the majority of patients had multi-level collapse (26/28, 92.8%) predominantly
at the palatal and tongue base levels [218]. Our manual analysis using airflow signals revealed that 33 of the 58 patients (57%) had more than one site-of-collapse. Therefore, methods based on analysing a single snore episode may not successfully identify the predominant site-of-collapse. My current method solves these issues by analysing the complete set of snore signals during hypopnoea events collected during natural sleep and determining the predominant site-of-collapse for a patient.

Previous studies suggested that identifying patients with tongue-based collapse could assist clinicians in choosing better treatment using OA therapy [21–25,97,102–108]. A study showed that the information regarding the posteriorly-located tongue collapse, identified using natural sleep endoscopy can predict the responders of OA therapy with an accuracy of 83% [21]. Marques et al. demonstrated that patients with palatal collapse were less likely to respond to OA therapy with a predictive accuracy of 74% using the information regarding the obstruction sites identified using the combination of airflow shape, age and BMI [22]. Lee et al. identified that the accurate information regarding the tonsil obstruction detected by DISE resulted in the high success rate of surgical treatment for OSA [206]. The classification results of the current study for categorising a patient into tongue and non-tongue based on the predominant sites showed that the model achieved an overall accuracy of 82% and 85% for the Learning Set (model development) and Hidden Set (testing on the best model), respectively. The accurate information regarding the tongue-base collapse for a patient will assist physicians in choosing the most appropriate treatment using OAs and thereby improving the success rate of treatment outcome.

My study has several limitations. Firstly, a gold standard method was not used to identify the site-of-collapse. Instead, I have used a published method to indirectly label the hypopnoea events based on the shape of the airflow signal. The authors report that inspiratory flow shape can identify the pharyngeal site resulted in collapse with high accuracy [182]. More details were discussed in section 2.4.6. Labelling was done twice with the help of a highly experienced sleep physician to validate the reliability. Even though high reliability (Cohen’s Kappa value was 0.82) was achieved for the indirect method to manually identify the site-of-collapse using the airflow signal, it introduces some labelling errors as this method bypassed the direct method (endoscopy or pressure catheter) to identify obstruction sites. To overcome the limitations of manual labelling, an automated system can be developed to classify the airflow signal based on airflow contour shape [259] or by a pattern matching algorithm and then extracting the corresponding audio signal.

The second limitation of the study was that the current method requires hypop-
noea event scoring (from PSG study), which limits the application of the method to PSG studies only. A complete automated screening tool that uses audio data only to predict apnoea events and the site-of-collapse can be introduced to overcome the limitation of hypopnoea labelling. Third, the snores were selected without considering the effect of body position and sleep stages. Sleeping position can affect the severity of snoring in terms of time and intensity, (For e.g. supine position resulted in more snoring and OSA than in lateral position [260], frequency of snoring increased during NREM sleep [261]), and can probably affect the site-of-collapse. Also, the sleep position may affect the signal quality and audio frequencies of the recording. Further studies are necessary for predicting the treatment outcome using the snore data and validate with longitudinal outcome data.
6.5 Summary

In this chapter, a novel approach for identifying the site-of-collapse was presented. A supervised machine learning algorithm was developed for categorising the site-of-collapse using snore signal recorded during hypopnoea events and then predicting the predominant site-of-collapse for a patient. For this, two multi-class classifiers based on GMM and LDA model were developed, and identified the best model using the most relevant features by comparing the results.

Performance comparison demonstrated that the LDA classifier outperformed the GMM classifier. Additionally, the feature selection method using the nested cross-validation process resulted in a higher-performing model compared to the model using all of the features. Our overall result on the Hidden Set achieved an overall accuracy of 67% for classifying snore signal into three classes using an LDA classifier. Performance evaluation based on the predominant site-of-collapse achieved an overall accuracy of 69% for classifying all the sites and an accuracy of 85% for classifying tongue and non-tongue-base collapse. My method achieved a similar result to the existing state-of-the-art research work using a simple classification algorithm and with very few features selected to build the model.

Our results demonstrated that the audio signal recorded during natural sleep can successfully identify the site-of-collapse in the upper airway. The additional information regarding the obstruction site may assist clinicians in deciding the most appropriate treatment for OSA, especially in patients with tongue-base collapse as they are far likely to be to gain a large therapeutic benefit from OAAs. We believe the novelty of our work is that it is an important first step in establishing the feasibility of a practical non-invasive, low-cost diagnosis tool for improving the selection of appropriate therapy for OSA patients.
Chapter 7

Cluster Analysis of OSA Related Snore Data

In Chapter 6, a novel approach for identifying the site-of-collapse using snore signals during hypopnoea events and then predicting the predominant site of collapse for an OSA patient was developed. The current chapter presents an unsupervised machine learning algorithm, which was developed to form clusters using the features extracted from snore data during hypopnoeas. The correlation between the clusters generated and the site-of-upper airway collapse was then analysed. The existence of a strong correlation between clusters and sites-of-collapse will add supporting evidence to the notion that snore signals are useful for site-of-collapse analysis.

The previous chapter presented the development of a supervised learning algorithm, which used a multi-class classifier to classify sites of collapse based on snore data. One of the limitations of the study (previous chapter) was that a gold standard method was not used to identify the site-of-collapse and the ground truth for the labelling employed was based on an indirect method that manually identifies the site of collapse based on airflow signal. This bypassed the clinical limitations of more invasive methods but at the expense of potentially introducing some labelling errors. Motivated by the possibilities of an unsupervised method in resolving the challenges in labelling, my objective in the current study was to introduce an unsupervised machine learning algorithm, which is outlined in this chapter. For this, an unsupervised cluster analysis algorithm was developed to assign data to groups or clusters based on similarity measurements. A $k$-means clustering analysis was performed to show the tendency of the data to form clusters, and the correlation between the site-of-collapse and clusters generated was then investigated. The data for this study was taken from
the audio database generated during the study period, as discussed in Chapter 4.

In section 7.1, the details of the existing method to identify the site-of-obstruction using snore data is discussed. The methods adopted in this study are described in section 7.2, where the details regarding study participants, feature extraction, the unsupervised learning approach adopted, feature selection algorithms and performance evaluation are outlined. In section 7.3, the performance of the \( k \)-means clustering model, which was generated using the relevant features selected, is evaluated. Internal and external cluster validation was performed to evaluate the performance and is thus outlined. Finally, the discussion and comparison of the cluster and classification algorithm is given in section 7.4.

### 7.1 Existing Methods

Acoustic analyses of snoring have been successfully implemented in the diagnosis of OSA, as well as in estimating the severity and variation of OSA during different sleep stages. An area in the literature that is currently largely unexplored is the identification of the site of collapse using snore signals. An important research project adopted a classifier to identify the four VOTE obstruction areas using an acoustic analysis of snoring episodes with the help of DISE [34, 207–210]. However, this method includes major limitations, such as inconsistencies in determining the obstruction site in natural and drug-induced sleep, and also that the method was not designed to analyse multi-level sites of collapse in full-night studies. More details were described in section 2.5.1. In contrast to these studies, a multi-class classification algorithm was developed to classify the snore signal into three sites-of-collapse (lateral, palate, tongue-base) using audio recorded during natural sleep and achieved a competitive result to the previous studies. The ground truth for the site of collapse was identified using an indirect method that bypasses the clinical limitations of invasive methods via manual analysis of the airflow signal, as discussed in 4.5.

Cluster analysis has been implemented in OSA research in the diagnosis of OSA, classifying patients based on severity [262], OSA patient characterisation of clinical phenotypes [123], treatment outcomes using physiological signals [262] and other demographic features. Respiratory signals have also been utilised in cluster-analysis approaches to classify snore and non-snore episodes and to determine the classification of OSA and non-OSA as well as the OSA severity [242, 243]. Even though different cluster analyses have been conducted to determine the correlation between snoring and OSA, the implementation of unsupervised algorithms to identify the relationship
between obstruction sites and snoring is currently limited to only two studies.

A $k$-means clustering model and Hidden Markov models were developed using audio recording in a full-night sleep study to cluster the audio signal based on formant frequency features [264, 265]. An analysis was then performed to identify the relationship between physiological or anatomical mechanisms and snoring. The results of this experiment were not generalisable, as the experiment was conducted with one patient, and the study lacked the ground truth annotation by an expert physician [264]. A 2-means clustering method was deployed to differentiate a palatal and non-palatal snore signal combination of the statistical moment coefficients of skewness and kurtosis extracted from the snoring sounds [265]. The results revealed that there is a potential link between the formant frequency of snore and changes in the physiology of the upper airway. This provided the hint that there is a possible association between first formant frequency and the position of the tongue.

This chapter presents the development of an unsupervised machine learning algorithm, which forms clusters using the features extracted from snore data during hypopnoeas. An outline of the analysis performed to determine the correlation between the clusters generated and the site-of-upper airway collapse is then given. A $k$-means clustering analysis was adopted with the help of a feature selection algorithm to select the most relevant features required for the clustering. A systematic evaluation was used to select the highest performing feature set and number of clusters.

7.2 Methodology

7.2.1 Study Subjects

The details of the study participants were described in section 4.1, but for convenience, I summarise the main facts. Fifty-eight patients who attended a full night sleep study and diagnosed with OSA were considered for this study. For this study, the audio recordings of all patients with 1807 hypopnoea events were used. The labelling information regarding each hypopnoea event and the predominant site-of-collapse for each patient were also selected from the database. This labelling information was used for external cluster validation to evaluate the cluster performance. Details of the data used for this study are given in Table 7.1, and more detailed information regarding the characteristics of the patients and labelling are given in Table 4.3 of Chapter 4.
Table 7.1: Participant characteristics details.

<table>
<thead>
<tr>
<th>Subject characteristics (Predominant site-of-collapse)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>58 (Lateral: 11, Palate: 10, Tongue: 32, Multilevel: 5)</td>
</tr>
<tr>
<td>Hypopnoea events</td>
<td>1807 (Lateral: 524, Palate: 357, Tongue: 926)</td>
</tr>
<tr>
<td>Duration (minutes)</td>
<td>849 (Lateral: 259, Palate: 176, Tongue: 414)</td>
</tr>
</tbody>
</table>

7.2.2 Feature Extraction

Fifty identical low-level descriptive features were derived from each hypopnoea event. The time-domain features: (1) Energy, (2) Entropy, and (3) ZCR. The frequency-domain feature consisted of (1) First three formant frequencies, (2) Thirteen MFCC and its first derivatives, (3) Twelve spectral chroma features, (4) Spectral Entropy,(5) Spectral Flux, (6) Spectral Centroids, (7) Spectral roll-off, and (8) Fundamental Frequency and Harmonic Frequency. The details of the features extracted were discussed in section 3.1. Two unsupervised feature selection algorithms were developed to select the highest performing feature set from the extracted features for the development of the cluster model and compared the performance.

7.2.3 Unsupervised Learning Approach

K-means Clustering

K-means clustering is a simple, fast and an efficient data clustering algorithm that works iteratively to allocate each data point to one of k subgroups (to the nearest cluster) based on the similarity of features provided. More details were discussed in section 3.2.3.

Silhouette analysis

Silhouette analysis can be used to analyse the separation distance between the clusters generated. It can also be used to validate the performance of the clusters by measuring how well a data point fits into its own cluster (cohesion) compared with other clusters (separation) [266]. Cohesion and separation are normally based on distances between in-cluster points. The silhouette value for $i^{th}$ data point is defined as
Figure 7.1: An illustration of the elements involved in the calculation of silhouette value \( s(i) \), where the data point \( i \) corresponds to the cluster A.

\[
\begin{align*}
    s(i) &= \begin{cases} 
        1 - \frac{a(i)}{b(i)}, & \text{if } a(i) < b(i) \\
        0, & \text{if } a(i) = b(i) \\
        \frac{a(i)}{b(i)} - 1, & \text{if } a(i) > b(i)
    \end{cases} \\
    \tag{7.1}
\end{align*}
\]

Or it can be expressed in a single formula as,

\[
    s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))},
    \tag{7.2}
\]

where \( a(i) \) is the intra-cluster distance and \( b(i) \) is the mean nearest-cluster distance. Distance can be measured using distance metrics such as Euclidean distance, squared Euclidean distance, or city block. For our study, squared Euclidean distance was used to measure the distance. To illustrate equation 7.1 for 3 clusters A, B and C, the intra-cluster distance \( a(i) \) in cluster A can be calculated as shown in Figure 7.1

\[
a(i) = \text{average distance from } i \text{ to all other data points in cluster A}.
\]

The inter-cluster distance between the data point \( i \) to the cluster B and cluster C can be computed as

\[
d(i, B) = \text{average distance from } i \text{ to all other data points of cluster B}. \\
d(i, C) = \text{average distance from } i \text{ to all other data points of cluster C}.
\]

After this, mean nearest-cluster distance can be identified \( b(i) \) as
\[ b(i) = \text{minimum } d(i, C) \text{, for all } C \text{ except cluster } A. \]

The silhouette value for a data point ranges from \([-1, 1]\). A value of +1 indicates that a data point is far from the neighbouring clusters and very close to its own cluster. Conversely, a silhouette value of -1 indicates that a point is closer to neighbouring clusters than its own cluster and it may be concluded that the data has been placed in the wrong cluster or misclassified. A value of 0 implies that the particular data point is located at the boundary of the two clusters and, hence, it is ambiguous to which cluster the data belong. Therefore, if most of the data points have a higher value, it represents the best cluster configuration [267]. Additionally, the mean silhouette score of a cluster can be computed by taking the average of the silhouette scores of all the data points to give a global measure of a particular cluster.

The silhouette plot provides a visual way to assess parameters such as the number of clusters based on the comparison of its tightness and separation [266–268]. The silhouette plot for a cluster is a plot of the \( s(i) \), ranked in decreasing order, for all data points \( i \) in that particular cluster. The width of a cluster in the silhouette plot represents the silhouette values of all of the points, and a large width indicates higher silhouette values. The height or thickness of the silhouette plot quantifies the cluster size with a similar thickness across clusters, indicating an equal number of data in the clusters. Figure 7.2 shows an illustration of the silhouette plot and clustered data points for some sample data. The sample data is clustered using different cluster numbers (2, 3 and 4) and a \( k \)-means clustering algorithm. The dotted line represents the mean silhouette score. The middle figure shows clustering using 4-means clustering algorithm, with a mean silhouette score of 0.64. From the silhouette plot in the left-hand panel of the middle figure, it can be seen that the plots have similar thickness, which indicates a similar number of data points in each cluster. The larger width of the black cluster indicates a higher silhouette value for the data points of this cluster. This is also demonstrated by the right-hand panel of Figure 7.2, where the black cluster is further separated from the other three clusters.

**Optimal Cluster Number**

Clustering is the process of grouping data by allocating each data point to one of the \( k \) subgroups (to the nearest cluster) based on the similarity of features provided. The most important step in the cluster analysis is to identify the optimal cluster number \( k \) in a dataset without any prior information, as the clustering performance may be highly correlated with the number of clusters [269,270]. Most of the existing clustering
Figure 7.2: Illustration of silhouette plot for a sample data. The sample data is clustered using different cluster number (2, 3 and 4) using a k-means clustering algorithm. The left-hand panel shows the silhouette plot and the right-hand panel shows the clustered data points. The dotted line represents the mean silhouette score (source: scikit-learn.org).
algorithms require a pre-defined cluster number. However, it is not easy to provide the most optimal cluster number in advance. Researchers have developed various methods to identify the optimal cluster number by exploiting various parameters such as depth difference, clustering gain and the Bayesian Information Criterion. Figure 7.2 shows a good example of a silhouette plot for data points clustered using a different number of clusters. The silhouette plot and clustered data indicate that clustering by three and five are not good for the current data points due to the presence of clusters with low average silhouette scores and variations in the size of the silhouette plot. Clustering by four resulted in a silhouette plot with similar thickness and all of the clusters have above average silhouette scores. This can be verified from the scatter plots presented on the right-hand panel. Hence, the optimal cluster number for this data point could be set as four.

For the current study, an algorithm was implemented to identify the optimal cluster number. The algorithm was based on the cohesion and separation property of the cluster, which is considered major characteristics to validate the performance of a cluster. Cohesion indicates how well data fits into its own cluster and the variation of the data within a cluster. Separation measures how well the data is isolated from the other clusters. Silhouette analysis is a simple method used to measure cohesion and separation for a cluster, and silhouette values were used to determine the optimal number of clusters. For this, the silhouette values for each feature and all features together were determined by varying the number of clusters, ranging from two to six. The optimal cluster number for a particular feature is the number with the maximum silhouette value. Using this method, the optimal cluster number for all of the features were separately identified. Finally, the optimal cluster number of the model was identified using the majority principle—that is, selecting the most common cluster number from individual feature results.

7.2.4 Feature Selection

For unsupervised learning, data variance and separability of the features were evaluated to determine the relevance of the features [255]. Most of the feature section algorithms in an unsupervised learning algorithm utilise the variance of the data (e.g., principal component analysis (PCA)). For this study, two algorithms for selecting the most relevant features for the model were implemented and the results were compared. The first method used was the PCA technique. For the second method, a novel algorithm combining silhouette analysis with the Laplacian score algorithm
was developed.

**Principal component analysis**

The snore sound features extracted from each hypopnoea event comprised 50 elements. However, some of these features could have been irrelevant or redundant. To remove irrelevant or redundant features and to select the more relevant features, PCA technique was adopted. PCA is a dimensionality reduction algorithm that projects the data represented by several inter-correlated quantitative dependent variables to a lower-dimensional space and retains most of the variation in the data [271]. The goal of PCA is to extract the most important information from the data, to represent it as a set of new orthogonal variables called principal components, and to represent the pattern of similarity of the data points by preserving as much ‘variability’ in the data as possible [272]. The PCA algorithm consists of the following steps.

(i) Compute zero mean data by subtracting the mean from each of the dimension.

(ii) Compute the covariance matrix Σ. A covariance matrix is a matrix whose \((i, j)^{th}\) element is the covariance between the \(i^{th}\) and \(j^{th}\) dimension of each trace. The covariance of two variables \(X\) and \(Y\) can be defined as

\[
\text{Cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}\tag{7.3}
\]

The covariance matrix is identified from the covariance as follows

\[
\Sigma = (c_{i,j}, c_{j,i} = \text{Cov}(D_i, D_j)),\tag{7.4}
\]

where \(D_i\) is the \(i^{th}\) dimension.

(iii) Compute the eigenvectors and corresponding eigenvalues of the covariance matrix using the equation

\[
\Sigma^{n \times n} = \bigwedge \ast \bigwedge \ast \bigwedge^{-1},\tag{7.5}
\]

where \(\bigwedge\) is the eigenvalue matrix and \(\bigwedge\) is the eigenvector matrix of the covariance matrix \(\Sigma\). The first principal component represents the eigenvector with the largest eigenvalue, and this component corresponds to the direction with the most variance. The number of principal components derived will be the same as the number of eigenvectors.
(iv) Sort the eigenvectors into decreasing order. Choose $k$ eigenvectors with the largest eigenvalues and form a new matrix with these eigenvectors in the columns.

With the new matrix generated of length $k$, two things can be done. The noise of the original data set can be reduced by down weighting some components while keeping all $n$ dimensions. Alternatively, the original data can be transformed to retain only $k$ dimensions. This can be achieved by selecting the first $k$ principal components with a particular threshold (typically 90% or 95%) of the total variations in the data. The variation is conveniently measured by the eigenvalues which measure the amount of variation in the total sample accounted for by each component.

For the current study, the threshold was selected as 95% and calculated it using as

$$\frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{d} \lambda_j} \geq 0.95, \quad (7.6)$$

where $\lambda_j$ represents the eigenvalues of the covariance matrix of dimension $j$ of data set and $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$. The first $k$ principal components were the eigenvectors corresponding to the highest eigenvalues explaining 95% of the total variation and the rest eigenvectors were neglected.

**Combination of Silhouette analysis and the Laplacian score algorithm**

In this feature selection algorithm, silhouette analysis and the Laplacian score algorithm were combined to build an optimal feature selection algorithm. The algorithm consisted of the following steps

(i) Determine the optimal number of clusters for the model using all of the features separately, as described in the previous section.

(ii) Using optimal cluster number, evaluate the mean silhouette value and silhouette plot for each feature.

(iii) Select the features with uniform cluster thickness on the silhouette plot, as the thickness of silhouette plot quantifies the cluster size, and a similar thickness indicates an equal number of data in the cluster. Mathematically, it is calculated by identifying the number of data points in a cluster. For example, for an optimum cluster number of two, features with the number of data points in a cluster of less than 60% were selected.
(iv) Rank the selected features from the previous step based on importance using Laplacian scores, a method to rank the features for unsupervised learning. The detailed information regarding the Laplacian score is given in the next section.

(v) Evaluate the performance by calculating the mean silhouette value through sequentially adding features based on the rank of the features from the previous step.

(vi) Select the features until the mean silhouette value reaches the first maximum.

**Laplacian Score**

Laplacian Score is a widely used feature ranking algorithm used in unsupervised learning to select the most important features to build the model. Laplacian Score evaluates the importance of features according to their locality preserving power and ranks the features based on the score [273]. An important feature should be the one on which two data points are close together if and only if there is an edge between these two points, and with a low Laplacian score.

Consider a data set \( x_r = [x_{r1}, x_{r2}, ..., x_{rn}]^T \), where \( x_{rn} \) represents the \( n^{th} \) observation of the \( r^{th} \) feature. The Laplace score can be calculated as

\[
L_r = \frac{\sum_{i,j} (x_{ri} - x_{rj})^2 S_{ij}}{\text{var}(x_r)},
\]

(7.7)

where \( \text{var}(x_r) \) is the estimated variance of the \( r^{th} \) feature. Minimising \( L_r \) implies that the algorithm prefers features with large variance reflects the importance of the features. \( S_{ij} \) identify the pairwise distance (\( \text{Dist}_{ij} \)) for all the neighbourhood data points \( i \) and \( j \) and convert the distance matrix into a similarity matrix \( S \) using a kernel transformation and is given as

\[
S_{ij} = \exp\left(\frac{-\left(\text{Dist}_{ij}\right)^2}{\sigma}\right),
\]

(7.8)

where \( \sigma \) represents the scale factor for the kernel.

**7.2.5 Performance Evaluation**

An unsupervised cluster analysis was developed using the features extracted from the snore signal during hypopnoeas to assign data into groups or clusters based on the similarity measurement and to investigate the correlation between the site-of-collapse
and the clusters generated. $K$-means clustering was adopted for this study using a feature selection algorithm to select the most relevant features required for the clustering. Generally, cluster validation can be categorised into two types (1) internal cluster validation and (2) external cluster validation, both of which are based on prior outcome information or whether the labelling is used or not [274].

**Internal Cluster Validation**

Internal cluster validation is based only on the intrinsic information of the data (compared only with the cluster generated) extracted from the structure of clusters generated and their relations with each other, without reference to any external information or labels. This information can be measured using the cluster properties of cohesion (how well a data fits in its own cluster) and separation (how well the data isolated from the other clusters), for which lower and higher values are better, respectively [275]. These are considered the major characteristics needed to validate the internal performance of the cluster. Different types of internal validation indices such as the Calinski-Harabasz index, Silhouette index, Davies-Bouldin index and Dunn index are available, which simultaneously measure cohesion and separation. For the current study, the silhouette index was used to evaluate the performance of internal cluster validation. Detailed information regarding the silhouette analysis is given in section 7.2.

**External Cluster Validation**

External cluster validation is based on evaluating the performance of the model using the reference result, which is considered as the ground truth [276]. This method is straightforward when the similarity between two clusters is well-defined. To investigate the correlation between the site-of-collapse and the clusters generated, I attempted to categorise the cluster into different sites-of-collapse by labelling all of the hypopnoea events in a cluster as one of the site-of-collapse classes. Additionally, labelling was performed for each participant’s data based on the predominant site-of-collapse (if more than 60% of total events came in one cluster, while the clusters were already labelled as one of the sites-of-collapse). The performance was evaluated for the label assignment for two cases by comparing the automatic and manually labelled data: 1) based on labelling of the hypopnoea events, 2) labelling the predominant site-of-collapse for each participant and then calculating the accuracy. Accuracy was calculated by determining the proportion of correct predictions and is given in equa-
tion 6.1. 95% CI is also provided and can be computed using the equation 5.6.

7.3 Results

7.3.1 Optimal Cluster Number

For identifying the optimal number of the clusters, the silhouette value for each feature and all features combined was determined by varying the number of clusters (ranging from two to six). The optimal cluster number for a particular feature is the number with the maximum silhouette value. Using this method, the optimal cluster number for all of the features were separately identified. Finally, the optimal cluster number of the model was identified using the majority principle—that is selecting the most repeated number from each feature. The results showed that the optimal number of a cluster was two for most of the features and in the case when all of the features were combined. Detailed results of the silhouette value for each feature with variation in the number of clusters (ranging from two to six) are shown in Figure 7.3. Table 7.2 shows the number of features with different optimum cluster numbers. The optimum cluster number was two for 36 out of 50 features (72%). Based on the results, the optimum cluster for the model was selected as two.

![Figure 7.3: Illustration of silhouette value for all the features by varying the cluster number using a $k$-means clustering. Each line represents the feature and $\Diamond$ represents the maximum silhouette value corresponding to the optimal cluster number for the feature.](image)
Table 7.2: Number of features with different optimal cluster numbers.

<table>
<thead>
<tr>
<th>Optimum Cluster Number</th>
<th>Number of Features (Total: 50 Features)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>36 (72%)</td>
</tr>
<tr>
<td>3</td>
<td>8 (16%)</td>
</tr>
<tr>
<td>4</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>5</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>6</td>
<td>3 (6%)</td>
</tr>
</tbody>
</table>

7.3.2 Feature Selection

Principal component analysis

PCA was implemented to extract the most important information from the features by removing irrelevant or redundant features. As the first three components consisted of 91% of the variation and the first four components consisted of 96% of the variation, the first four principal components were selected to build the clustering model.

Feature selection using Silhouette analysis and the Laplacian score algorithm

In the first phase of the optimal features process, the optimal cluster number for the model was identified as two. In the next phase, the mean silhouette value and silhouette plot for each feature was evaluated using a 2-means clustering. The "best" features were selected based on the thickness of the silhouette plot (uniformity thickness for two clusters), where a similar thickness indicates an approximately equal number of data in the cluster. As the optimal cluster number was two for the model, the best features comprised of clusters with less than 60% of the total data points. Based on the number of data points in a cluster, the algorithm selected 27 features (shown in Table 7.3), by discarding the features where the number of data points was greater than 60% of the total data points in a cluster.

Figure 7.4 shows an example of good features selected for the final model. Both the 1st coefficient of the MFCC and the spectral chroma feature had a uniform width for the clusters with a mean silhouette value of 0.77 and 0.72 respectively. The 1st coefficient of MFCC feature resulted in two clusters, where 968 out of 1807 (53%) data points were contained in Cluster 1 and 839 (47%) of data points were contained in Cluster 2. The clusters generated using the 1st coefficient of spectral chroma feature consisted of 868 (48%) data points in Cluster 1 and 939 (53%) data points in Cluster...
Table 7.3: Features selected based on the number of data points in a cluster.

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7 MFCC Coefficients</td>
</tr>
<tr>
<td>2</td>
<td>7 First derivative MFCC Coefficient</td>
</tr>
<tr>
<td>3</td>
<td>5 spectral chroma Coefficient</td>
</tr>
<tr>
<td>4</td>
<td>First 3 Formant frequency</td>
</tr>
<tr>
<td>5</td>
<td>Energy</td>
</tr>
<tr>
<td>6</td>
<td>ZCR</td>
</tr>
<tr>
<td>7</td>
<td>Spectral Entropy and Flux</td>
</tr>
<tr>
<td>8</td>
<td>Fundamental Frequency</td>
</tr>
</tbody>
</table>

2. Figure 7.5 shows an example of bad features discarded for the final model. Even though the ZCR feature had a high mean silhouette value of 0.989, the width of the two clusters was not uniform, with 1722 out of 1807 (95%) data points in Cluster 1 and 85 (5%) data points in Cluster 2. For the energy feature, mean silhouette value was 0.73 with 602 (22%) data points in Cluster 1, and 1409 (78%) data points in Cluster 2 resulting in a non-uniform width for the clusters. Furthermore, there was a large number of outliers in Cluster 1, with a negative silhouette value for some data points. This indicates that the data was allocated to the wrong cluster.

In the next phase, the Laplace score was identified for the 27 selected features and the selected features were ranked based on the Laplacian score. Although the ranking was performed based on the score, there existed a small difference between the Laplace score for all features. Therefore, in the final phase of the feature selection algorithm, the performance of the cluster generated (calculated using the mean silhouette value) was evaluated by sequentially adding features based on the rank of the features and then selecting the features until the mean silhouette value reached the first maximum.
Figure 7.4: Example of good features selected for clustering. The dotted line represents the mean silhouette value. Both features, 1\textsuperscript{st} coefficient of MFCC and 1\textsuperscript{st} coefficient of spectral chroma has uniform width for the clusters with a high mean silhouette coefficient of 0.77 and 0.72 respectively.

Figure 7.5: Example of bad features discarded for the final model. The dotted line represents the mean silhouette value. Even though zero crossing rate and energy feature has a high mean silhouette value of 0.99 and 0.73 respectively, the number of data points in two clusters are not uniform for both the features. Also, Energy feature has a few number of outliers with negative silhouette value for some data points.
The results showed that the maximum mean silhouette value was achieved (0.79) when the cluster generated with the first 17 features, and therefore, the 17 features were selected to build the final model. Table 7.4 shows the features that were selected. The selected features consist of seven MFCC coefficients, the six first derivatives of the MFCC coefficients, the three spectral chroma features and the second formant frequency.

Table 7.4: Final features selected for the model.

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1^{st}$, $2^{nd}$, $4^{th}$, $6^{th}$, $9^{th}$ and $12^{th}$ Coefficient of MFCC</td>
</tr>
<tr>
<td>2</td>
<td>$2^{nd}$, $4^{th}$, $5^{th}$, $7^{th}$ and $9^{th}$ Coefficient of First derivative MFCC</td>
</tr>
<tr>
<td>3</td>
<td>$1^{st}$, $4^{th}$, $5^{th}$ and $7^{th}$ Coefficient of spectral chroma</td>
</tr>
<tr>
<td>4</td>
<td>$1^{st}$ and $2^{nd}$ Formant frequency</td>
</tr>
</tbody>
</table>

### 7.3.3 Internal Cluster Validation

Based on the features selected (PCA and Laplacian score algorithms), the final clustering model using $k$-means clustering was developed, and an internal cluster validation was performed to evaluate how well the cluster fit the data without reference to any external information.

**Principal component analysis**

Using PCA, the first four principal components were selected to build the final model, as they contained 96% of the total variations. To cross-verify the optimal cluster number, the silhouette value for the selected principal components was determined by varying the number of clusters (ranging from two to six). The optimal cluster number obtained was two. Finally, the 2-means cluster model was developed using the first four principal components. The performance of the internal cluster validation revealed that the data had a strong tendency to form two clusters with a mean silhouette value of 0.72. Additionally, 1275 (71%) data points had a silhouette value higher than the mean silhouette value (0.72). Figure 7.6 shows the silhouette plot with the first four principal components. Figure 7.7 shows the plot for the number of clusters vs mean silhouette value using the selected features.
Based on the silhouette analysis and the Laplacian score algorithm, the 17 most relevant features (as shown in Table 7.4) were selected to build the final cluster model. To cross-verify the optimal cluster number, the silhouette value for the features selected was determined by varying the number of clusters ranging from two to six using a $k$-means clustering and comparing the mean silhouette values. As shown in Figure 7.7, the optimal cluster number using this method was two. For the final system, a 2-means cluster model was developed using the optimal features. Internal cluster validation showed that the data had a strong tendency to form two clusters, with a mean silhouette value of 0.79. Additionally, 1442 (80%) data points had silhouette values higher than the mean silhouette value (0.79). Figure 7.8 shows the silhouette plot with the selected features.

The performance of the internal cluster validation for the clusters generated using $k$-means algorithm showed that the data had a natural tendency to form two clusters with a high mean silhouette value. $K$-means algorithm using PCA to select the most relevant feature revealed that a mean silhouette value of 0.72, with 71% of the data points having silhouette values greater than the mean silhouette value. $K$-means algorithm using the features selected via silhouette analysis and the Laplacian
score algorithm achieved a mean silhouette value of 0.79, with 80% of the data points having silhouette values greater than the mean silhouette value. When comparing the internal cluster validation performance of two feature selection methods (see Figures 7.6 and 7.8), the 2nd algorithm (silhouette analysis and the Laplacian score algorithm) was found to outperform the PCA method, with a mean silhouette value of 0.79 versus
0.72 for the PCA method.

7.3.4 **External Cluster Validation**

External cluster validation was performed to evaluate the performance of the model using the manual labels. In this phase, the aim was to analyse the correlation between the cluster formed and the site-of-collapse. As the optimal cluster number was identified as two for the model, the clusters were related to tongue and non-tongue related collapse. Non-tongue related collapse data was generated by combining all of the hypopnoea events from the lateral wall and palate related collapse. This process resulted in 881 (49%) hypopnoea events related to non-tongue-base collapse and 926 (51%) tongue-base collapse. For the predominant site-of-collapse, non-tongue related patients comprised patients labelled with lateral wall, palate and multi-level site-of-collapse. This resulted in 26 (45%) non-tongue base labelled patients and 32 (55%) with tongue base labelled patients. The main reason for labelling tongue-base and non-tongue-base was that previous studies have revealed that information regarding tongue-base collapse in patients could assist clinicians in choosing better treatment using OA therapy, with a high success rate. Furthermore, this combination resulted in a balanced dataset. To validate the external cluster validation performance, the clusters were labelled as tongue and non-tongue based collapse and then identifying the predominant site-of-collapse for each patient. Accuracy was calculated by comparing the manual label and the labels assigned to the cluster. The predominant site-of-collapse for a patient was identified using the same rule in the manual labelling of the patient (i.e., if more than 60% of total events comes in one cluster, while clusters were labelled as tongue and non-tongue base collapse).

**Principal component analysis**

To validate the performance of the $k$-means algorithm using the first four principal components, clusters 1 and 2 were labelled as tongue and non-tongue base collapse respectively to calculate the overall accuracy. A representative example of $k$ means clustering on a 2-dimensional subset of features is given in Figure 7.9. The results showed that the model achieved an overall accuracy of 61% (1105/1807) (95% CI, 59%-63%) for categorising tongue/non-tongue related collapse based on hypopnoea events. The model achieved an accuracy of 68% (39/58) (95% CI, 54%-79%) for categorising tongue/non-tongue related collapse based on the predominant site-of-collapse. Detailed results are given in Table 7.5 and 7.6.
Figure 7.9: 2D representation of \(k\)-means clustering. Clustering was done with the first two principal components for the simple visualisation. For the external cluster validation cluster 1 and 2 were labelled as tongue and non-tongue respectively. ○ represents the misclassified event based on the manual labelling.

Table 7.5: Confusion matrix for hypopnoea event based site-of-collapse (Tongue/non-Tongue) classification. Data set comprises of 1807 (Non-tongue: 881, Tongue:921) labelled hypopnoea events. For the cluster labelling, cluster 1 and 2 were labelled as tongue and non-tongue base collapse respectively.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>Automatic Classification</th>
<th>Non-Tongue</th>
<th>Tongue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Tongue</td>
<td>542</td>
<td>339</td>
<td></td>
</tr>
<tr>
<td>Tongue</td>
<td>363</td>
<td>563</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.6: Confusion matrix for the predominant site-of-collapse (Tongue/non-Tongue) classification.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>Automatic Classification</th>
<th>Non-Tongue</th>
<th>Tongue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Tongue</td>
<td>15</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Tongue</td>
<td>8</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Feature selection using Silhouette analysis and the Laplacian score algorithm

A 2-means cluster model was developed using 17 features selected through silhouette analysis and the Laplacian score algorithm. For performance validation, cluster 1 and
2 were labelled as tongue and non-tongue base collapse, respectively, to calculate the overall accuracy. A representative example of k means clustering using the first two relevant features is given in Figure 7.10.

Figure 7.10: 2D representation of k—means clustering. Clustering was done with features identified using Laplacian scores for the simple visualisation. For the external cluster validation, cluster 1 and 2 were labelled as tongue and non-tongue respectively. • represents the misclassified event based on the manual labelling.

Table 7.7: Confusion matrix for hypopnoea event based site-of-collapse (Tongue/non-Tongue) classification. Data set comprises of 1807 (Non-tongue: 881, Tongue:921) labelled hypopnoea events. For the cluster labelling, cluster 1 and 2 were labelled as tongue and non-tongue base collapse respectively.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>Automatic Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Tongue</td>
<td>553</td>
</tr>
<tr>
<td>Tongue</td>
<td>324</td>
</tr>
<tr>
<td></td>
<td>328</td>
</tr>
<tr>
<td></td>
<td>602</td>
</tr>
</tbody>
</table>

This model achieved an overall accuracy of 64% (1155/1807) (95% CI, 62%-66%) for categorising tongue/non-tongue related collapse based on hypopnoea events and an accuracy of 68% (39/58) (95% CI, 54%-79%) for categorising tongue/non-tongue related collapse based on the predominant site-of-collapse. Detailed results are given in Tables 7.7 and 7.8.
Table 7.8: Confusion matrix for the predominant site-of-collapse (Tongue/non-Tongue) classification.

<table>
<thead>
<tr>
<th>Manual Labelling</th>
<th>Automatic Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Tongue</td>
</tr>
<tr>
<td>Non-Tongue</td>
<td>16</td>
</tr>
<tr>
<td>Tongue</td>
<td>9</td>
</tr>
</tbody>
</table>

External cluster validation for the clusters generated using $k$-means algorithm showed that the model achieved satisfactory results. In comparing the performance of the model using two feature selection methods (PCA, silhouette analysis and the Laplacian score algorithm), it was found that both methods showed similar results when identifying the predominant site-of-collapse (Tongue/non-Tongue) with an accuracy of 68%. Hypopnoea event classification (Tongue/non-Tongue) accuracy was slightly higher (64%) for the model using features selected by silhouette analysis and the Laplacian score algorithm compared with the PCA method (61%).

### 7.4 Discussion

For the current study, an unsupervised cluster analysis algorithm to assign data into groups or clusters based on similarity measurements was performed. A $k$-means clustering analysis was developed to show the tendency of the data to form clusters and to investigate the correlation between the site-of-collapse and the clusters generated. The high performing features for the model were identified using two feature selection algorithms. Internal cluster validation showed that the data exhibited a strong tendency to form clusters, with the optimal cluster number of two and a mean silhouette value of 0.79. External cluster validation showed that the model achieved an accuracy of 68% when classifying patients based on the predominant site-of-collapse (tongue/non-tongue).

To resolve the problem of using an invasive method to obtain the ground truth for the site-of-collapse along with the snore data, a simple, non-contact and inexpensive technique to identify the site-of-collapse using snore signals was introduced (in Chapter 6). The probable site-of-airway collapse was indirectly identified using manual analysis of the shape of the airflow signal during hypopnoea. However, while the indirect method bypassed the clinical and cost limitations of the invasive methods, it did introduce some labelling errors. Even though different classification algorithms
have been developed in classifying snore signals into different obstruction sites, unsupervised algorithms to identify the relationship between snoring and the obstruction sites have received no attention in the literature. Therefore, the current results are now compared with the classification method discussed in the previous chapter.

The classification model (leave-one-patient-out nested cross-validation using 58 data) achieved an overall accuracy of 67% when classifying the site-of-collapse (tongue and non-tongue) based on hypopnoea events. This achieved an accuracy of 81% for labelling the predominant site-of-collapse for a patient. The corresponding accuracy achieved using external cluster validation was 64% and 68%, respectively. The performance comparison showed that the classification model outperformed the clustering model. However, there are some advantages to the clustering algorithm. The clustering algorithm does not require labels or ground truth information, whereas the classification methods require ground truth information. Another advantage of cluster analysis is that the computation complexity is less compared with the classification method, as the classification algorithm requires a training set to tune the model.

Feature selection for the clustering model indicated that the best-suited feature set consists of frequency-related features, as shown in Table 7.4. MFCCs tend to play an important role in clustering of snore data based on the site-of-collapse similar to the classification model, which strongly support the hypothesis that the characteristic of snore signal varies depending on different excitation locations as described in the previous chapter. Another important feature selected for the clustering model is formant frequency. Depending on the shape and physical dimensions of the upper airway, different snoring sounds with diverse acoustic properties, or formant frequencies, are produced. Since OSA patients have weaker dilator muscle tone and narrower airway than non-OSA patients, the feature selection indicates that formant frequencies can provide important information on the excitation location. Comparing the optimal features selected for the clustering and classification models indicates that frequency related features provide the most discriminating information regarding the site-of-collapse. As the optimal features selected predominantly comprise MFCC features (five out of seven features in classification and 11 out of 17 features in clustering), MFCCs tend to play an important role in identifying the obstruction site, which has successfully been used in speech and snore analysis.

There are several limitations to this study. The main limitation arose due to the implementation of k-means clustering, as it is highly sensitive to initial values and there is difficulty in predicting the optimal number of clusters. Further studies are necessary to investigate the correlation between the cluster generated using snore and
the site-of-collapse by utilising advanced cluster models (e.g., hierarchical clustering [277], latent class analysis [278] and Gaussian mixture model clustering), combining more physical and clinical participant characteristics, and by implementing more advanced feature selection algorithms. Additionally, further studies are needed to predict treatment outcomes using snore data and to validate these with longitudinal outcome data.

7.5 Summary

This chapter presented the development of an unsupervised machine learning algorithm, which forms clusters using the features extracted from snore data during hypopnea. An analysis of the correlation between the clusters generated and site-of-upper airway collapse was then given. The chapter then outlined the deployment of a k-means clustering algorithm, which forms clusters using the features extracted from snore data. A systematic evaluation was developed using silhouette analysis to identify the optimal cluster number. Two feature selection algorithms using PCA and a novel method combining silhouette analysis with the Laplacian score algorithm were developed to select the optimal features for the model.

Internal cluster validation showed that the data exhibited a strong tendency to form the clusters, and the data was found to fit well in two clusters, with a mean silhouette value of 0.79. External cluster validation showed that the model achieved an accuracy of 68% in classifying the participants based on the predominant site-of-collapse (tongue/non-tongue). The results indicate there exists a correlation between the clusters and the site of collapse, which supports the evidence that the snore signal is useful in determining the site of collapse, as discussed in the previous chapter. Further studies are needed to investigate the correlation between clusters generated and the snore signal, which could include more patient characteristics and predictions regarding treatment outcomes.
Chapter 8

Conclusion and Future Work

8.1 Conclusion

Acoustic analysis of snoring sounds is a simple, non-invasive and inexpensive way to assess the upper airway. Conventional methods in determining the sites-of-collapse in the upper airway include the use of an endoscope to directly visualise the airway collapse or a pressure catheter to infer the site-of-collapse during natural or drug-induced sleep. Unfortunately, these methods have limited clinical application, as they are not well tolerated by patients, are time consuming and expensive, and the diagnosis obtained using drug-induced sleep may have limited relevance in natural sleep. Sometimes non-invasive imaging techniques such as MRI and CT scanning are also used, but these methods are not suitable for routine clinical practice due to their cost and the challenges associated with scanning during sleep.

In this thesis, the acoustic analysis of snoring was proposed as a simple, non-invasive and inexpensive way to identify the obstruction site in OSA patients during natural sleep. Supervised and unsupervised machine learning algorithms were used to identify the site-of-collapse from the snore recordings during hypopnoea events. The results indicate that the acoustic properties of hypopnoea related snoring sounds during natural sleep vary with the site-of-collapse, and therefore acoustic analysis can potentially be used as an auxiliary method to identify the predominant site-of-collapse in OSA patients with high accuracy. Information regarding the site-of-collapse could help clinicians in deciding the most appropriate treatment in OSA patients.

For this study, an audio database was generated using audio recordings obtained from 58 OSA patient’s undergone an overnight sleep test. The audio signals were synchronised with the PSG data using a resampling technique. The spectral subtraction
method was deployed to remove the background noise from the audio recordings to improve the signal-to-noise ratio. The probable site-of-collapse was determined using manual analysis of the shape of the airflow signal during hypopnoea, which has been reported to correlate with the site-of-collapse. The labelling was performed twice with the help of an experienced sleep physician to validate the reliability.

An automatic machine learning algorithm using an LDA classifier was deployed to extract OSA related snore events from the nocturnal audio recordings. The results indicate that the model achieved an accuracy of 87% when identifying snore events from the audio recordings and an accuracy of 72% when identifying OSA related snore events from the snore events. The results showed that the model achieved a decent accuracy in extracting the OSA related snore events.

We investigated whether the audio signal recorded during hypopnoea events could predict the site-of-collapse of the upper airway. Two multi-class classifiers (LDA and GMM) were developed to classify the snore during hypopnoea as lateral wall, palate and tongue-base related collapse and to predict the predominant site-of-collapse for each patient (lateral wall, palate, tongue-base related collapse or multi-level site-of-collapse). Feature selection was employed using an unbiased nested cross-validation to boost the classification performance. Performance comparison demonstrated that feature selection resulted in a higher-performing model than using all features, and hence the feature selection system was chosen for evaluation on the Hidden Set. Performance comparison of two classifiers showed that the LDA classifier outperforms the GMM classier in classifying the site-of-collapse using snore signals. The performance evaluation based on the predominant site-of-collapse achieved an overall accuracy of 69% when classifying all of the sites and an accuracy of 85% when classifying tongue and non-tongue-base collapse. The results indicate that the acoustic properties of hypopnoea related snoring sounds during natural sleep vary in association with the site-of-collapse, therefore can potentially be used as an auxiliary method to identify the predominant site-of-collapse in OSA patients.

Finally, the correlation between the site-of-collapse and the clusters generated using the features extracted from the snore during hypopnoeas was investigated. An unsupervised machine learning algorithm was developed using a \( k \)-means clustering to form clusters. An analysis of the correlation between the clusters generated and the site-of-collapse in the upper airway was then performed. Two feature selection algorithm using PCA, and a novel method combining silhouette analysis with the Laplacian score algorithm were developed. External cluster validation was performed by labelling the cluster as one of the three sites-of-collapse and it was then compared
with manual labelling. A feature selection algorithm combining silhouette analysis and the Laplacian score algorithm showed a stronger tendency to forming clusters compared with using PCA. Internal cluster validation showed that the data exhibited a strong tendency in forming clusters, with the optimal cluster number of two and a mean silhouette value of 0.79. External cluster validation showed that the model achieved an accuracy of 68% when classifying participants based on the predominant site of collapse (tongue/non-tongue). The results indicate that there exists a correlation between clusters and the site of collapse, which adds evidence to the snore signal being a useful signal for the site-of-collapse analysis.

In conclusion, this study proposes the use of a non-invasive method to automatically identify the site-of-upper airway collapse in OSA patients using acoustic signal processing and machine learning methods. A non-contact, inexpensive technology exploiting audio-signal recording during hypopnoea events in natural sleep was used to indirectly identify the site of collapse. The novelty of our work is that it is an important first step in establishing the feasibility of a practical non-invasive, low-cost diagnosis tool for improving the selection of an appropriate therapy for OSA patients.

### 8.2 Future Work Recommendations

The findings of this thesis highlight the application of the acoustic analysis of snoring during hypopnoea events in natural sleep towards the identification of the site-of-collapse. Although the current study has shown encouraging and promising results, the dataset is a modest size due to the time constraints of a PhD study. Therefore, this study requires more in-depth research and has several possibilities for enhancement.

- One important aspect for future studies relating to the current study would be the possibility of predicting treatment outcomes using snore data and validating this using longitudinal outcome data. Even though this study showed that the acoustic properties of snoring can be successfully used to identify the site of collapse, this approach has yet to be applied to predicting treatment response. Understanding the responders of the treatment based on the site-of-collapse phenotype could help to improve patient selection and thereby enhance treatment outcomes.

- Another area of research could be the development of a screening tool that uses only audio data to predict apnoea events and the site-of-collapse using supervised machine learning algorithms. This could be developed by combining
the method used to identify hypopnoea events from the raw audio signal using advanced techniques and then identifying the site-of-collapse. Furthermore, more advanced machine learning techniques such as deep learning, LSTM, CNN can be deployed to develop a more accurate model with more data.

- It would also be worth conducting a study on the acoustic properties of snoring relating to the obstruction site and anatomic factors. As various anatomical factors such as the size of tonsils and tongue base are different in patients with OSA, this can affect the spectral character of the snoring sound by affecting the shape and size of the upper airway.

- More studies are required to exploit a greater number of potential acoustic features (e.g., wavelet features, non-stationary features) and higher-level statistical model-based features(e.g., bi-spectrum). Some fundamental work in exploring the correlation between acoustic features and the anatomical changes in the upper airway could provide a better understanding of the snore generation mechanism. Further studies are necessary in exploring the features extracted from snore events rather than considering the snore during hypopnoes as a single event.

- It would also be interesting to observe the performance of the system using audio recordings with an advanced microphone and placement of the microphone nearer to the bed, as there are no standards available for snore measurement. Furthermore, studies should incorporate the effect of body position and sleep stages during the incidence of snoring.

- Further studies should also investigate the correlation between the cluster generated using snore and the site of collapse by utilising advanced cluster models (e.g., hierarchical clustering, latent class analysis and Gaussian mixture-model clustering) and by combining more physical and clinical patient characteristics. This should also occur by implementing more advanced feature-selection algorithms and by validating the model with longitudinal outcome data.

- Creating a large scale snoring signal database would generate opportunities to develop more robust models. With more data, advanced signal processing and machine learning/deep learning methods could be adopted. Although snore data is easier to collect, labelling is expensive, time-consuming, and may not be sufficiently accurate. This could be resolved by introducing a synthetic data
generation algorithm such as semi-supervised conditional generative adversarial networks. Additionally, the limitation associated with manual labelling in the current study could be overcome by implementing a complete automated system. Specifically, this could be achieved by introducing an automated system to classify the airflow signal using a classifier or by a pattern matching algorithm based on the shape and then extracting the corresponding audio signal.

- It will be worth conducting a study using gold standard method (endoscope data) to identify the site-of-collapse and compare the results with the current method. This can be achieved by conducting a standard overnight PSG and DISE in the same patient to see whether audio recording in normal PSG predicts the site-of-collapse as observed in DISE. This could provide more insight into the effectiveness of the current method.
Bibliography


despite previous tonsillectomy and adenoidectomy and normal controls”, *Pediatr Radiol.*, vol. 36, no. 6, pp. 518-523, 2006.


