An Acoustic Monitoring System for Adherence Measurement and Analysis of Inhaler Technique

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy in the
Faculty of Science and Engineering

2019

Wenying Xie

School of Engineering
Department of Electrical and Electronic Engineering
Contents

Abstract ................................................................................................................................. 9
Declaration .............................................................................................................................. 11
Copyright Statement ............................................................................................................ 12
Acknowledgements .............................................................................................................. 14
Publications .......................................................................................................................... 15
Chapter 1  Introduction ...................................................................................................... 16
  1.1 Setting the scene .......................................................................................................... 16
  1.2 Chapter summary ......................................................................................................... 18
Chapter 2  Literature review ............................................................................................ 21
  2.1 Overview .................................................................................................................... 21
  2.2 Review of inhaler technique ....................................................................................... 22
  2.3 Review of automatic identification of breath sounds ................................................... 26
  2.4 Review of electronic adherence measurement ............................................................ 29
  2.5 Conclusion ................................................................................................................ 34
Chapter 3  Background Theory ....................................................................................... 35
  3.1 Overview .................................................................................................................... 35
  3.2 Foundation of Signal Processing .............................................................................. 35
  3.3 Mel frequency cepstral coefficient (MFCC) ............................................................... 39
  3.4 Statistical Models ....................................................................................................... 45
  3.5 Machine Learning ..................................................................................................... 49
  3.6 Neural Networks ....................................................................................................... 60
  3.7 Cross-validation methods .......................................................................................... 65
  3.8 Conclusion ................................................................................................................ 65
Chapter 4  Hardware System Design .............................................................................. 67
  4.1 Overview .................................................................................................................... 67
  4.2 Technical background ............................................................................................... 68
  4.3 Evaluation tools ......................................................................................................... 75
  4.4 Prototype – first version ............................................................................................. 78
  4.5 Prototype – Second version ....................................................................................... 89
List of Figures

Figure 2.1 Dry Powder Inhaler (DPI) – Turbuhaler .............................................. 23
Figure 2.2 Dry Powder Inhaler (DPI) – Diskus .................................................. 23
Figure 2.3 Pressurised Metered Dose Inhaler (pMDI) ............................................. 23
Figure 2.4 the procedure of acoustic identification ................................................ 27
Figure 2.5 Verihaler provides wireless connectivity (Source: adapted from [65]) ...... 31
Figure 2.6 Hailie™ sensor (Source: adapted from [66]) ......................................... 31
Figure 2.7 Propeller sensor shown with the mobile application (Source: adapted from [67]) .......................................................................................................... 32
Figure 2.8 AIM was developed for effective training of inhaler use (Source: adapted from [68]) .......................................................................................................... 32
Figure 2.9 T-Haler training device (Source: adapted from [70]) .............................. 33
Figure 3.1 Original signal 1 Hz, sampling frequency 10 Hz ..................................... 36
Figure 3.2 Original signal 10 Hz, sampling frequency 10 Hz .................................. 36
Figure 3.3 Output signal y[n] was convolved by x[n] and h[n] ................................. 37
Figure 3.4 The typical FIR filter of order N block diagram ..................................... 39
Figure 3.5 MFCC extraction steps ......................................................................... 40
Figure 3.6 the Ratio of Actual Frequencies to Mel Scale ....................................... 42
Figure 3.7 Spectrum of the mel filter bank (normalised) ....................................... 43
Figure 3.8 Gaussian probability distribution .......................................................... 45
Figure 3.9 State space of Markov chain of latent variables .................................... 47
Figure 3.10 Transition diagram of latent variables .................................................. 48
Figure 3.11 The transition probabilities matrix ....................................................... 49
Figure 3.12 Maximum boundary in binary classification ....................................... 50
Figure 3.13 This dataset is a non-linear dataset and consist of two features (X, Y) ..... 55
Figure 3.14 The additional feature $Z = (X - Y)^2$ makes the non-linear dataset separable .............................................................................................................. 56
Figure 3.15 Binary tree classification (100 samples) .............................................. 57
Figure 3.16 Top- partitioning views for four classes; Bottom – binary decision tree for four classes ....................................................................................................... 58
Figure 3.17 Simple convolutional neural network architecture ................................ 61
Figure 3.18 2D convolution with filters $w_0, w_1$ .................................................. 61
Figure 3.19 Max pooling operation ......................................................................... 62
Figure 3.20 The fully connected layer receives encoded features from the previous layer and outputs the result to the final layer [112] ............................................ 62
Figure 3.21 A recurrent neural network with forwarding computation. .......................... 64
Figure 4.1 Typical FAT file system partition ................................................................. 68
Figure 4.2 Typical operation flow with FatFs Module .................................................. 69
Figure 4.3 4-Wire SPI with multiple slaves ................................................................. 70
Figure 4.4 Timing of SPI clock polarity and phase ...................................................... 71
Figure 4.5 Pinout of Micro SD Card ............................................................................. 71
Figure 4.6 R1 & R3 Response Format ........................................................................... 73
Figure 4.7 SDIO (SPI mode) Initialization Procedure .................................................. 74
Figure 4.8 The design of the touchpad ....................................................................... 75
Figure 4.9 32L152CDISCOVERY Evaluation Board .................................................. 76
Figure 4.10 EVAL-ADXL362Z .................................................................................... 76
Figure 4.11 SD Card Module ....................................................................................... 77
Figure 4.12 Digital MEMS microphone module (MP34DB02) ...................................... 77
Figure 4.13 PDM to PCM Module ............................................................................... 78
Figure 4.14 The first version of the hardware design overview ...................................... 79
Figure 4.15 The schematic of MCU for the first version .............................................. 80
Figure 4.16 MCU PCB Layout for first version ............................................................. 80
Figure 4.17 Schematic of the power management unit .................................................. 81
Figure 4.18 Schematic of the accelerometer circuit ...................................................... 81
Figure 4.19 Schematic of capacitive sensor .................................................................. 82
Figure 4.20 Capacitive sensor PCB layout ................................................................. 82
Figure 4.21 Pre-amplifier circuit of the analogue microphone ...................................... 83
Figure 4.22 PCB Layout of microphone pre-amplifier ................................................ 84
Figure 4.23 Connection of micro-SD with SPI mode .................................................... 84
Figure 4.24 the Schematic of Connectors .................................................................... 85
Figure 4.25 Top layer of PCB ...................................................................................... 86
Figure 4.26 Bottom layer of PCB ............................................................................... 86
Figure 4.27 Firmware structure .................................................................................. 87
Figure 4.28 The device is activated by motion detection; it then starts recording if the user touches the specific area of the capacitive sensor .......................................... 88
Figure 4.29 Schematic of MCU for the second version ................................................ 90
Figure 4.30 Schematic of digital MEMS microphone ................................................... 90
Figure 4.31 Top layout (red) and bottom layout (blue) .................................................. 91
Figure 4.32 Second version ......................................................................................... 91
Figure 4.33 SD card was lift and more space can be used (yellow area) ......................... 92
Figure 4.34 Schematics of MCU for the final version .................................................. 93
Figure 4.35 Schematics of the memory unit for the final version .................................... 94
Figure 4.36 PCB of the final version; left is top side; right is the bottom side.  
Figure 4.37 The final version has a smaller size and cost-effective design. MCU, SPI, Flash and digital microphone are placed on (a); SD card and accelerometer are on another side (b); (c) the SD card is above the other components.  
Figure 4.38 The final version is activated by pressing the button; the device working status indicated by a flashing LED.  
Figure 5.1 (a) and (c) are waveform of exhalation and inhalation sounds; right two figures are waveform envelopes of exhalation (b) and inhalation (d).  
Figure 5.2 Exhalation spectrum (a) and inhalation spectrum (b).  
Figure 5.3 The magnitude of the spectrogram (a) decreases over time, but the spectrogram (b) changes more slowly.  
Figure 5.4 The temporary and spectral characteristic of the actuation sound.  
Figure 5.5 Audio augmentation methods.  
Figure 5.6 Latent states with observed sequence.  
Figure 5.7. Top is the architecture of CLDNNs; middle is a repeating module in LSTM; bottom is the typical structure of a LSTM cell.  
Figure 5.8 (a) is one of the architectures of ResNet34; (b) is modified ResNet for 1-dimentional dataset.  
Figure 5.9 C++ software for motion signal characterisation in real-time.  
Figure 5.10 The ADXL362 has three sensing axes $A_x$, $A_y$ and $A_z$ (left); the axes of acceleration sensitivity are flipped on the device (right).  
Figure 5.11 The output response while the device was shaken along the X-axis.  
Figure 5.12 The output response while the device was shaken along the Y axis.  
Figure 5.13 The output response while the device was shaken along the Z-axis.  
Figure 5.14 The output response while inhaler was shaken as normal.  
Figure 5.15 The upper waveforms are acceleration response of X-axis, Y-axis and Z-axis; the bottom step wave is grms output ($N = 32$).  
Figure 5.16 The upper waveforms are acceleration response of X-axis, Y-axis and Z-axis; the bottom step wave is grms output ($N = 200$).  
Figure 6.1 Pre-processing the accelerometer outputs.  
Figure 6.2 The accelerometer samples with the different activities.  
Figure 6.3 The left plots show accelerometer signals with shake motions and the right are the outputs of the shake detection algorithm.  
Figure 6.4 The acceleration signals (left) were averaged by the moving average method with a window of 195 values; The yaw and roll angles were used to describe the holding orientation (right).  
Figure 6.5 The sign of the yaw angle is reversed if the rotation of the yaw extends over $\pm 180^\circ$. 
Figure 6.6 The device is considered as vertical if its yaw and roll angles are within $\pm 35^\circ$ (red area) ................................................................. 130

Figure 6.7 The black arrows show the model training process and the estimation of models is shown in red ................................................................. 132

Figure 6.8 Overlapping a clean breath sound and a noisy sound at SNR = 5dB, 15dB, 25dB and 35dB ........................................................................... 133

Figure 6.9 The acoustic record collected from a participant .......................... 137

Figure 6.10 (a) shows the participants’ self-assessments and (b) shows the electronic assessments ................................................................. 140

Figure 6.11 (a) Length of inhalation time; (b) actuation timing; (c) length of breath-holding time; (d) percentage of exhalation before inhalation; (e) percentage of cough in holding breath step; (f) percentage of successful actuation ........................................................................ 141

Figure 6.12 The summary of average correct proportion of the recommended steps .................................................................................. 142
List of tables

Table 2-1 Comparison between pMDIs and DPIs .......................................................... 24
Table 2-2 Overview of recommended steps in pMDI, Turbuhaler and Diskus use .......... 25
Table 2-3 Comparison of the cost of the medicinal form ............................................. 26
Table 2-4 Summary of smart inhalers ............................................................................ 34
Table 3-1 Typical MFCC features .................................................................................. 44
Table 3-2 Random forest pseudocode .......................................................................... 59
Table 4-1 SPI Mode Command Set .............................................................................. 72
Table 6-1 Experiment 1: inhaler triggered record form ................................................. 122
Table 6-2 Summary of volunteer recordings ................................................................. 123
Table 6-3 The test result of the shake detection algorithm ........................................... 126
Table 6-4 The test result of the tilt algorithm ............................................................... 130
Table 6-5 Comparison of breath recognition models .................................................... 134
Table 6-6 Performance of the classifiers – actuation identification ......................... 135
Table 6-7 Daily log template ......................................................................................... 136
Table 6-8 Electronic records assessment standard [37][154][155] ................................ 138
Table 6-9 The summary of participants’ daily logs ....................................................... 139
Abstract

Chronic respiratory diseases, such as asthma and chronic obstructive pulmonary disease (COPD) are significant and growing causes of morbidity and mortality worldwide. Poor disease control causes a substantial burden on patients, their families and society. Inhaler therapy is the most popular asthma and COPD treatment, but patients are compromised from inappropriate medication use. How to measure adherence to medication is a huge global problem.

This thesis presents an original audio-based monitoring system for the pressurised metered-dose inhaler (pMDI), which has been developed to track and analyse the patient’s dose delivery and inhaler technique. Prior to this study, there existed no automated system, based on acoustic monitoring, to track and assess pMDI inhaler technique. In terms of the hardware, this system combines accelerometer and acoustic sensors to enable a comprehensive assessment of the pMDI technique. For recognition of the breath phases, this research initially employed a hidden Markov model (HMM) with a Gaussian mixture model (GMM) to identify the phases of the breath sound. Ultimately a model was developed that concatenating two deep learning models (1-D ResNet18 and CLDNNs), which improved the recognition accuracy of acoustic signals degraded by noise. Respecting motion events, this thesis introduces a method based on a root mean square (RMS) framing window in combination with a rotation matrix to achieve robust shake detection and vertical holding detection, respectively. This thesis further describes a series of experiments. Experiment 1 to 3 was designed to characterise the strengths and weaknesses of the system, in which testing also provided useful information for further system development. The final experiment was a real-patient study that involved the recruitment of six patients diagnosed asthma or COPD. The system tracked their pMDI use for three weeks and established the deviations of pMDI usage and technique that existed between the patients’ self-reports and independent electronic reports. Characterisation of these differences, and the ability of the system to not only to monitor adherence but also to identify usage failure points suggest that the
monitoring system will, in the future, be a highly desirable component of digital healthcare provision.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DoculInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
To my father Xie Mingqing and my mother Chen Jihong

To my wife Mou Yang, her mother Mou Yan and her father Fang Yongfei

for your love, encouragement and patience
Acknowledgements

I would like to thank my supervisor Prof. Patrick Gaydecki and Prof. Ann-Louise Caress who provided me with valuable instructions, guidance and encouragement. I would like to thank Dr. Jean Hennings, who organised stakeholder meetings and provided the support for my experiment.

I would like to thank my parents, who encourage me to explore the truth of the world and always teach me to stay on integrity and kindness, my wife Yang Mou for her patience and support. Without my family’s support, it is impossible to complete my PhD.

I would also like to say many thanks for my dear colleagues (Antony Barton, Cong Sun, Erdem Atbas, Jiaxuan Wang, Lorenza Curetti, Lubos Marcinek, Mohammed Abdulaal, Sheheera Ismail, Thomas Lloyd, Zhenzhou Yang and many others) for their comments, ideas and friendships.

Last, I would like to express my gratitude to all whose kindness and advice have made this work successfully.
Publications


Chapter 1
Introduction

1.1 Setting the scene

1.1.1 Background

A complex array of factors, including tobacco smoking, outdoor and indoor air pollution, has led to a massive global increase in morbidity and mortality from respiratory diseases, such as asthma and chronic obstructive pulmonary disease (COPD) [1]. In 2015, the world health organisation (WHO) estimated 3.17 million and 0.38 million deaths caused by COPD and asthma, respectively. In 2017, chronic respiratory diseases were ranked the third cause of global death; the global burden of disease (GBD) dataset (1990-2017) reported that COPD and asthma accounted for 6.6% of deaths worldwide [2]. In terms of the UK, the NHS reported premature mortality in the UK is twice the EU for COPD and more than 1.5 times for asthma in 2012. At the same time, UK asthma patients accounted for 5.9% of its total population, nearly the highest in the whole world [3]. Furthermore, in the same year, the North West of the UK had approximately three million patients of both asthma and COPD, a prevalence that is above the national average [4]. The individual costs of asthma and COPD for patients and their families are significant, especially if the conditions cannot be well controlled [1][3][4]. Breathlessness is a distressing symptom for patients and families. The physical symptoms of asthma and COPD cause disturbance of patients’ daily life. Patients find difficulty in participating in sports or other social activities. In the UK, COPD causes around 23,000 deaths per year, i.e. one person every 20 minutes on average, while three people (adults and children) die of asthma every day. However, 90% of asthma deaths would be preventable if these patients had better management of diseases [3][5]. Inhaled therapy is the major treatment for COPD and asthma, but its therapeutic efficacy is related to long-term adherence to inhaler use and technique [6][7]. Poor adherence not only increases the risk of hospitalisation and disease
exacerbation but also causes patients and their families to suffer from economic pressure and poor quality of life [7][8]. Moreover, it also results in significant economic burdens to societies. In 2012, The UK paid £964.9 million for healthcare costs associated with asthma [9]. According to the European Lung White Book report, in 2011, the annual economic burden to EU countries was €82.3 billion for COPD and asthma combined, in terms of healthcare and lost production costs [10].

There can be a serious consequence of the poor inhaler use; hence, this thesis aims to explore an intelligent method to monitor and record digitally adherence to medication. The system offers the potential to make a significant improvement in patients’ life quality and reduce social and economic burdens on the healthcare system. Prior studies of smart inhalers focused on usage monitoring of the dry powder inhalers (DPIs), in particular, the Diskus inhaler [11]–[13]. However, there are only a few studies on pressurised metered-dose inhalers (pMDIs), which are the most commonly prescribed inhaler devices [14][15]. Additionally, patients commonly prefer to choose an inhaler with the simplicity of use [28]. Therefore, this thesis focuses on the easy use device of digital adherence monitoring of pMDI. Two aspects are considered: the engineering of the acoustic monitoring system and deployment of the system for the assessment for patients. The development work is described in the following chapters and the technique assessment is presented in the exploratory clinical study later in this thesis.

1.1.2 Problem definition

Inhaled and oral medicine are typical treatments for COPD and asthma. Tablets are usually prescribed to patients whose symptoms are not controlled by inhaler alone. [16][17]. To prevent disease exacerbation and relieve mild dyspnoea, the typical treatment is inhaled therapy [12][15]. This method uses an inhaler to deliver the drugs to the site of the pathology directly and efficaciously [18][19]. In terms of inhaled medicine, asthma inhalers can be grouped into reliever and preventer. The reliever inhalers are used to relieve asthma attack when they occur. If symptoms occur more often, the preventer will be prescribed for patients and be asked to use it regularly to
stop exacerbation [16]. The COPD inhalers have three types: short-acting bronchodilator, long-acting bronchodilator and steroid. The short-acting bronchodilator inhaler is only used to widen airways (maximum of four times a day) when patients feel breathless. This type of inhaler usually The long-acting bronchodilator inhalers are similar to the short-acting inhalers, but each dose lasts for longer time (up to 12 hours), which is usually prescribed if patients’ symptoms lost control by short-acting bronchodilator inhalers [17]. In the inhaler market, pMDIs and DPIs are the most popular inhaler types [20][21]. The pMDI type is one of the most common inhalers in the UK and globally (especially in resource-poor countries) owing to its low cost. However, use of the pMDI needs good coordination of exhalation and inhalation to ensure the efficiency of drug delivery [15]. A study reported an average of 85.5% of patients with asthma and COPD could not use their pMDI appropriately [22]. Patients repeatedly failed to use the pMDI correctly, even after instruction [23]. In an attempt to measure adherence to the pMDI technique, inefficient and subjective approaches, such as patients’ self-reports, interviews and telephone interventions, are employed to help professionals or healthcare nurses understand their patients’ situation [6][24][25]. Further, these common methods cannot provide details of inhaler use.

As already discussed, there are a number of challenges associated with adherence measurement. The research described here explores the development of instrumentation and software for and effectively and objectively determining adherence to medication in COPD and asthma patients. In addition, the discrepancy between pMDI adherence as given by self-reports and the electronic reports generated by this new method is also explored in this thesis.

1.2 Chapter summary

Chapter 2 covers the details of the inhaler device and its relevant intelligent features. This chapter commences with a review of the major types of inhaler and how they should be used. A subsequent review of the cost-effectiveness of inhalers explains
why this research chose the pMDI as the primary research object. Next, relevant publications relating to smart inhalers are discussed, which includes smart electronics, automatic identification methods and study results. Furthermore, competitor smart inhalers are also reviewed at the end of this chapter.

Chapter 3 considers supporting theory and principles, such as relevant signal processing foundations, acoustic feature extraction and machine learning methods; these are applied later as algorithms for analysing the information provided by the hardware developed. In particular, the theory and mathematics of signal processing are introduced as a means of characterising the acoustic signals. Then an acoustic feature extraction technique is described in detail. The extraction output is then used as the input for the recognition models. Subsequently, statistic models and machine learning classifiers are explored for the identification of acoustic events. The theory of deep learning neural networks is also explored for further breath recognition improvement.

Chapter 4 focuses on the development of the low power-consumption acoustic monitoring device hardware. Before describing the engineering work, the requirements of the tracking device are discussed in the first section. This is followed by an explanation of the relevant technical details of the prototypes. The hardware development comprises four stages: device evaluation, the first prototype, the optimised prototype and the final prototype. The evaluation stage made use of existing electronic modules to check the feasibility of the initial hardware design. The remaining sections describe the iterative development of prototype devices, which also represent good reference sources for further development of more complex devices.

Chapter 5 considers the characterisation and identification of the acoustic signal and the motion signal signals. The chapter starts with a description of the analysis of the acoustic signals and discusses the methods of acoustic recognition used. Software-based on Qt is then developed for characterising motion signals. In the final section,
an RMS window method and Euler angle method are covered, whose purpose is to
detect valid motion for inhaler use.

Chapter 6 presents a series of experiments which were used to evaluate the
performance of the inhaler hardware and software. The first three experiments
focused on assessing the whole system, the results of which would be employed to
further improve the feasibility and reliability of the system. The final experiment
involved patient volunteers (under ethical approval granted by North West - Greater
Manchester East Research Ethics Committee). This experiment explored the deviation
between self- and electronic reporting.

Chapter 7 reflects on the strengths and weakness of the developed system with
respect to both hardware and software. In addition, the study results, and the
limitations of these experiments are considered in details.

Chapter 8 summaries the work completed in the thesis and its contributions and
findings. It recommends future clinical studies to further explore the benefits of
intelligent devices to assist with adherence to medication. This chapter contains
proposals for system improvement and offers guidelines for the development of a
mature product. In addition to these, this chapter proposes further research that
should focus on how the system may correct inhaler technique and establish a
network community, and explains how the work will contribute to this field of
research.
Chapter 2

Literature review

2.1 Overview

Asthma and COPD both are chronic respiratory diseases and characterised by airflow limitation. In the UK alone, asthma causes around 1000 preventable deaths a year [1][3][5]. Problems with inhaled medication use can be of two kinds: (a) not taking the right number of doses and (b) ineffective use of medication [12][15]. Adherence is defined as “a patient’s behaviour of taking inhaled medications corresponds to a doctor’s instruction” [26]. It implies whether patients comply with the health providers’ recommendation. In contrast, low adherence reflects the poor use of inhaled medications for respiratory patients, which is also a significant factor for prescribed treatment [27]. According to an educational program for asthma control conducted by de Oliveira et al., good instruction in using inhalers could help reduce asthma emergency visits [28]. Correct inhaler technique increases the efficacy of the medication. Ensuring patients can follow instructions for using their inhaler device is a particular challenge for asthma and COPD treatment [29].

Currently, the ability to assess how regularly and effectively patients use their inhaled medications on an everyday basis is very limited [12]. For assessment of inhaler technique, a health provider can only take a face-to-face approach, which is, in practice, very inefficient. Some inhalers have a dose counter for patients’ self-management, but none can provide an indication of whether doses were inhaled correctly, or at all [11]. There is no device which allows accurate and objective assessment of how patients use the full range of inhalers on a day-to-day basis during routine care [11][12].

Deposition of medication in the lungs is influenced by improper inhalation technique [30]. One study concluded that only 5% of patients are able to use pMDI correctly [31].
Another study also showed 86% misuse in MDI and 71% in with the Diskus [22][24]. However, the standard pMDI is inexpensive, hence patients may be given a pMDI in the first instance. This may lead to unintentional under-use of medication, arising from poor inhaler technique. If inefficacious inhaled treatment cannot be identified in patients, it may result in poor asthma control and, in turn, increase unscheduled healthcare use [27]. Only a limited number of inhaler devices have in-built dose counters and other features which can mitigate poor inhaler technique – and usually such devices are more expensive and therefore not the inhalers that would routinely be prescribed in practice. There is consequently considerable scope for improvement in assessing both whether the patients are taking the right number of doses and also whether inhaler devices are being used effectively. This is clearly important, given the rate of non-adherence in asthma and COPD and its known adverse, potentially life-threatening, consequences. Currently, acoustic monitoring of inhalers can measure and analyse adherence by recorded breath sounds [32]. These systems usually have two main parts: a. recording hardware, b. automatic identification of signals.

2.2 Review of inhaler technique

2.2.1 Major types of inhaler

There are many inhaler devices used to prevent or relieve patients’ symptoms; these are divided into two major categories: the dry powder inhaler (DPI) and the pressurised metered-dose inhaler (pMDI). DPIs contain medical powder inside that needs deep and fast breathing to ensure the success of drug delivery. pMDIs use aerosol medication mixed with the propellant, which requires a long and slow inhalation to move the drug into the lung. Figure 2.1 and Figure 2.2 show DPIs and a pMDI is shown in Figure 2.3. The pMDI is used globally, as it is a cheap and widely available device type. A report showed that pMDI, in 2017, had around 70% market share in the global inhaler market and will dominate the market for the next decade [33]. However, pMDI is also the most difficult to use, requiring very good co-ordination of device activation and inhalation [34]. The advantages and limitations of pMDIs and DPIs are shown in Table 2-1.
Figure 2.1 Dry Powder Inhaler (DPI) – Turbuhaler

Figure 2.2 Dry Powder Inhaler (DPI) – Diskus

Figure 2.3 Pressurised Metered Dose Inhaler (pMDI)
Table 2-1 Comparasion between pMDIs and DPIs

<table>
<thead>
<tr>
<th>Inhaler type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
</table>
| pMDIs        | • Portable and compact  
              • Cheap  
              • Use the spacer to help drug delivery | • Require good coordination of actuation and inhalation  
              • Chlorofluorocarbons (CFCs) propellant causes the cold-Freon effect  
              • Emit the greenhouse gas |
| DPIs         | • Portable and compact  
              • No propellant  
              • Actuated by inhalation  
              • Include dose counter | • Require a high respiratory rate  
              • Cost more than pMDIs  
              • Store in a dry and cool place |

2.2.2 Comparison of inhalers using steps

The inhalers require different techniques for use, but they all have three common parts, which includes exhalation, drug inhalation and breath-holding. Table 2-2 lists the steps, which are adapted from the websites of Asthma Canada and Asthma UK [35][36], for pMDI (aerosol), Turbuhaler (dry power) and Diskus (dry power) inhalers. A perfect inhaler technique requires a patient to strictly comply with the recommended six steps. However, previously discussed, it is very difficult for most patients to do this. According to Table 2-2, dry powder inhalers (Turbuhaler and Diskus) have similar steps, which can be briefly summarised as cap off, exhalation, fast deep inhalation and hold breath. However, the pMDI requires different operations to maximise the efficiency of drug delivery. Unlike the Turbuhaler and Diskus, the pMDI inhaler must be shaken well before inhaling the drugs. Then good coordination of actuation and inhalation is compulsory but also the most difficult step for pMDI users [37].
Table 2-2 Overview of recommended steps in pMDI, Turbuhaler and Diskus use

<table>
<thead>
<tr>
<th>Step</th>
<th>pMDI</th>
<th>Turbuhaler</th>
<th>Diskus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a) Shake the inhaler well before use (3 or 4 shakes). b) Unscrew the cap and take it off. Hold the inhaler upright. c) Open the DISKUS: Hold it in the palm of your hand, put the thumb of your other hand on the thumb grip and push the thumb grip until it &quot;clicks&quot; into place. d) Remove the cap and hold the inhaler upright.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>b) Twist the coloured grip of your Turbuhaler as far as it will go. Then twist it all the way back. You have done it right when you hear a &quot;clicks&quot;. c) Slide the lever away from you as far as it will go to get your medication ready.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>a) Breathe out away from the device. b) Breathe out away from the device. c) Breathe out away from the device</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>b) Put the mouthpiece between your teeth, and close your lips around it. c) Place the mouthpiece gently in your mouth and close your lips around it.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>a) Start to breathe in slowly. Press the top of your inhaler once and keep breathing in slowly until you have taken a full breath. b) Breathe in forcefully and deeply through your mouth. c) Breathe in deeply until you have taken a full breath.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>a) Remove the inhaler from your mouth, and hold your breath for about 10 seconds, then breathe out. b) Remove the Turbuhaler from your mouth, and hold your breath for about ten seconds, then breathe out. c) Remove the DISKUS from your mouth, and hold your breath for about ten seconds, then breathe out.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[^a\] - pMDI; \[^b\] - Turbuhaler; \[^c\] - Diskus

2.2.3 Review of the inhaler cost-effectiveness

Poorly controlled COPD and asthma results in heavy economic burdens, both for societies and individuals. In terms of healthcare, the cost-effectiveness of different inhalers is a significant research topic; of particular significance is the balance between financial cost and treatment efficacy [38]. Table 2-3, from the British National Formulary (BNF) [39], shows a cost comparison of Fluticasone and Beclometasone - ingredients both commonly used for prophylaxis of asthma. The table shows that the costs of the powder drug are greater than those of aerosol medicine. Other research revealed that pMDI inhalers have poorer asthma control and a higher rate of disease exacerbation, compared with DPIs [40]. Although DPIs are generally superior to pMDIs, pMDIs are still the most commonly prescribed therapy [41]. Between 2002 and 2008, pMDIs accounted for 47.5% of all inhaler sales in European countries, whilst DPIs accounted for 39.5% [42].
Table 2-3 Comparison of the cost of the medicinal form

<table>
<thead>
<tr>
<th>Active ingredient</th>
<th>Inhalation powder</th>
<th>Pressurised inhalation</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>NHS indicative price / 30 days¹ (Manufacture)</td>
<td>NHS indicative price / 30 days¹ (Manufacture)</td>
</tr>
<tr>
<td>Beclometasone dipropionate 200 mg/dose</td>
<td>£9 (Orion Pharma (UK) Ltd)</td>
<td>£8.52 (GLENMARK²)</td>
</tr>
<tr>
<td>Fluticasone propionate 50 mg/dose</td>
<td>£8.04 (GSK³)</td>
<td>£6.48 (GSK³)</td>
</tr>
<tr>
<td>Fluticasone propionate 250 mg/dose</td>
<td>£51.02 (GSK³)</td>
<td>£36.14 (GSK³)</td>
</tr>
</tbody>
</table>

¹Based on 4 doses/day; ²Glenmark Pharmaceuticals Europe Ltd; ³GlaxoSmithKline UK Ltd;

2.2.4 Conclusion

As a popular treatment for respiratory conditions, inhaled medications face a particular challenge of achieving appropriate inhaler technique [12][15]. Some inhalers are easier to use than others, but the pMDI is the most commonly prescribed in the UK because of its low cost. However, it is also the most difficult to use well, as it requires good coordination of actuation and inhalation, as well as a reasonable degree of manual dexterity [15].

2.3 Review of automatic identification of breath sounds

2.3.1 Previous work

In work reported by Alshaer et al. [43], a microphone was attached to a mask that fits over mouth and nose. The inspiratory Fourier spectrum has two major peaks, the first is located between the 30 Hz - 270 Hz band; the next centre frequency is in the region of 1400 Hz. The expiratory spectrum has only one sharp peak between 30 Hz and 180 Hz, and then its magnitude drops to very low amplitude above 500 Hz. Thus, a frequency Bands Ratio (BR) can be used to distinguish the two phases. It is given by
The experiment analysis concluded that the BR of the mean inspiration was 2.27 and the BR of the mean expiration was only 0.15. This algorithm achieved an accuracy of 97.4% for breath phase detection [43]. However, the breath phase cannot be simply identified by spectrum features, especially in complex environments, because breath sounds are non-stationary signal whose amplitude varies over time. Therefore, advanced feature extraction methods need to be considered in actual applications.

There is a range of applications for the automatic identification of breath sounds (AIBS). Sleep studies use AIBS to detect breathing-related sleep disorders, such as snoring and sleep apnoea. A study reported that the correct detection rate reached approximately 92% when patients were sleeping with the window open; a detection rate of 98% was achieved in a quiet sleeping environment [44]. Research has also been conducted to remove breath sounds from speech and musical signals for improvement of song quality; this work exploited a template based on the MFCC of breathing samples, achieving an accuracy of 92.7% [45]. In addition to the use of MFCC features, the power spectral density (PSD) has also been used to detect breath sounds, with a reported accuracy of 92.8% in a quiet environment [32]. The processing flow of most automatic identification methods is summarised in Figure 2.4.

![Figure 2.4 the procedure of acoustic identification](image-url)
2.3.2 Classification models

Many classifiers including Naïve Bayes (NB), decision tree (DT), random forest, support vector machine (SVM) and the artificial neural network (ANN), have been used for the classification of exhalation and inhalation. NB is a probabilistic classifier based on Bayes Theorem, which is often used to calculate the most possible output for high-dimensional inputs [46]. NB only needs a small number of parameters to estimate a result, and the algorithm is relatively simple [47]. The DT model is one of the machine learning algorithms used to solve the decision problem. It uses the training data set to construct a decision tree, and the tree can classify unknown samples. The DT is efficient and useful for predicting unknown samples [48]. However, DT uses the “divide and conquer” strategy, which has poor performance for highly-correlated features. Moreover, DT is prone to over-fitting if it constructs a very deep tree [49]. Random forest is ensembled using a set of decision trees, which has a low risk of overfitting and a good generalised representation [50]. There is another typical machine learning classifier named SVM, whose classification strategy is to construct a hyperplane that contains the maximum boundary for discriminating the data. It is popularly applied in various areas, including medical diagnosis, wind power prediction and acoustic event detection [51]–[56]. ANNs, also referred to as connection models, are mathematics models intended to mimic (simplistically) the human brain. The models rely on the complexity of the system, whose nodes are connected to each other in particular arrangements. By self-adjusting the coefficients (weights) between the internal hidden layers, the problems of recognition and classification may be addressed with considerable effectiveness [57]. At present, deep learning methods are evolving rapidly, often based on the convolutional neural network (CNN) and the recurrent neural network (RNN); both are commonly used in industrial applications. The details of neural networks will be introduced in Chapter 3.

2.3.3 Conclusion

This section reviewed Alshaer et al.’s contributions to spectral characterisation and its use in the recognition of breath sounds [43], despite this method having poor generalisability. The other reviewed research on automatic identification all employ
stages that can be summarised in four steps: pre-processing, feature extraction, model training and prediction [32][44][45]. These studies used different feature extraction methods and classification models to recognise breathing for various applications. Ultimately, these contributions provided great insights on advances in research for breath recognition.

2.4 Review of electronic adherence measurement

2.4.1 Previous work conclusions

As has been stated, poor adherence to the inhaler technique may have severe consequences [38][39]. Electronic monitoring is a recent development and offers the potential to assess adherence and improve the quality of treatment. In 2011, Nelson et al’s study reported the use of telephonic monitoring to assess participants’ inhaler technique. The results confirmed that the inhaler technique can be established by sound [60]. Later, Foster et al. discussed the feasibility and acceptability of an electronic device to monitor patient dose uptake. They concluded the device could help professionals or researchers understand their patients in more details [61]. Holmes et al. used an acoustic monitoring device, named Inhaler Compliance Assessment (INCA), to track patients who used the Diskus inhaler. They developed an algorithm that used PSD, MFCC, singular value decomposition (SVD) and zero-crossing rate (ZCR) features to identify different acoustic events [42][43]. Following the study on the Diskus inhaler, Taylor et al. employed the continuous wavelet transform (CWT) to characterise and identify the acoustic actuation of pMDI inhalers [32]. In 2018, Taylor et al. used the INCA device to objectively assess pMDI use. This study employed quadratic discriminant analysis (QDA) to classify acoustic events and, further, investigated an audio-based approach to estimate the peak inspiratory flow rate (PIFR) [13].

Compared with traditional adherence measurement studies for asthma and COPD, automated acoustic assessment of inhaler technique still is a new field of study and research. The previous studies, reviewed in this paper, successfully assessed
adherence with the different approaches. This implies that acoustic monitoring is potentially a feasible methodology for quantifying adherence to inhaler use and technique.

2.4.2 Smart inhaler devices

Acoustic monitoring devices should be able to establish medication usage and determine the quality of adherence. Scott et al.’s study found that electronic devices determined that the median use of medication was 70.5%, but verbal reports and questionnaires overestimated the usage at 85.1% and 84.2% respectively [64]. Patients sometimes provided unreliable feedback, so data derived from direct questioning are not considered accurate for the purpose of measuring adherence [26]. Between 2016 and 2019, several smart inhaler devices were released on the market but most of them do not provide technical details. This section briefly introduces some smart inhaler devices as follows:

**INCA**

This device is manufactured by Vitalograph. The device includes a microphone, microcontroller and battery. It records the sound from the Diskus™ DPI device for medical analysis. The audio files are stored in flash memory and can be read via a USB connection. INCA commences recording audio when the Diskus inhaler is opened and stops when it is closed [12].

**Verihaler**

VeriHaler is a wireless acoustic monitoring inhaler; it provides feedback to the user or clinician for a range of metrics including peak inspiratory flow and cannister press timing. It can work with both pMDI and DPI inhalers [65]. However, VeriHaler is not an attachable device, which needs the patient to buy an extra inhaler designed by them. It cannot track the patient’s hand motion including vertical holding and shaking. The device is shown in Figure 2.5.
Hailie™

The Hailie™ provides medication reminders to the user. Further, it interfaces to smartphones, using an application that enables review of usage. However, it is only compatible with specific inhaler types and cannot assess the patient's inhaler technique. Hailie™ is shown in Figure 2.6.

Propeller

The Propeller sensor, shown in Figure 2.7, is not an acoustic monitoring device. It has been developed by Propeller Health, a digital therapeutics company. This smart sensor counts the doses taken and uploads the data to the cloud server via a mobile application. Propeller Health combines the smart device with the online service to
establish an online ecosystem for patients and health specialists. The purpose is to improve the quality of the patient’s asthma or COPD treatment [67].

![Image](image1.png)

Figure 2.7 Propeller sensor shown with the mobile application (Source: adapted from [67])

AIM (Aerosol inhalation monitor)
The AIM is an inhaler training device to provide clear feedback and information to support correct technique regarding inspiratory acceleration, the timing of actuation, inspiratory flow rate and breath-holding time. However, it cannot be implemented on an actual inhaler to track the patient’s use. The AIM device is displayed in Figure 2.8.

![Image](image2.png)

Figure 2.8 AIM was developed for effective training of inhaler use (Source: adapted from [68]).
**T-Haler**

The T-Haler is a training inhaler device developed by Cambridge Phenomenon Ltd. It includes interactive software, linked to the wireless device, that assesses the user’s pMDI technique. The real-time feedback is presented as an interactive video to instruct the user on the correct inhaler technique [69].

![T-Haler training device](image)

**Figure 2.9 T-Haler training device (Source: adapted from [70])**

### 2.4.3 Conclusion

A number of smart inhalers have appeared in recent years but few acoustic monitoring inhalers have been developed to assess inhaler technique in detail. Moreover, several devices have been developed for specific inhalers, such as INCA. In addition to an acoustic recording device, breathing analysis is urgently required. Based on the literature available to date, it is concluded that few studies have been conducted on the analysis of breath sounds associated with inhaler usage (for asthma and COPD sufferers); even fewer focus on the custom design of algorithms intended for acoustic signal processing and recognition. In terms of cost, most of the major manufacturers of inhalers are developing their own smart inhalers, but these are typically expensive, which presents challenges for publicly funded healthcare systems and in resource-poor countries. These features of smart inhalers are summarised in Table 2-4.
Table 2-4 Summary of smart inhalers

<table>
<thead>
<tr>
<th>Smart inhaler</th>
<th>Adapted inhaler</th>
<th>Inhaler technique</th>
<th>Hand motion</th>
<th>Track usage</th>
<th>Attachable</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCA</td>
<td>Diskus (^\text{TM})</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Verihaler</td>
<td>None</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hailie (^\text{TM})</td>
<td>pMDIs</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propeller</td>
<td>pMDIs</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIM</td>
<td>None</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>Training device</td>
<td></td>
</tr>
<tr>
<td>T-Haler</td>
<td>None</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>Training device</td>
<td></td>
</tr>
<tr>
<td>Smart inhaler developed in this thesis</td>
<td>pMDIs (To cover the most types of the inhaler in future)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>Training device</td>
</tr>
</tbody>
</table>

2.5 Conclusion

With an increasing number of patients with asthma and COPD, it is now desirable to develop a small, attachable and intelligent acoustic monitoring for use with inhales. This will improve adherence to medication and therefore reduce the financial burden on the healthcare system. This literature review summarised the relevant background and contributions to this field. The first part of the review attempted to clarify the challenges associated with inhaler usage, faced by both patients and health specialists. A comparison of usage steps and cost-effectiveness for the common types of inhaler were reviewed in this section. Following this discussion, a review was presented of ongoing research into breath sound analysis and common classification models for breath identification. A review of electronic adherence measurement devices concluded the chapter; this discussed both hardware and software devices and methods. These reviews not only considered the contribution but also provided a reference for further research.
Chapter 3

Background Theory

3.1 Overview

In relation to the development and analysis of the smart system, certain fundamental principles of signal and signal processing are discussed in this chapter. The first section covers basic signal processing theory, including Shannon sampling theory, the convolution operation, Fourier analysis and types of digital filer. This section provides essential background knowledge for analysis and characterisation of the acoustic signals. A typical speech feature extraction method - the mel frequency cepstral coefficients (MFCC) is introduced in the following section. To establish an effective method of acoustic recognition, this chapter discusses three types of classification methods, including statistical models, machine learning and neural networks. At the end of this chapter, two cross-validation methods are described to evaluate these trained models.

3.2 Foundation of Signal Processing

3.2.1 Shannon Sampling Theory

When sampling a signal, the minimum frequency of the sample should be considered. Otherwise, the inappropriate rate will cause distortion or aliasing. The Shannon sampling theory provides a principle to avoid aliasing. Regarding the Shannon sampling theory, the sampling frequency should be at least twice the highest of the real signal. The theory may be displayed in mathematical terms as follows

\[ f_s \geq 2f_c \]  
(3.1)

Where \( f_s \) is the sampling frequency, \( f_c \) is the highest frequency contained in the real signal.
Consider a single sine wave in which $f_c = 1 \text{ Hz}$; it is shown in Figure 3.1 (a). Setting the sampling frequency $f_s$ at 10 Hz, Figure 3.1 (b) and (c) confirm that the sampled signal contains all of the information from the original. However, if $f_s$ lower than $2f_c$, for example, as shown in Figure 3.2 (b), then information is lost, or aliased. This means the samples cannot describe key features, such as the peak of each, in the original signal. Therefore, when an under-sampled signal is reconstructed, as shown in Figure 3.2 (c), it will lose information.

![Figure 3.1](image1.png) Original signal 1 Hz, sampling frequency 10 Hz

![Figure 3.2](image2.png) Original signal 10 Hz, sampling frequency 10 Hz
3.2.2 Convolution

In a linear system, if a system function $h(t)$ is known, then the system response $y(t)$ can be obtained by convolution of $h(t)$ with the input $x(t)$. Mathematically, it is a close relative of the correlation operation, which is a time-reversal of convolution [71].

Assuming a system starts from time zero, the continuous system is given by

$$y(t) = \int_{-\infty}^{\infty} x(\tau) h(t-\tau) d\tau$$

(3.2)

In discrete space, if $x[n]$ and $h[n]$ are finite sequences, their convolution is expressed by

$$y[n] = \sum_{m=0}^{M-1} x[m] h[n-m]$$

(3.3)

Where $M$ is the length of the $x$.

Equation (3.3) can be explained as the new output $y[n]$ is obtained with the sum of the product of two finite sequences [71]. Figure 3.3 illustrates an example of a discrete convolution.

![Convolution](image)

Figure 3.3 Output signal $y[n]$ was convolved by $x[n]$ and $h[n]$
The convolution of input $x[n]$ and system impulse response $h[n]$ usually uses the asterisk as a shorthand notation. For example, the convolution in discrete space is

$$y[n] = h[n] * x[n]$$  \hspace{1cm} (3.4)

### 3.2.3 Fourier analysis

The Fourier analysis is a common mathematical tool used in signal processing; it decomposes a continuous or discrete time-domain signal into a frequency domain signal. This transform reveals the energy distribution of a signal as a function of frequency [71]. Fourier analysis includes the Fourier transform, Fourier series, discrete-time Fourier transform (DTFT) and discrete Fourier transform (DFT). The four transform functions are used to process different properties of signals [71][72]. In digital electronic systems, relevant signals are processed in discrete space. For example, breath sounds considered here are sampled 16000 times per second and stored on a memory card for further analysis. Thus, the DTFT and DFT are significant in this study. The DTFT and DFT are given by Equations (3.5) (3.6), respectively. According to the equations, the DTFT generates a continuous output as spectral representation, but the output of DFT is a discrete version of the spectrum [73]. In fact, the continuous output of the DTFT is rarely used to process discrete signals, because it is difficult to be computed in a digital system [71].

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n}$$  \hspace{1cm} (3.5)

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-\frac{2\pi jkn}{N}}$$  \hspace{1cm} (3.6)

Compared with DTFT, the DFT deconstructs discrete signal using a finite sum instead of an infinite sum. However, its $O(N^2)$ computational complexity is unacceptable for real implementation [74] in direct form. To reduce the computational burden, a more efficient algorithm, named the fast Fourier transform (FFT), optimises the DFT method
and decreases the complexity to $O(N \log_2 N)$ [71]. The key use of the FFT throughout the thesis is to analyse and characterise the acoustic signal, the results of which are then used for sound recognition.

### 3.2.4 The finite impulse response filter (FIR)

The use of digital filters is ubiquitous because of their flexibility and reliability in digital signal processing. In contrast to a digital filter, the components of an analogue filter will drift over time, temperature and voltage. FIR filter coefficients can be produced by the window function and the frequency sampling method, which are the most commonly applied techniques [71]. For an FIR filter of order $N$, each output value is a weighted sum of the $N$-length input $x[n]$. Namely, the process is a convolution of the filter coefficient sequence $h[k]$ and input sequence $x[n]$. The most common FIR filter form is shown in Figure 3.4, in which the FIR filter operation is described by

$$y[n] = h_0 x[n] + h_1 x[n-1] + \ldots + h_{N-1} x[n-N]$$

$$= \sum_{k=0}^{N} h[k] x[n-k]$$

(3.7)

![Figure 3.4 The typical FIR filter of order N block diagram](image)

### 3.3 Mel frequency cepstral coefficient (MFCC)

Mel frequency cepstral coefficient (MFCC) is a classic operation in speech signal processing, and it is based on the idea of cepstrum. The MFCCs feature extraction has
six steps to process each sequence of a segmented speech signal [75]. Figure 3.5 illustrated the extracting steps.

**Step 1 Pre-emphasis**
For the human voice, the physical characteristics of the glottal source act as a low-pass filter with an attenuation of approximately -6 dB/decade slope, [76]. To compensate for this spectral attenuation, the speech signal \( s(n) \) passes through a high-pass band filter \( H(Z) \); the Z-transform of the filter is given by

\[
H(z) = 1 - \alpha z^{-1}
\]  
(3.8)

where the parameter \( \alpha \) controls the attenuated slop of high pass filter.

The pre-emphasised signal \( s_2(n) \) is given by

\[
s_2(n) = s(n) - \alpha s(n-1)
\]  
(3.9)

Where the typical values of the filter coefficient \( \alpha \) are 0.95 or 0.97.

**Step 2 Framing and Hamming window**
The spectrum of a speech signal is continuously changing. It is impractical to extract sufficient information from an entire conversation as a single epoch. Therefore, it is
necessary to use a Hamming window to separate the signal into time fragments of a constant length. Firstly, we set $N$ sample points into one observation unit called a frame, i.e. one frame consists of $N$ sample points. The value of $N$ is chosen so that each frame is between 20 and 30 ms in duration [76]. To avoid excessive changes in between adjacent frames, frames are overlapped. The overlapped region includes $M$ sampling points, which is typically half and one third of the value of $N$. It is assumed that every signal frame is represented by $s_i(n)$. Thus, the windowed frame $x(n)$ can be described by

$$x(n) = s_i(n)w(n)$$

(3.10)

where the $w(n)$ is a Hamming window function and $s_i(n)$ represents the $i$-th frame of the pre-emphasised signal $s_2(n)$.

The Hamming window $w(n)$ is given by

$$w(n) = \begin{cases} (1 - \alpha_0) - \alpha_0 \cos\left(\frac{2\pi n}{N-1}\right), & 0 \leq n \leq N \\ 0, & \text{otherwise} \end{cases}$$

(3.11)

where $N$ represents the length of the window function; the parameter $\alpha_0$ decides the sharpness of Hamming window which is typically set to 0.46.

**Step 3 Fast Fourier transform (FFT)**

Following the framing and window step, the $x(n)$ is processed by the FFT, which yields the energy distribution of the windowed signal for different frequencies.

**Step 4 mel filter bank**

The characteristics of human hearing cannot represent a linear function. Therefore, the spectrum may be represented on the mel scale [77]. The mel is a unit of pitch defined by the relationship between the actual and perceived pitches [78]. Figure 3.6 illustrates the relationship between the mel scale and actual frequencies. The mel function is related to the actual frequency by
\[ Mel(f) = 2595 \log_{10}(1 + f / 700) \] (3.12)

Figure 3.6: The Ratio of Actual Frequencies to Mel Scale

Human hearing is more sensitive to differences in the low-frequency range than in the high-frequency range. During the MFCC computation, the Mel filter bank will output the energy of a single frame. The bank has \( M \) triangle filters, whose centre frequencies are \( f_m, m = 1, 2 \ldots M \). Prior speech recognition studies report that the value of \( M \) is usually taken as 20, 24, 29 and 40 [79]. These triangular band-pass filters are spaced unequally along the mel scale. Figure 3.7 illustrates the MFCC FB-40 filter bank that was introduced in Slaney’s study. The first 13 filters, below 1000 Hz, are spaced linearly and the following 27 filters, above 1000 Hz, are spaced logarithmically [79]. The MFCC FB-40 filter bank is defined as

\[
H_m(k) = \begin{cases} 
0 & , k < f(m-1) \\
\frac{2(k - f(m-1))}{(f(m+1) - f(m-1))(f(m) - f(m-1))} & , f(m-1) \leq k \leq f(m) \\
\frac{2(f(m+1) - k)}{(f(m+1) - f(m-1))(f(m) - f(m-1))} & , f(m) \leq k \leq f(m+1) \\
0 & , k > f(m+1)
\end{cases} 
\] (3.13)

where filter \( m \) is the range from 1 \( \ldots \) \( M \), \( f(m) \) represents the \( m \)-th mel-spaced frequency, and \( k \) is the bin of \( N \)-point FFT.
The next step computes the logarithm of \( H_m(k) \). The expression of log energy is given by:

\[
\text{log}_{10}(\sum_{k=0}^{N-1} (X(k))^2 H_m(k))
\]

(3.14)

where \( X(k) \) is the fast Fourier transform of the frame \( x(n) \).

**Step 5 Cepstral coefficients: Discrete Cosine Transform (DCT)**

In the previous step, the overlapped triangular band-pass filters cause the adjacent log energy \( s_{log}(m) \) to be correlated. The typical MFCC extraction employs DCT type II to decorrelate \( s_{log}(m) \) and only keeps the major coefficients [79]. Therefore, the cepstral coefficients are described as:

\[
c(n) = \sum_{m=1}^{M} s_{log}(m) \cos(n \frac{\pi(m-1/2)}{M}), \quad n = 1, 2, ..., C
\]

(3.15)

where \( M \) is the number of filters in filter bank and \( C \) represents the number of cepstral coefficients.
Step 6 Dynamic features

The standard cepstrum can only extract the statistic characteristic of a speech, but the
dynamic characteristic is also an important feature in speech recognition. Therefore,
dynamic information can be calculated by the first and second-order derivatives of
cepstral coefficients. The expression can be simply expressed as:

\[ \Delta d = \frac{c(n+1) - c(n-1)}{2} \]  

(3.16)

where values \( c(n) \) at sequence \( n \) are the cepstral coefficients.

The energy within each windowed frame \( x(n) \) also contains useful information. Thus,
it should also be considered. It is given by

\[ \text{Log Energy} = 10 \log_{10} \left( \sum_{n=0}^{N-1} x(n)^3 \right) \]  

(3.17)

Assuming the number of cepstral coefficients \( C = 12 \), the typical MFCC computation
finally output features according to table 3.1.

Table 3-1 Typical MFCC features

<table>
<thead>
<tr>
<th>12</th>
<th>Cepstral Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Derivatives of Cepstral Coefficients</td>
</tr>
<tr>
<td>12</td>
<td>The Second Order of Derivatives of Cepstral Coefficients</td>
</tr>
<tr>
<td>1</td>
<td>Logarithmic Energy Coefficients</td>
</tr>
<tr>
<td>1</td>
<td>Derivatives of Energy Coefficients</td>
</tr>
<tr>
<td>1</td>
<td>The Second Order of Derivatives of Energy Coefficients</td>
</tr>
<tr>
<td>39</td>
<td>MFCC features</td>
</tr>
</tbody>
</table>
3.4 Statistical Models

3.4.1 Gaussian mixture model (GMM)

A probabilistic model, which is used to describe normal distributions of subdivisions within an overall population, is considered a Gaussian mixture model. Gaussian distribution is popularly applied in various engineering areas and science disciplines [80]. When $x$ is a continuous random variable, its probability density function of Gaussian distribution is defined by

$$p(x) = \mathcal{N}(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$  (3.18)

If variable $x$ is normally distributed, it can be notated by

$$x \sim \mathcal{N}(\mu, \sigma)$$  (3.19)

where mean $\mu$ and covariance $\sigma$ are parameters in this probability density function.

According to the law of large numbers, the Gaussian distribution has the desirable property of fitting natural, real-world data. The curve in Figure 3.8 shows a typical Gaussian density probability distribution.

![Figure 3.8 Gaussian probability distribution](image-url)
For the Gaussian-mixture distribution, if variable \( X \) is a random variable vector, the multi-dimensional GMM is defined by

\[
p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)
\]

\[
= \sum_{k=1}^{K} \frac{\pi_k}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)}, (\pi_k > 0)
\]

where \( \pi_k \) is the mixture weights, \( K \) is the number of components, expectation \( \mu \) and covariance matrix \( \Sigma \) are given by

\[
E(x) = \mu
\]
\[
E[(x - \bar{x})(x - \bar{x})^T] = \Sigma
\]

### 3.4.2 Hidden Markov model (HMM)

Machine learning provides several methods to process sequential data; these methods include the recurrent neural network (RNN) and the hidden Markov model (HMM). HMM is still one of the most popular methods in the application of speech recognition and other sequential predictions; the HMM method has existed for a long time [81]. Essentially, HMM is a statistical model. It assumes that a sequence of latent variables, in which previous and next variables are conditionally dependent given a current one, generates a Markov chain. Further, each latent state corresponds to the observed state [82]. Figure 3.9 shows a graphical representation of the Markov state space structure, which joins distributions of observations in this model; the joint distribution function is given by

\[
p(x_1, \ldots, x_N, z_1, \ldots, z_N) = p(z_1) \prod_{n=2}^{N} p(z_n \mid z_{n-1}) \prod_{n=1}^{N} p(x_n \mid z_n)
\]

where \( z_n \) represents the \( n \)-th latent state and \( x_n \) is the \( n \)-th observed value.
The hidden Markov model can be considered a specific instance of the aforementioned state-space model [82]. A diagram Figure 3.10 illustrates that three hidden variables have three possible states corresponding to coloured circles. The arrows represent the transitions of states.

Since there are various types of the latent variables, such as labels and integral numbers, it is common to express these variables in the form of a binary format. Thus, one-hot encoding, also termed one-of-k, is employed to represent the latent variable $z_n$. One-hot encoding is a $K$-dimensional binary vector, which only has one element equal to one and others must be zero [83]. Namely, the encoding vector indicates $K$ possible states and the $k$-th nonzero element corresponds to the state $z_k$. Therefore, if the latent variables are $K$-dimensional binary vectors, their transition probabilities are put into matrix $A$ that has $K(K-1)$ elements, excluding the self-loop probability. The transition matrix $A$ is shown in Figure 3.11 and it is given by

$$A_{jk} = p(z_{nk} = 1 \mid z_{n-1,j} = 1)$$

$$\sum_k A_{jk} = 1$$  \hspace{1cm} (3.23)

The conditional probabilities can be written in terms of transition probabilities as
\[
p(z_n \mid z_{n-1}, A) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^z_{n-1, jk} \tag{3.24}
\]

However, there is no previous node before \(Z_1\) while \(n = 1\). Therefore, in this situation, using a vector of probabilities \(\pi\) to represent marginal distribution \(p(z_1)\) denote by

\[
p(z_1 \mid \pi) = \prod_{k=1}^{K} \pi_k^{z_1} \sum_k \pi_k = 1 \tag{3.25}
\]

where \(\pi\) is an HMM parameter that initialising each first possible state.

For latent and observed variables, their joint distribution is defined by

\[
p(X \mid Z, \theta) = p(z_1 \mid \pi) \prod_{n=2}^{N} p(z_n \mid z_{n-1}, A) \prod_{m=1}^{N} p(x_m \mid z_m, \phi) \tag{3.26}
\]

Where \(\theta = \{\pi, A, \Phi\}\), \(\Phi\) is a set of parameters governing the distribution. For instance, if this probability subject to multivariate Gaussian distribution, \(\Phi\) is denoted by \(\{\mu, \Sigma\}\).

![Transition diagram of latent variables](image.png)
3.5 Machine Learning

3.5.1 Support Vector Machine (SVM)

In machine learning, a support vector machine (SVM) is a typical classifier that constructs a hyperplane and then finds the maximum boundary to discriminate data, especially for sparse dataset [84]. It has been applied in various areas, including medical diagnosis, object detection, image segmentation and wind power prediction [51][52][54][85][86].

For a linear SVM, if the dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}, y \in \{-1, 1\}$ is linearly separable, a decision function is defined by

$$f(x) = \omega^T \phi(x) + b$$

where $\omega$ is the coefficient and $b$ is the $y$-intercept; $\phi(x)$ represents a transformation for $\phi$-space [87].

Figure 3.12 illustrates the basic conception of a linear SVM [88]. The $x$ locates on hyperplane while $f(x) = 0$; if $f(x) > 0$, $x$ can be classified as $y = 1$, otherwise $y = -1$ when $f(x) < 0$. Therefore, $|f(x)|$ indicates the distance from the hyperplane, and it is also equal to the functional margin denoted by $yf(x)$. Although the functional
margin indicates the position of x with respect to the separating plane, it does not give an exact magnitude.

![Figure 3.12 Maximum boundary in binary classification](image)

The geometric distance is employed to solve this problem. According to Equation (3.27), \( \omega \) is orthogonal for all vectors on the hyperplane. Assuming \( x' \) is an orthogonal projection of the arbitrary point \( x \) onto the hyperplane. Thus, the point \( x \) of \( \phi \)-space is given by

\[
\phi(x) = \phi(x') + L \frac{\omega}{\|\omega\|}
\]  

(3.28)

where \( L \) is the perpendicular distance from point \( x \) to the hyperplane.

As aforementioned, \( x' \) lies on the plane, which means \( f(x') = \omega^T x' + b = 0 \). Taken (3.28) into (3.27), we can get
\[ f(x) = \omega^T \phi(x) + L \frac{\omega}{\|\omega\|} + b \]  
(3.29)

\[ L = \frac{|f(x)|}{\|\omega\|} = \frac{|y(x)|}{\|\omega\|} \]  
(3.30)

The perpendicular distance \( L \) of the point \( x \) is given by (3.30), which is also known as the geometric distance. As shown in Figure 3.12, dataset \( X \) is separated by a red line. The maximum margin \( \gamma \) is the geometric distance of the closest points. In terms of the classified decision, the large value of \( \gamma \) ensures a high certainty of the decision and vice versa. Assuming there is an existence of maximum \( \hat{\gamma} \), the constraint inequality is given by

\[ \frac{y_n f(x_n)}{\|\omega\|} = \gamma_n \geq \hat{\gamma}, \quad n = 1, \ldots, N \]  
(3.31)

The classification problem is converted to find a maximum margin. Thus this basic SVM classifier is also called a maximum margin classifier. This algorithm aims to search the appropriated \( \hat{\gamma} \) by optimisation of \( \omega \) and \( b \). The progress is

\[ \arg \max_{w, b} \frac{1}{\|\omega\|} \min_n y_n f(x_n) \]  
(3.32)

If \( \omega \) and \( b \) are rescaled by ratio \( \kappa \), the geometric distance of \( x_n \), in (3.31), will not be changed, because the functional margin \( y_n f(x_n) \) is normalised by the Euclidean distance of \( \omega \). Namely, \( \kappa \) changes \( y_n (\kappa \omega^T \phi(x) + \kappa b) \) to a different scale, but it is invariant for the hyperplane [89]. To simplify (3.32), the boundary of the closest points is defined as

\[ y_n (\omega^T \phi(x_n) + b) = 1 \]  
(3.33)

For all data points, the inequality must be satisfied as follow:
Equation (3.32) reduces to finding the maximum of $\|\omega\|^{-1}$, which is same as minimising $\|\omega\|^2/2$. Therefore, the problem can be written as

$$\text{minimise} \quad \frac{1}{2}\|\omega\|^2, \quad \text{subject to} \quad y_n(\omega^T \phi(x_n) + b) \geq 1, \quad n = 1, \ldots, N$$

where multiplied $\frac{1}{2}$ is used to simplify the calculations of the following equations.

Equation (3.35) is shown a special case of convex optimisation called the quadratic programming (QP) problem, in which the function is subjected to a linear inequality constraint. To solve this problem, the general solution usually causes high computational complexity. Therefore, Lagrange multipliers are introduced to transform the problem into dual space, that makes the optimisation more efficient and more easily applied to high-dimensional feature space [83][90]. In this way, each of the constraints is multiplied by Lagrange multipliers $a_n$; constructing the Lagrangian function

$$L(\omega, b, a) = \frac{1}{2}\|\omega\|^2 - \sum_{n=1}^{N} a_n (y_n(\omega^T \phi(x_n) + b) - 1),$$

subject to $a_n \geq 0, \quad n = 1, 2, \ldots, N$,

finds a stationary point, which is a minimum of $\|\omega\|^2/2$ with respect to $\omega, b$ and is maximised with respect to $a_n$. To compare with Equation (3.35), the Lagrangian function merges the linear constraint into itself. Thus, the optimisation is equivalent to

$$\arg \min_{\omega, b} \left\{ \arg \max_{a_n \geq 0} L(\omega, b, a) \right\} = p^*$$

where $p^*$ represents the optimal solution. If the minimising and maximising steps are swapped, the dual representation can be evaluated as
\[
\text{arg max}_{a_{i,0}} \left\{ \text{arg min}_{w,b} L(\omega,b,a) \right\} = d^*
\]

(3.38)

The primal problem \(p^*\) is transformed into the dual problem \(d^*\) and the values satisfy weak duality \(d^* \leq p^*\).

As aforementioned, the optimal problem (3.35) is a special convex optimisation; if the inequality constraints hold, it also holds with Slater’s condition [91]. Accordingly, if \(g_n(x)\) is denoted by \(y_n(\omega^T \phi(x_n) + b - 1)\), the optimisation has a stationary point while satisfying the Karush-Kuhn-Tucker (KKT) conditions

\[
g_n(\hat{x}) \geq 0, \quad n = 1, \ldots, N \\
a_n g_n(\hat{x}) = 0, \quad n = 1, \ldots, N \\
a_n \geq 0, \quad n = 1, \ldots, N
\]

(3.39)

where \(\hat{x}\) is an arbitrary point in the primal function or dual function [83]. According to the KKT condition, if \(a_n = 0\), any point \(x_n\). Moreover, when the KKT conditions holds, the duality gap equals to zero, and then a strong duality \(d^* = p^*\) holds.

In order to meet the maximum and minimum conditions, partial derivatives of \(L(\omega,b,a)\) with respect to \(\omega\) and \(b\) are equal to zero, i.e.

\[
\frac{\partial L(\omega,b,a)}{\partial \omega} = \omega - \sum_{n=1}^{N} a_n y_n \phi(x_n) = 0
\]

(3.40)

\[
\frac{\partial L(\omega,b,a)}{\partial b} = -\sum_{n=1}^{N} a_n y_n = 0
\]

(3.41)

Substituting (3.40) and (3.41) into the target function (3.36) gives

\[
\tilde{L}(\omega,b,a) = -\frac{1}{2} \sum_{n,m=1}^{N} a_n y_n \phi(x_n)^T a_m y_m \phi(x_m) + \sum_{n=1}^{N} a_n y_n - b \sum_{n=1}^{N} a_n y_n
\]

\[
\tilde{L}(\omega,b,a) = \sum_{n=1}^{N} a_n y_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m y_n y_m \phi(x_n)^T \phi(x_m)
\]

(3.42)
Although this section only discusses the linear-separable condition, how SVM discriminates non-linear features is worth mentioning at this stage. The direct method, to construct a non-linear SVM, is to map the original features onto \( \Phi \)-space where mapped features are separable [92]. However, if the dimension \( D \) of \( \Phi \)-space is much larger than the dimension \( d \) of the original space, the large number of parameters in the higher space imposes an impractical computational burden. To avoid explicitly calculating the \( \Phi \)-space transformation, the kernel function \( k(x_n, x_m) \) is applied to SVM [93], which can be treated as a similar representation of the transformed space. This simplifies the computation of \( \Phi(x_n)^T \Phi(x_m) \) [92][93]. Thus, Equation (3.42) can be written as

\[
\tilde{L}(\omega, b, a) = \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m y_n y_m k(x_n, x_m)
\]  

(3.43)

From the above equation, maximising \( \tilde{L}(\omega, b, a) \) solely depends on \( a_n \) and is subject to \( a_n \geq 0 \) and \( \sum_{n=1}^{N} a_n y_n = 0 \). In this QP problem, an approximate value for \( a_n \) can be quickly obtained by the sequential minimal optimisation (SMO) algorithm, which solves the QP problem with lower computational complexity [94].

Having found values for \( a_n \), \( \omega \) is obtained by (3.40). For any \( \tilde{x}_n \), if its \( a_n \neq 0 \) and the condition \( \tilde{y}_n f(\tilde{x}_n) = 1 \) is satisfied, it is called a support vector [83]. Thus all support vectors \( \tilde{x}_n \) have a set of indices \( S \); then \( b \) is given by combination of (3.27) and (3.40), i.e.:

\[
b = \frac{1}{N_S} \sum_{n \in S} \left( y_n - \sum_{m \in S} a_m y_m k(x_m, x_n) \right)
\]  

(3.44)

where \( N_S \) represents the number of support vectors.

In terms of the binary classification, the class of input \( x \) can be identified by the

\[
f(x) = \sum_{n=1}^{N} a_n y_n k(x_n, x) + b
\]  

(3.45)
In Equation (3.45) as stated, if $f(x) > 0$, the input $x$ can be classified to 1 and otherwise it outputs 0 when $f(x) < 0$. According to Equation (3.36), the multiplication only needs to operate on support vectors, because $\alpha_n$ equals to zero while the trained datapoint is a non-support vector. In this case, the classification function is always equal to $b$ and does not need extra computation.

If feature-space function $\phi(x)$ equals to $x$ itself, the classifier is known as a linear SVM classifier. However, the traditional linear SVM cannot be implemented on a non-linear dataset (shown in Figure 3.13). To solve this problem, SVM maps, by using kernel functions $k(x_n, x_m)$, non-linear separable dataset up to high dimensions (illustrated in Figure 3.14), and then finds a hyperplane to separate dataset [93].

Figure 3.13 This dataset is a non-linear dataset and consist of two features (X, Y).
Figure 3.14 The additional feature $Z = (X - Y)^2$ makes the non-linear dataset separable.

3.5.2 Random Forest

Random forest is an ensemble learner bagged by a set of decision trees. For a better understanding of the random forest method, it is necessary to briefly introduce the decision tree-based method – classification and regression trees (CART) [95].

To build a tree model, the basic idea is to split the dataset, by given conditions, into homogeneous groups. For example, there is a diagnostic dataset, in Wisconsin, from the UCI repository, which contains features of breast cancer, such as radius, texture, perimeter and other identified features [96]. The generated decision tree is obtained by training the algorithm with 100 samples; its simplified decision-making progress is illustrated in Figure 3.15.
Figure 3.15 Binary tree classification (100 samples)

For tree-based methods, the significant key-point is split. The algorithm of the decision tree is responsible for splitting the nodes appropriately [97]. Here, an impurity metric, such as Gini index, entropy and information gain, is used to provide the splitting threshold based on a measurement of purity of each node [98]. To simplify the understanding of CART, consider that there is a $N$-sample dataset $X = (X_1, X_2)$ and the dataset contains four groups. In this decision process (see Figure 3.16), the Gini index is used to calculate the impurity of $N_t$ samples in the $t^{th}$ node, defined by

$$Gini(t) = \sum_k p(k \mid t)(1 - p(k \mid t))$$

$$= 1 - \sum_k p(k \mid t)^2$$

$$= 1 - \sum_k \left( \frac{|C_k|}{|N_t|} \right)^2 \quad k = 1, 2, ..., K$$ (3.46)

where $C_k$ represents the amount of the $k^{th}$ class in node $t$. Equation (3.46) shows that the amount of categories contained in the samples is proportional to the Gini index. Therefore, the split node depends on the value of the Gini index.
Figure 3.16 illustrate that the decision tree has understandable and clear decision progress; it is a popular predictor method with applications in several areas, including power quality disturbance detection, financial assessment and disease risk ranking [99][100][101]. However, the tree predictor, because of its topology, has high-variance. If parental nodes changed slightly, it will generate a totally different series of children nodes. Furthermore, a decision tree suffers from the risk of overfitting and loss of generality for new dataset [98].

In order to reduce the overfitting of a single decision tree and improve its accuracy, the random forest is an appropriated algorithm. As a modification of the bootstrap
**aggregating** (bagging) predictor, the random forest builds several independent weak-learners (decision trees) and outputs result by averaging or voting [50][102].

Comparing with Leo Breiman’s bagging method, the random method not only replicates bootstrap samples but also randomly draws learning features for each node of the generated tree [102]. Consider $X\{(x_n, y_n)\}, n = 1,2, \ldots, N$, where $x_n$ is a $K$ features vector and $y_n$ is a class label or numerical value. Booststrap samples $X^{(B)}$ are randomly drawn with replacement from the original dataset $X$. Then, each tree in the random forest is individually trained on these samples. For a single tree, the random forest only uses a subset of total features to split each node, where the typical size of the subset is equal to round$(\log_2(K + 1))$ [50]. The procedure of generating the random forest is shown in Table 3-2.

<table>
<thead>
<tr>
<th>Table 3-2 Random forest pseudocode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bootstrap samples $X^{(B)}$ are randomly drawn from original $X$.</td>
</tr>
<tr>
<td>2. For $b = 1$ to size $(X^{(B)})$</td>
</tr>
<tr>
<td>Generate a tree:</td>
</tr>
<tr>
<td>Repeat steps (a), (b) until the tree approach maximum size:</td>
</tr>
<tr>
<td>(a). $k$ random features drawn from total $K$ features.</td>
</tr>
<tr>
<td>(b). Split node by $k$ features.</td>
</tr>
<tr>
<td>End</td>
</tr>
<tr>
<td>3. Ensemble model $T_b$ outputs result by voting or averaging.</td>
</tr>
</tbody>
</table>
3.6 Neural Networks

3.6.1 Convolutional neural networks (CNNs)

As one of a class of deep learning neural networks, convolutional neural networks are the most popular in computer vision [103][104][105]; recently they have found application in acoustic areas, such as speech recognition and acoustic events classification [106][107]. Compared with neural networks layer [108], the convolutional layer uses local connections with shared weights for the input dataset, so convolutional neural networks have a significant reduction in the number of input parameters [109]. An ordinary convolutional neural network is shown in Figure 3.17. The convolution layer, pooling layer and fully connected layer are basic layers of convolutional neural networks.

The convolutional layer is the core layer of the convolutional neural network; it will output the dot product of the input with a set of learnable filters that is also called a kernel. Output properties are depended on serval hyperparameters, such as kernel size $K$, stride $S$, zeros-padding $P$ and depth $D$ [109]. Figure 3.18 illustrates the convolution procedures. Zero-padding is used to ensure that the output has the same width and height of the input; the stride represents the filter shifting distance at one time; the depth (kernel) is the number of the filters. Assuming $W$ is the input width-height, the size of the output $W'$ is given by

$$W' = \frac{W - K + 2P}{S} + 1$$  \hspace{1cm} (3.47)

For example, in Figure 3.18, $K = 3, S = 1, P = 1$ and $W'$ is calculated by $\frac{5-3+2\times1}{1} + 1 = 5$. There are two filters, which means the output depth $D = 2$. Thus, the final output dimension is $(5 \times 5 \times 2)$. 

60
The pooling layer is obtained using a down-sampling operation. It is used to reduce the input scale and the number of training parameters; it can further control overfitting. There are two operations in general pooling layers: average pooling and max pooling. The max-pooling layer was the most popular in the majority of deep learning architectures; its operation is illustrated in Figure 3.19. More recently, average pooling is frequently implemented in new deep learning architectures, such as the residual network, GoogleNet and Xception [104][110][111].
The fully connected layer is a layer that connected to all the neurons of adjacent layers (see Figure 3.20). In the traditional artificial neural network, it is also named as the hidden layer. In the deep learning network, this layer is used to combine the high-level features from the previous layer and output the result to the final layer.

Finally, the activation functions are introduced, which imitate the transmission of synaptic signals from the axon to the dendrites. These functions map the output of the previous layer into specific ranges and transmit then to the next layer. The sigmoid function was commonly used for this purpose several years ago, but it was replaced
by a rectified linear unit (ReLU), which has more efficient training when involving many layers \([108][113]\). The activation functions are given below

- **Sigmoid**

  \[
  \sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.48}
  \]

- **Softmax**

  \[
  f_i(x) = \frac{e^{x_i}}{\sum_{j=1}^{J} (e^{x_j})}, \quad i = 1, ..., J \tag{3.49}
  \]

- **Tanh**

  \[
  \tanh(x) = 2\sigma(2x) - 1 \tag{3.50}
  \]

- **ReLU**

  \[
  f(x) = \max(0, x) \tag{3.51}
  \]

ReLU can considerably improve the performance of deep learning networks and is widely used in recent deep learning studies \([110][111][114]\). However, if \(x\) becomes negative, the ReLU neuron will stick and always output zero; this is termed dying ReLU \([115]\). To avoid this problem, the leaky ReLU was introduced and is given by

\[
 f(x) = \begin{cases} 
 x & \text{if } x \geq 0 \\
 \alpha x & \text{if } x < 0 
\end{cases} \tag{3.52}
\]

3.6.2 **Recurrent neural networks (RNNs)**

A sequence to be analysed, such as an audio signal, has variable lengths of input and output, which causes a huge challenge for standard neural networks. Historically, Hidden Markov models (HMMs) were applied in the recognition of such sequences, including speech, biological sequences and handwriting \([77][116][117]\). However, HMMs are only able to describe the dependency between adjacent states \([118]\), which causes a lack of contextual information and limits the improvement of recognition.
RNN has been successfully implemented for natural language processing (NLP) [119][120]. An RNNs is shown in Figure 3.21. Successive outputs \( y_t \) correspond to computations of previous inputs and outputs. In step \( t \), the hidden layer calculates \( a_t \) using the previous step state \( a_{t-1} \) and input \( x_t \). In this way, the previous states can be remembered and continuously have influence on forward states.

![Figure 3.21 A recurrent neural network with forwarding computation.](image)

For the \( t^{th} \) hidden layer, the forward propagation is given by

\[
\begin{align*}
a_t &= g(a_{t-1}W_{aa} + X_tW_{xa} + b_x) \\
\tilde{y}_t &= g(a_tW_{ya} + b_y)
\end{align*}
\]

(3.53)

where \( g(x) \) denotes activation functions, such as ReLU, Sigmoid and Tanh; \( W_{aa} \) is the weight of the previous state, \( W_{xa} \) represents the weight of its input value and \( W_{ya} \) is the weight at the output; \( a_t \) is the hidden state at step \( t \).

Although the RNN is a very effective method of sequence processing, it does not work on long-term sequences. In RNN training, the weight of each neuron is updated by the backpropagation algorithm which calculates the partial derivative of the loss function backwards. Due to the chain-connection structure of the RNN, the long sequence causes the gradient to easily vanish, and finally prevents the weights changing [121].
Recent research uses advanced RNN architectures, such as the gated recurrent unit (GRU) or long short-term memory (LSTM) units, to avoid its disadvantage [122][123]. The LSTM units will be introduced in chapter 5.

3.7 Cross-validation methods

Cross-validation is an evaluation method for training learners, which can improve the final predicted results [124]. The basic idea of this method is to split the training dataset into two, with unequal sizes. The large one is used for training and then the trained model is evaluated using the small dataset.

3.7.1 Hold-out method

The most convenient implementation of cross-validation is the Hold-out method. The input dataset is divided into a training set and test set. The learner is trained on the training set and the test set is only used for evaluation. However, any inappropriate separation may cause the imbalance of training samples, thereby increasing the number of test errors [125][126].

3.7.2 K-fold method

K-fold is a very typical cross-validation method. The whole training progress will repeat K times. For each time, the dataset is split into K equal size sets; the $k^{th}$ set is selected for evaluation and the other $K - 1$ sets are used for the training set.

3.8 Conclusion

The methods and theories described above supported the development of the whole software system of the automated inhaler monitor. However, this chapter covered only the most pertinent aspects of signal processing theory and analysis. The extraction method based on MFCC feature analysis was discussed, whose features are
a good representation of the major components of the original acoustic signals. Following the feature extraction, the background of typical recognition models, e.g. HMM, SVM, and CNNs, were introduced in detail. This chapter also considered two cross-validation methods, Holdout and K-fold, which are used to evaluate learning models.
Chapter 4

Hardware System Design

4.1 Overview

In general, the system comprises an acoustic monitoring device attachable to an inhaler, and software to track patients’ usage and technique. Hence the hardware and algorithms are two significant parts of this system. The hardware provides access to collect patients’ inhaler use data; the algorithms are used to automatically identify whether patients use their inhaler in compliance with recommended steps [127]. This chapter will consider techniques, evaluation tools and hardware designs from the initial idea to the final design.

As the previous chapter mentioned, inhalers have different types and shapes. Therefore, the design needs to be compatible with different shapes of inhalers. The recording device also needs to be small (<40x40mm), due to the size of the inhaler. Furthermore, the space constraints of the device will limit the size of a lithium battery. For a lithium battery with the expected dimension, its capacity would be no more than 300 mAh. Therefore, the recording device requires the use of low-power components and optimized firmware to reduce power consumption. Moreover, it is a challenge to identify reliably when a patient uses the inhaler. The acoustic monitoring device is required to have the following features:

- Ultra-low power consumption
- Rechargeable
- Record patients sound in high quality for advanced analysis
- Small size
- User acceptability (e.g. comfort and ease of use)
- Low interference with the usual usage of the inhaler
4.2 Technical background

4.2.1 FAT32 File System

The FAT32 standard is based on the File Allocation Table (FAT) file system, which can mount a drive as large as 127 GB. The maximum possible size for a single file is 4 GB [128]. The file system has five partitions: Boot Sector, Superblock, File Allocation Table, Root Directory and Data Field [129]. Figure 4.1 shows the FAT file system partition. The first section is the Boot Sector that contains address information of the disk data. The second, Superblock, contains a small fixed-sized cluster, which stores system information, such as version information, total number of blocks and so forth. Following the Superblock is the File Allocation Table; it stores block indices that map out blocks to the physical disk. The Root Directory is the main entry, namely the first block of each file and folder. The last space is divided into 4kB blocks for storing file data.

![Figure 4.1 Typical FAT file system partition](image)

4.2.2 FatFs system module

FatFs is a multi-platform system module for FAT / exFAT file, which is programmed in compliance with ANSI C (C89) and completely isolated with a physical disk. Therefore, FatFs can be used as an independent platform transplant into microcontrollers, such as 8051, PIC, STM32 etc. [130]. Figure 4.2 illustrates the module system organisations. User application uses the f_open() function to open the target file and operate the file. The advantage of this is the user application does not need to know how to control the low-level I/O.

Some of the module characteristics [130] are listed as follows:

- File system type: FAT12, FAT16, FAT32 and exFAT.
• The number of open files: Unlimited. (depends on available memory)
• Volume size: 2 TB at 512 bytes/sector.
• File size: 4 GB - 1 on FAT volume and virtually unlimited on exFAT volume.
• Cluster size: 128 sectors on FAT volume and 16 MB on exFAT volume.
• Sector size: 512, 1024, 2048 and 4096 bytes.

Figure 4.2 Typical operation flow with FatFs Module

4.2.3 SPI communication protocol

The Serial Peripheral Interface (SPI) is a serial protocol using four wires to communicate data between SPI Master and SPI Slaves[131]. An SPI device uses Master-Slave mode to exchange data with other SPI devices. The SPI interface separates the clock and the data line - it has two wires for sending and receiving data (MOSI, MISO), the third one is SCLK for synchronising the clock, and the last one is for selecting slaves (CS). A diagram of the master connecting to multiple slaves is shown in Figure 4.3.
When devices begin communication, the master device configures the clock and then transmits the clock signal to synchronise with the slave devices. If data transmission is working in full duplex mode, the master device sends a bit, in each SPI clock cycle, to the slave via the MOSI line. The slave will receive the bit from the master and then sends a bit back to the master indicating successful reception.

The SPI clock signal has four operational modes that depend on the clock polarity (CPOL) and phase (CPHA). The master usually configures it in compliance with the slave transmission method. When CPOL=0, the idle state of the clock is 0 and the active state is 1. For CPHA=0, data are captured on the clock's rising edge. For CPHA=1, data are captured on the clock's falling edge. When CPOL=1, the base value of the clock is high. Namely, the active state is 0 and the idle state is 1. For CPHA=0, data are captured on the clock's falling edge. For CPHA=1, data are captured on the clock's rising edge. Figure 4.4 shows the timing of the four modes. SCKL, SCLK1, SCLK2 and SCL3 show CPOL=0 & CPHA=0, CPOL=0 & CPHA=1, CPOL=1 & CPHA=0 and CPOL=1 & CPHA=1 respectively.
4.2.4 SDHC card protocol

The secure digital high capacity (SDHC) Card is defined in version 2.0 of the SD specification. The second version supports SPI, 1-bit SD transfer modes and 4-bit SD transfer modes within a clock range of 0 to 25 MHz [132]. Figure 4.5 shows the contact pads of the micro-SD card. In SD mode, all eight pads are connected to host-controller. However, in SPI mode, there are only six pins needed to connect to the host [133]. All the MMC/SDCs (including SDHC card) operate between 2.7 to 3.6 volts. Therefore, during the card initialization, the first thing is to check and confirm the supply voltage range, which can be read from the operating conditions register (OCR).

Figure 4.5 Pinout of Micro SD Card

SPI command set

This section introduces the protocol of SDHC cards working in SPI mode. Table 4-1 shows basic commands for SPI mode, such as generic read/write and initialization. (Note: SPI mode usually uses clock configuration 0 (CPHA=0, CPOL=0) to control SDHC cards.)
### Table 4-1 SPI Mode Command Set

<table>
<thead>
<tr>
<th>Command</th>
<th>Argument</th>
<th>Response</th>
<th>Data</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMD0</td>
<td>None</td>
<td>R1</td>
<td>No</td>
<td>GO_IDLE_STATE</td>
</tr>
<tr>
<td>CMD1</td>
<td>None</td>
<td>R1</td>
<td>No</td>
<td>SEND_OP_COND</td>
</tr>
<tr>
<td>ACMD41</td>
<td>2</td>
<td>R1</td>
<td>No</td>
<td>APP_SEND_OP_COND</td>
</tr>
<tr>
<td>CMD8</td>
<td>3</td>
<td>R7</td>
<td>No</td>
<td>SEND_IF_COND</td>
</tr>
<tr>
<td>CMD9</td>
<td>None</td>
<td>R1</td>
<td>Yes</td>
<td>SEND_CSD</td>
</tr>
<tr>
<td>CMD10</td>
<td>None</td>
<td>R1</td>
<td>Yes</td>
<td>SEND_CID</td>
</tr>
<tr>
<td>CMD12</td>
<td>None</td>
<td>R1b</td>
<td>No</td>
<td>STOP_TRANSMISSION</td>
</tr>
<tr>
<td>CMD16</td>
<td>Block length [31:0]</td>
<td>R1</td>
<td>No</td>
<td>SET_BLOCKLEN</td>
</tr>
<tr>
<td>CMD17</td>
<td>Address [31:0]</td>
<td>R1</td>
<td>Yes</td>
<td>READ_SINGLE_BLOCK</td>
</tr>
<tr>
<td>CMD18</td>
<td>Address [31:0]</td>
<td>R1</td>
<td>Yes</td>
<td>READ_MULTIPLE_BLOCK</td>
</tr>
<tr>
<td>CMD23</td>
<td>Number of blocks [15:0]</td>
<td>R1</td>
<td>No</td>
<td>SET_BLOCK_COUNT</td>
</tr>
<tr>
<td>ACMD23</td>
<td>Number of blocks [22:0]</td>
<td>R1</td>
<td>No</td>
<td>SET_WR_BLOCK_ERASE_COUNT</td>
</tr>
<tr>
<td>CMD24</td>
<td>Address [31:0]</td>
<td>R1</td>
<td>Yes</td>
<td>WRITE_BLOCK</td>
</tr>
<tr>
<td>CMD25</td>
<td>Address [31:0]</td>
<td>R1</td>
<td>Yes</td>
<td>WRITE_MULTIPLE_BLOCK</td>
</tr>
<tr>
<td>CMD55</td>
<td>None</td>
<td>R1</td>
<td>No</td>
<td>APP_CMD</td>
</tr>
<tr>
<td>CMD58</td>
<td>None</td>
<td>R3</td>
<td>No</td>
<td>READ_OCR</td>
</tr>
</tbody>
</table>

---

1. ACMD<n> means a command sequence of CMD55-CMD<n>.
2. HCS[30], reserved[31] and [29:0].
3. Supply Voltage Range [11:8], check pattern(0xAA) [7:0], reserved [31:12].

### SPI response format

When a host sends a command to an SD card, it will receive the command and return a response to the host. The responses are defined as R1, R2, R3 and R7 which combine with different formats and contain some information, such as errors, supply voltage and operation state [132]. Most commands return R1 (7-bit data). The R3/R7 response (R1 + trailing 32-bit data) is for only CMD58 and CMD8. If some commands require more than the standard response timeout value (NCR), they will respond with R1b. R1 and R3 formats are shown in Figure 4.6.
SD card initialisation in SPI mode

Before reading or writing to the SD card, it needs to be initialised by a series of commands; earlier versions of SD card do not support some commands, so the validation of these commands can be used to identify the SD card version. The initialisation of SD card in SPI mode is illustrated in Figure 4.7.

a) When inserting the SD card or power-on, wait for a minimum of one millisecond. The MCU sets the SPI clock rate between 100 kHz and 400 kHz, then pulls up DI and CS, and waits for least 74 SCLK cycles.

b) Send command CMD0 with pull-down CS to reset the SDHC card. If the software reset is successful, the CS pin should be still in low until the card responds R1 with 0x01, which means the card enters SPI mode. (Note: In SPI mode, the CRC is disabled so that sending the arguments of command CMD0 and CMD8 should be zero).

c) After resetting the SDHC card, the host sends CMD8 to the card. If the card returns data without an illegal command, this means the card is 2.0 Version or later SD Memory card.

d) The host sends CMD58 to read OCR to confirm compatible voltage ranges and then sends ACMD41 to set HCS (If the host supports high capacity, HCS is set to 1).

e) If the previous step returns ‘in_idle_state’=0, CMD58 will be sent to the card and it will return register data of CCS (CCS=1 High Capacity SD card, CCS=0 Standard Capacity SD card). Finally, the host receives the CCS data, which means the SDHC card initialization is successful.
4.2.5 Capacitive touch sensing design

Capacitive touch sensing is based on capacitive coupling of foreign objects which can influence the electric field of the capacitor, such as a finger [134]. Figure 4.8 shows a cross-section of a PCB-based capacitor.

The circular copper region and the surrounding ground pad represent a PCB-based capacitor. Leakage of an electric field occurs above the pad surface and forms a baseline capacitance when there are no foreign objects interfering with the field. The
typical capacitance is usually in the range of 10pF and the sensor cross-sectional size is typically 8mm - 10mm [135].

Figure 4.8 The design of the touchpad

4.3 Evaluation tools

4.3.1 MCU evaluation board

The 32L152CDISCOVERY (see Figure 4.9) uses an STM32L152RCT6 as its microcontroller, which features ultra-low power consumption. The board features include [136]:

- STM32L152RCT6 microcontroller, which is an ultra-low power consumption chip and has a 256 KB Flash memory, 32 KB RAM and 8 KB EEPROM
- On-board ST-LINK/V2 programming and debugging function
- Power supply on board: USB bus or external power supply from 3.3V or 5V
- External application power supply: 3V and 5V
- One capacitive sensor
- Extension header for the chip I/Os for quick connection to prototyping board and easy probing
4.3.2 Accelerometer Module

The accelerometer chip used consumes less than 2 µA at a 100 Hz output data rate and 270 nA in wake-up mode. Moreover, it has autonomous interrupt processing, without the need for microcontroller intervention. Its important features include the following[137]:

- SPI digital interface
- Adjustable threshold sleep/wake modes for motion activation
- Wide supply and I/O voltage range: 1.6 V to 3.5 V
- 10 nA standby current
4.3.3 Memory module

This module reads and writes data to the Nano-SD card controlled by the MCU via an SPI digital interface.

![Figure 4.11 SD Card Module](image)

4.3.4 Microphone modules

An electret condenser microphone was selected. Its module contained a pre-amplifier circuit with a frequency response range from 22 Hz to 15 KHz. This is a digital MEMS microphone and outputs a pulse-density modulated signal. The microphone has an acoustic overload point of 120 dBSPL with a 62.6 dB signal-to-noise ratio and -26 dBFS sensitivity[138]. It is shown in Figure 4.12.

![Figure 4.12 Digital MEMS microphone module (MP34DB02)](image)
4.3.5 PDM to PCM converter modules

ADAU7002Z is designed for converting a PDM bitstream into a PCM output. Its module is displayed in Figure 4.13. Its detailed features include:

- I/O supply operation: 1.62 V to 3.6 V
- 64× output sample rate PDM clock
- Automatic BCLK ratio detection
- Output sample rate: 4 kHz to 96 kHz
- Automatic power down with BCLK removal
- 0.67 mA operating current at 48 kHz and 1.8 V IOVDD supply
- Shutdown current: <1 µA
- 8-ball, 1.56 mm × 0.76 mm, 0.4 mm pitch WLCSP

![Figure 4.13 PDM to PCM Module](image)

4.4 Prototype – first version

The first prototype used an analogue microphone to record the sound. The top-level structure of the hardware is shown in Figure 4.14. The SD card and accelerometer connect to the SPI bus. The microphone connects to the microcontroller’s integrated amplifier, which feeds the analogue-to-digital converter (ADC). The touch sensor connects to two general-purpose I/O (GPIO) pins. The universal
Synchronous/Asynchronous Receiver/Transmitter (USART) is used to print out system logs. The serial Wire Debug (SWD) interface provides access to download or to update the firmware of the microcontroller.

Figure 4.14 The first version of the hardware design overview

4.4.1 Microcontroller unit

STM32L152CC is an ultra-low power consumption device. In standby mode, the chip only consumes 0.29 uA and can be activated by three wakeup pins. Moreover, it incorporates two operational amplifiers whose outputs can be programmed and connect to the internal ADC directly. The left part of Figure 4.15 illustrates the basic circuit of the MCU, including the crystal, boot mode and power supply. The right part shows various interfaces and connections with peripheral components.

In order to improve the analogue performance, the supply voltage sources VDD and VDDA were electrically separated, and the decoupling capacitors were placed as close as possible to the MCU. In Figure 4.16, the ‘A’ is a decoupling capacitor for digital block VDD, and ‘B’ is a label for decoupling capacitors for the analogue block VDDA. The power supplies should be placed close to the ground line to minimise the loop noise from the supply current. The reason is that the supply loop will act as an antenna may cause EMI emissions, which reduces the stability of the power supply.
4.4.2 Power management unit

The power management unit provides functions which include battery management, power consumption control and low-dropout through a regulator (LDO), Figure 4.17. The battery management IC U1 provides around 150 mA fast-charging current to the battery, and the termination threshold of current is 10% of fast charging current; the LDO U2 outputs 3.3 volts to the whole system; the digital and analogue ground have
separate planes and are bridged by the resistor R5. Load switches U3 and U4 are used to control the power supply of peripheral components.

4.4.3 Accelerometer circuit

The accelerometer circuit is shown in Figure 4.18. It transmits measured acceleration data to the MCU via the SPI bus. When the accelerometer detects motion, it will generate an interrupt event and wake up the MCU for further processing.
4.4.4 **Touch sensing circuit**

The capacitive touch sensor AT42AT101 is a digital burst mode charge-transfer sensor that can detect a touch with the proper electrode and circuit design. When the SYNC pin connects to ground, the chip will work in a low-power consumption mode. The sensitivity can be adjusted by choosing appropriate values of C22 and C23. Typical values are 2 nF to 50 nF, but larger values demand higher stability and better dielectric to ensure reliable sensing. The schematic is shown in Figure 4.19.

![Figure 4.19 Schematic of capacitive sensor](image)

The Cs (C22, C23) capacitor should be a high-stability type, such as X7R ceramic or PPS film. X7R ceramic types are low-cost and can be obtained in 5% tolerance. In some high sensitivity applications, PPS capacitors are the best choice. The PCB layout is shown in Figure 4.20.

![Figure 4.20 Capacitive sensor PCB layout](image)

*Error! Reference source not found.*
4.4.5 Microphone pre-amplifier circuit

The electret microphone has an excellent frequency response and is of reasonable cost. It produces a signal via varying its capacitance due to changes in sound pressure [140]. The schematic of the pre-amplifier is shown in Figure 4.21. R18 and R19 provide a DC bias for the internal JFET of MIC1. C14 and R22 couple the AC signal of the microphone to the amplifier. The input signal of op-amplifier is DC biased by MIC_3.3V/2 and for AC above around 88.42 Hz. C16 reduces the noise of the power supply (and of the equivalent R21 and R24 resistors), otherwise, the noise is likely to be amplified by the op-amp, along with the signal. R25 and C19 create a high-frequency gain. C18 is used to avoid parasitic oscillation and increases the stability of the amplifier circuit.

![Pre-amplifier circuit of the analogue microphone](image)

Figure 4.21 Pre-amplifier circuit of the analogue microphone

Every block (noisy, low-level sensitive, digital, etc.) should be grounded individually, and all ground returns should aim to a single point. If the microphone ground directly connects to the digital ground, the digital noise may interfere with the voice signal. Figure 4.22 shows that all the microphone signal tracks are separated from other parts by a copper pour layer. Furthermore, the central four capacitors highlighted in the
The figure below are tantalum, making the pre-amplifier circuit receptive to low-frequency signals.

![PCB Layout of microphone pre-amplifier](image1)

**Figure 4.22 PCB Layout of microphone pre-amplifier**

### 4.4.6 Memory unit

Figure 4.23 shows the Micro-SD circuit; it has eight pins, hence different to the normal SD circuit. CD1 (pin 9) is a pin for card insertion, i.e. the pin will be grounded when a card is inserted.

![Connection of micro-SD with SPI mode](image2)

**Figure 4.23 Connection of micro-SD with SPI mode**
4.4.7 Connector circuit

Figure 4.24 shows the firmware download and debug port, USB connector, touch interruption connector and UART connector.

![Schematic of Connectors](image)

Figure 4.24 the Schematic of Connectors

4.4.8 PCB layout

The PCB has four layers, i.e. the top and bottom are signal layers, the second is a ground plane and the third is a power plane. The MCU, analogue microphone, PMU circuit and accelerometer are on the top layer as shown in Figure 4.25 and the SD card connector and USB connector are on the backside (see Figure 4.26).
4.4.9 Firmware structure

The program has five modules to control different peripherals of the device. The `ff.c` file contains the Fatfs file system. When `audiorecord.c` the record function is executed, `ff.c` will mount the external memory, and then initialize the SD card via the driver named `sdio.c`. The structure of the other modules is shown in Figure 4.27.
Figure 4.27 Firmware structure
4.4.10 Hardware workflow

Initially, the device works in standby mode, in which the MCU shuts down most of the internal modules; all peripherals are also turned off, except for the accelerometer. If the motion is detected, the accelerometer generates an interrupt and activates the MCU. Then the MCU will turn on the capacitive sensor to detect whether a touch event occurred. This step is repeated 20 times and the result averaged. After a 20 ms delay, the device turns on the microphone and SD card, and then records 10 seconds of audio data, storing this to the SD card. Finally, the device returns to standby mode. The flowchart is illustrated in Figure 4.28.

Figure 4.28 The device is activated by motion detection; it then starts recording if the user touches the specific area of the capacitive sensor
4.5 Prototype – Second version

The first version has low power cost, motion detection, hand touch sensing and it uses an analogue microphone to record the sound, which is a typical and mature design, but it needs several components to make sure it works. This causes the first device too big to attach on an inhaler and cannot be implemented in the actual use. In order to reduce the size. The second version of the prototype optimised the PCB layout and replaced the previous analogue recording with a digital solution, which saves space for reducing the dimension of the hardware and improves the quality of the audio quality. Moreover, the newer prototype used the STM32L4 series which have lower power consumption and higher processing performance than the STM32L1 series. In terms of the firmware, the second prototype used the same processing flow to the first version.

4.5.1 Microcontroller unit

The STM32L4 series are ultra-low-power microcontrollers based on the ARM Cortex-M4 32-bit RISC core operating at a frequency of up to 80 MHZ. They feature a digital filter for sigma-delta modulators (DFSDM), converting an input stream to digital data at the hardware level. They also offer various communication interfaces for accessing peripheral components. As for optimisation, compared the previous design, the current one added an extra inductor consisting of a low-pass filter with a pair of capacitors C14 and C15; thus it could provide a more stable and clean voltage to the internal ADCs. The schematic is shown in Figure 4.29.
4.5.2 Digital microphone (MEMS microphone)

The MEMS microphone features small size, high reliability and good sound quality. It is fabricated with a capacitive sensing element and an IC interface, outputting digital data in pulse-density modulation (PDM) format. Figure 4.30 illustrates that the MEMS microphone needs fewer peripheral components for operating. The pin 4 DOUT and pin 3 CLK are directly connected to the DFSDM module on MCU.
4.5.3 PCB Layout

The second version was designed using a four layer PCB; the top and bottom are signal layers; the middle layers are ground plane and power plane, respectively. The top and bottom sides of the PCB layout are shown in Figure 4.31.

Compared to the analogue microphone, the digital microphone only required two decoupling capacitors; this significantly reduced the board size (see Figure 4.32).

Figure 4.31 Top layout (red) and bottom layout (blue)
The first and second prototypes have some limitations cannot be ignored. In the processing flow, the device wakes up by the motion of accelerometer and then uses the capacitive sensor to check whether the user holds the inhaler in their hand. If all the conditions are satisfied, the device will start the recording process. Ideally, it is a good mechanism to track patients’ use in a motionless condition, such as sitting on a chair. However, its false-trigger rate will rise up if users are in activities, such as walking and running. Moreover, continuously waking up the device will cause extra power costs. In another side, the current version cannot provide enough memory for advanced computation. Thus, the extra memory should be considered in the next design. At last, the two prototypes only used the motion sensor to trigger themselves but did not record its data, though the motion is one of the assessment factors for the most pMDIs. In order to make up these defects, the third prototype replaced the SD card connector by a shorter and higher connector, which increased the placement density of other components and further reduced its board size. Figure 4.33 shows how the memory connector saved space. Furthermore, this version employed a smaller MCU. It was 5 mm x 5 mm, and cheaper than previously used MCUs. An SPI Flash memory was added for high-speed reading and writing. Although one extra component was added, this version was still 15% smaller than the second version. In order to decrease power consumption and avoid false-triggering, the capacitive sensor was replaced by a mechanical button. For the software, the workflow also
changed. Once the was button pressed, the system started to record 25 seconds of audio and motion data, and then store the data into local memory.

![Figure 4.33 SD card was lift and more space can be used (yellow area)](image)

4.6.1 Microcontroller unit

In the final version, the microcontroller unit uses STM32L442KC, which belongs to the same series of MCU as in the second version but integrates fewer internal modules and thus has a smaller size. In Figure 4.34, the MCU offers fewer pins and they are fully connected; the digital microphone was connected to serial audio interfaces (SAI) instead of the DFSDM; the SWD connection was removed. Further, the firmware was programmed via a micro USB using device firmware upgrade (DFU) mode.

![Figure 4.34 Schematics of MCU for the final version](image)
4.6.2 Memory unit

Serial NOR flash memory was added in the final version; the memory chip has QUAD SPI and provides 16 Mbyte of high-speed buffer to temporarily store data. The circuit of the SD card connector and SPI NOR Flash are shown in Figure 4.35.

Figure 4.35 Schematics of the memory unit for the final version

4.6.3 PCB Layout

The dimensions of the final version were only 19 mm x 19mm; the PCB layers kept the same structure as the first and second versions. In Figure 4.36, the PMU components were placed on the left of both sides; the digital microphone was on the down-right corner of the top side; the accelerometer was on the centre of the board, and the SD card was mounted above it.

Figure 4.36 PCB of the final version; left is top side; right is the bottom side
4.6.4 **Hardware workflow**

In the first and second version, the device was woken by motion detection, and then a double check was performed by the capacitive sensor to decide whether to trigger recording. This is a good mechanism to track patients’ inhaler use from beginning to end in motionless conditions. However, if patients are in motion, such as walking and running, the device will be continually activated and turn on the capacitive sensor. In this case, power consumption will be high and false-triggering can easily happen. In addition, the previous working sequence operates sequentially. It is blocked by the completion of each step and then moves to the next step. However, a good technique needs to ensure that the inhaler is held vertical at all the time, and this is monitored.
by the accelerometer. Hence the accelerometer and microphone must operate simultaneously. In order to optimize operation, the device in the final version is activated by only a mechanical button. Compared to capacitive sensing, this method can significantly reduce power consumption and decrease the rate of false-triggering. The motion and acoustic sensor work in parallel. The workflow is displayed in Figure 4.38.

Figure 4.38 The final version is activated by pressing the button; the device working status indicated by a flashing LED.
4.7 Conclusion

The hardware is the key system responsible for performance measurement and adherence to medication. This chapter introduced the progression of the hardware design and summarised the technique used in the system, including the file system, SPI communication and capacitive sensing. In the initial stages, the combination of motion detection and touch sensing was used to trigger the system, but the system consumed too much power and also was prone to false triggering. The mechanical button was therefore introduced and the following experiments proved its reliability. In order to collect all useful information, audio recording and motion detection were conducted in parallel. To extend the battery life, high power-consumption components are managed by the power management unit, and components with low power consumption and low quiescent current were selected for the final design. The final version design satisfied the aforementioned requirements and showed reliable performance as described in later chapters. Furthermore, the final design also provides a good reference point for later and more sophisticated devices.
Chapter 5

Signal characterisation and recognition

5.1 Overview

This chapter introduces the characterisation and recognition methods for the recorded signals. First, the acoustic signals were analysed both in the temporal and spectral domains. The following sections describe the methods of audio feature extraction and audio augmentation. The system outputs data corresponding to the patient’s usage and technique, allowing the analysis algorithms to interpret and classify the waveforms and also to establish the quality of compliance. An HMM-GMM model and two deep learning model are discussed in the following section. The last two sections discuss the output response of the accelerometer by the software developed for this project. Finally, a nested RMS window method and Euler angle were used to detect the device status.

5.2 Acoustic signal characterisation

5.2.1 Audio analysis

The breath sounds and inhaler actuation sound were analysed using temporal and spectral features. In the time domain, the breath-out phase (exhalation) was characterised by a fast ascent and then slow descent (Figure 5.1 (a)). Compared to the exhalation, the inhalation, in Figure 5.1 (c), was a flat curve and had slow change. To describe these trends, the peak detection algorithm [141] was applied to extract envelopes of the waveform, which are shown in Figure 5.1 (b) and (d). The waveforms were then transformed into the frequency domain (see Figure 5.2) using the Fourier transform (FT).
Figure 5.1 (a) and (c) are waveform of exhalation and inhalation sounds; right two figures are waveform envelopes of exhalation (b) and inhalation (d).

Figure 5.2 Exhalation spectrum (a) and inhalation spectrum (b).

Intuitively, the inhalation and exhalation shared a similar spectral distribution, thus it is difficult to identify them. Furthermore, the breath sound is a non-stationary signal; its amplitude varies over time. However, a traditional FT analysis cannot provide a
time description of spectral components [142]. In this case, the short-time Fourier transform was used to analyse this signal. Figure 5.3 shows that the exhalation and inhalation have identifiable features in the time-frequency domain. To represent these properties of time-frequency, the Mel-frequency cepstral coefficients (MFCC) were employed to describe their features.

![Figure 5.3](image)

**Figure 5.3** The magnitude of the spectrogram (a) decreases over time, but the spectrogram (b) changes more slowly.

The device records the dosing by catching the actuated sound. A successful actuation of the inhaler is indispensable for the effective delivery of drugs. Moreover, dose statistics can help specialists understand the patient’s condition.

The sound of actuation is generated by the mixed aerosol-cloud passing through the inhaler nozzle [143]. As shown in Figure 5.4 (a), the waveform displays the recording of repeated firings of the inhaler. Each actuation lasts approximately 250 ms and the identified frequency is around 1658 Hz (see Figure 5.4 (b) & Figure 5.4 (c)). Acoustic features of the sound may be classified using machine learning models [55].
The temporary and spectral characteristic of the actuation sound

5.3 Acoustic signal recognition

5.3.1 Audio feature extraction

The MFCC method is a useful feature for analysis of audio signals (detailed in Chapter 3). MFCC feature extraction can be summarised as four steps: FT transformation, Mel-scale mapping, cepstral coefficients extracting and dynamic characteristics computing.

For the inhaler actuation detection, additional features (zero-crossing rate, spectral centroid, spectral spread and energy entropy) are extracted and are classified by forest (RF) and Support-Vector Machine (SVM) classifiers. Compared with the neural network models, the two machine learning methods have less computation cost and memory cost. They can, therefore, be deployed on low-cost MCUs to implement local detection.

5.3.2 Audio pre-processing

In the training step, the inhalation and exhalation audio clips are processed by a band-pass filter (BPF) to remove noisy spectral components. The filtered audio is then
divided into segments and labelled manually. Finally, augmented operations are implemented on the clips to generate the training dataset.

5.3.3 Audio augmentation

The audio augmentation is an approach to cope with small data sets used in deep network models and to improve recognition accuracy [144]. Figure 5.5 (b), (c) and (d) illustrate three augmentation operations - time-shifting, speed tuning and noise overlay. The data augmentation simulates the data in a real environment, in which features are distributed over a broader range of values. Appropriately combining these operations not only expands the sample size artificially but also improves the generalisation of the model.

![Audio augmentation methods](image)

*Figure 5.5 Audio augmentation methods*
5.3.4 Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM)

Historically, HMM-GMM was a popular method for automatic speech recognition (ASR). In acoustic signal recognition, an audio clip is split into several frames and then each frame generates an MFCC matrix. Assuming the probability densities of these features combine with Gaussian components, then the distribution can be estimated by a multivariate Gaussian mixed model and is given by

\[
p(x) = \sum_{k=1}^{K} \left( \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \right)
\]

\[
= \sum_{k=1}^{K} \left( \frac{\pi_k}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \right) \exp \left\{ -\frac{(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}{2} \right\} \cdot \pi_k > 0 \tag{5.1}
\]

where \(\pi\) is mixture weights, \(\Sigma\) represents the covariance matrix, and \(\mu\) is mean of \(X\).

In order to fit the given features, the parameters are estimated by the expectation-maximization (EM) algorithm. The trained GMM is considered as probabilities of the HMM latent states with respect to observed \(x\). The probability from latent states to observation \(x\) is called emission probability and expressed by

\[
b_j(x) = \sum_{k=1}^{K} \pi_{jk} \mathcal{N}(x|\mu_{jk}, \Sigma_{jk}) \tag{5.2}
\]

In Figure 5.6, the observed distributions of acoustic feature vectors are the emission probability \(b_j(x)\) and are also expressed as \(p(x|S_j)\). The symbol \(A_{(j-1)j}\) is used to describe the transition probability of the successive state. According to the Bayesian theorem, the posterior probability of the latent state \(S_j\) is given by equation 5.3

\[
p(S_j|x, \Theta) = \frac{p(X,S_j|\Theta)p(S_j|\Theta)}{p(X|\Theta)}
\]

\[
\Theta = \{A, \mu_j, \Sigma_j\} \tag{5.3}
\]
Equation 5.3 takes the form in equation 5.4 by using the conditional independence property [82].

\[
p(S_j | X, \theta) = \frac{\alpha_j(n) \beta_j(n)}{p(X | \theta)}
\]

\[
\alpha_j(n) = p(x_1, ..., x_n, S_j | \theta)
\]

\[
\beta_j(n) = p(x_{n+1}, ..., x_N | S_j, \theta)
\]  

In this thesis, the observation sequence X is the extracted MFCC feature vector of audio clips and is known. Therefore, the Baum-Welch algorithm is employed to estimate parameters and train the HMM-GMM model.

![Figure 5.6 Latent states with observed sequence](image)

### 5.3.5 CLDNNs

CLDNN (Convolutional, Long Short-Term Memory, Deep Neural Network) is a neural network architecture used for acoustic classification of short audio clips. This framework learns the frequency features from convolutional layers; long short-term (LSTM) layers can learn from the temporal sequence [145].

Long Short-Term Memory (LSTM) is an extended architecture of RNN that adds gate units to control information whether to transfer to the next layer. As shown at the top of Figure 5.7, the convolutional layers extract feature maps and output these to the LSTM layers. In order to solve the long-term dependencies problem, each LSTM cell
introduces gate units to manage the time span of information in the network chain [146]. The bottom graph of Figure 5.7 illustrates the structure of the LSTM cell. The concatenation of the input \( x_t \) and the previous output \( y_{t-1} \) are connected to the sigmoid activation layers \( \sigma \) and then multiplied by the other signal flows. The sigmoid function output is limited between zero and one. Therefore, the gate units can be implemented by multiplication of the output by the \( \sigma \) layer. The forget gate \( f_t \) decides which information will be discarded from state \( c_{t-1} \); the input gate \( i_t \) controls what information can be reserved in the state \( c’_t \). The current state \( c_t \) represents the useful parts of the information. The output \( y_t \) is obtained from the interpretation of \( c_t \) [147].

The last LSTM layer connects a global average pooling layer to reduce the output dimension. The following output passes through a fully connected layer (FC) to classify the feature maps obtained by the previous layer.

5.3.6 Residual networks

The residual network (ResNet) is one of the deeper neural networks, published in 2016. It reduced the difficulty of training in deeper networks and has lower classification errors in the typical image datasets[114]. In Figure 5.8, one of the architectures of 34-layers residual network (ResNet34) is displayed and is mainly stacked by convolutional layers. The residual block creates a shortcut to add output and input \( x \). This block provides a path to forward useful identity features into the next layers and increases the depth of the neural networks. Due to its good performance in image classification, it also improves the accuracy of audio clip recognition.

To implement ResNet34 for acoustic recognition, its architecture is slightly modified to fit 1-dimensional operation. As shown in Figure 5.8 (b), the input layer connects to a Convolutional layer with seven kernels and sixty-four filters. The input data are downsampled twice by this layer with a stride of 2 and the following Maxpooling layer. The curve line is a shortcut that directly connects the residual block (blue box); the dotted line has an additional operation of adjusting the dimension to fit the
downsampled output. The last residual block connects to a global average pooling layer and 1000 units of a fully connected layer (dense layer) with softmax, whose activation function is introduced in chapter 3. Finally, the decimal probability of each class is output.

5.3.7 Random forest and Support Vector Machine

The ensemble classifier RF and the discriminative classifier SVM were introduced in Chapter three. Compared to deep neural networks, the two machine learning models have a smaller amount of parameters and lower computation costs. The RF and SVM classification models were constructed using the scikit-learn library [88]. The training progress has two steps: parameter tuning and model training with cross-validation. First, the performance of the model corresponds to the hyperparameters chosen. A parameter search grid is created by those specified hyperparameters and a small training subset is used for model evaluation. This iterative search is called the grid search method. Then, hold-out cross-validation is applied in the training progress. The dataset is randomly split into two subsets of unequal size, and then the large subset is used for model training and the model is evaluated by the small one. The evaluation of models will be discussed in the next chapter.
Figure 5.7. Top is the architecture of CLDNNs; middle is a repeating module in LSTM; bottom is the typical structure of a LSTM cell.
Figure 5.8 (a) is one of the architectures of ResNet34; (b) is modified ResNet for 1-dimentional dataset

5.4 Motion signal characterisation

5.4.1 The serial data visualisation tool

The accelerometer sensor sends the motion signal to the MCU via the Serial Peripheral Interface (SPI) protocol. This protocol is widely used for low-speed devices, but on its own, it is difficult to visualize the data in real-time. To characterise the signal patterns of shaking and vertical holding, real-time plotting software using C++ was developed for visualisation of the accelerometer data (Figure 5.9). The graphical user interfaces
were programmed using the QT widget toolkit, which is very popular for cross-platform applications. This real-time plotting software uses the serial port to read CSV format data from terminals and its streamflow module is designed for the base interface of input and output. Namely, it provides a flexible communication framework to add other stream interfaces, such as TCP/IP and file streams, for future development.

![Real-time plotting software](image)

Figure 5.9 C++ software for motion signal characterisation in real-time

### 5.4.2 The coordinate system and outputs of the accelerometer

The device used a MEMS ADXL362 accelerometer to detect the inhaler motion. This sensor has a maximum output data rate (ODR) of 400 Hz, with a selectable
measurement range. It outputs the acceleration forces along with three axes. In the 3-axis coordinate system, the Z-axis is vertical and points up, the X-axis and Y-axis are horizontal and they are mutually perpendicular. The coordinate system of the accelerometer is illustrated in the left of Figure 5.10. In the normal (flat) orientation, the accelerometer outputs $X_{\text{out}} = 0 \, g$, $Y_{\text{out}} = 0 \, g$ and $Z_{\text{out}} = 1 \, g$. In the right of Figure 5.10 the accelerometer outputs $X_{\text{out}} = 0 \, g$, $Y_{\text{out}} = 1 \, g$ and $Z_{\text{out}} = 0 \, g$. The reason for the different output is that the hardware board is vertically attached to the inhaler, which causes the Y axis to be vertical to the ground.

![Diagram of accelerometer axes](image)

**Figure 5.10** The ADXL362 has three sensing axes $A_X$, $A_Y$ and $A_Z$ (left); the axes of acceleration sensitivity are flipped on the device (right)

### 5.4.3 Shaking motion analysis

The accelerometer measures the proper acceleration in a scalable range. If the ADXL362 is accelerating it outputs a positive value and vice versa for deceleration. Intuitively, the shaking behaviour, in the short term, can be considered as a periodic oscillation. To investigate the output response of the accelerometer, the device was held in the hand and shaken separately along the X-axis, Y-axis and Z-axis. The to-and-fro movement over time made the induced oscillatory acceleration, similar to a vibration signal, with a period of approximately 250 milliseconds. The waveforms are shown in Figure 5.11 to Figure 5.13. In a real situation, patients will not shake the inhaler along a single axis only, and there is often a rotation during the shaking
progress. Figure 5.14 depicts plots from shaking in normal usage conditions; the uncertain behaviour causes the acceleration components to change continuously.

Figure 5.11 The output response while the device was shaken along the X-axis
Figure 5.12 The output response while the device was shaken along the Y axis
Figure 5.13 The output response while the device was shaken along the Z-axis
Figure 5.14 The output response while inhaler was shaken as normal
5.5 Motion detection

5.5.1 Shake detection

In practical usage, acceleration changes in three-dimensional space, rather than simply along a single axis. Further, the acceleration changes rapidly no matter which direction it is. In order to describe this change, the moving root mean square (RMS) is calculated using

\[ f_{\text{rms}}(x) = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N}} \]  \hspace{1cm} (5.5)

where N is the length of the moving window, whose value governed by the sensitivity to changes in acceleration.

To combine the D axes of acceleration, the RMS vector \( X = \{ f_{\text{rms}}(x_1), f_{\text{rms}}(x_2), \ldots, f_{\text{rms}}(x_i) \} \) is summed, from which we subtract \( b_g \) to remove the effect of gravity on the vertical direction. It is therefore

\[ g_{\text{rms}}(X) = \text{sign} \left( \frac{\sum_{i=1}^{D} f_{\text{rms}}(x_i) - b_g}{D} - T_2 \right), \hspace{1cm} (5.6) \]

\[ D = 3, b_g = 1 \]

where the value of the threshold \( T_2 \) is twice a standard gravity (g).

The output of \( g_{\text{rms}} \) indicates whether the intensive shaking behaviour occurred. However, if the acceleration increases sharply in a short time, this detection will cause false triggering (see Figure 5.15). To avoid false detection, the size \( N \) of the moving window should be longer than the shaking period. In Figure 5.16, RMS uses a larger moving window \( (N = 200) \), which reduces sensitivity to jitter and provides a reliable indicator for shaking motions.
Figure 5.15 The upper waveforms are acceleration response of X-axis, Y-axis and Z-axis; the bottom step wave is $g_{\text{rms}}$ output ($N = 32$)
Figure 5.16 The upper waveforms are acceleration response of X-axis, Y-axis and Z-axis; the bottom step wave is $g_{\text{rms}}$ output ($N = 200$)
5.5.2 Orientation calculation

Not vertically holding the pMDI inhaler is one of the critical errors of inhaler techniques; such errors can significantly reduce the efficiency of the delivery of medication[30]. Therefore, it is necessary to calculate the orientation of the device. This is one of the features used to quantify adherence to medication.

The device orientation can be described by the rotation matrix $R$ corresponding to the real-world coordinate system [148]. As the previous section mentioned, the reference coordinate system is shown in the right of Figure 5.10, where the acceleration along the Y-axis outputs one unit of the standard gravity and the other axes are equal to zero. Therefore, the accelerometer output is

$$\begin{bmatrix} G_x \\ G_y \\ G_z \end{bmatrix} = R(\alpha, \beta, \gamma) \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} = R(\alpha, \beta, \gamma) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (5.7)$$

In order to express the rotation of the device in three-dimensional space, the Euler angles roll $\gamma$, pitch $\beta$ and yaw $\alpha$ are employed to represent the rotations about X-axis, Y-axis and Z-axis, respectively [149][150]. The matrices of rotation about the three axes are given by

$$R_x(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\gamma) & -\sin(\gamma) \\ 0 & \sin(\gamma) & \cos(\gamma) \end{pmatrix} \quad (5.8)$$

$$R_y(\beta) = \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix} \quad (5.9)$$

$$R_z(\alpha) = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5.10)$$

To rotate a 3D object, the object will be rotated sequentially in 2D space along the principal axes. If the pitch $\beta$ is rotated first, then roll $\gamma$ and yaw $\alpha$ is last. This operation is implemented by multiplying $R_y$, $R_x$ and $R_z$: 

118
\[
R_{zxy}(\alpha, \beta, \gamma) = R_z(\alpha)R_x(\gamma)R_y(\beta) = 
\begin{pmatrix}
R_{11} & R_{12} & R_{13} \\
R_{21} & R_{22} & R_{23} \\
R_{31} & R_{32} & R_{33}
\end{pmatrix}
\] (5.11)

According to Equation 5.8, Equation 5.12 can be rewritten by

\[
\frac{G_{xyz}}{||G_{xyz}||} = \frac{1}{\sqrt{G_x^2 + G_y^2 + G_z^2}} \begin{pmatrix} G_x \\ G_y \\ G_z \end{pmatrix} = R_{zxy}(\alpha, \beta, \gamma) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}
\]

\[
= \begin{pmatrix} R_{12} \\ R_{22} \\ R_{32} \end{pmatrix} = \begin{pmatrix} \cos(\gamma) \sin(\alpha) \\ \cos(\gamma) \cos(\alpha) \\ -\sin(\gamma) \end{pmatrix}
\] (5.12)

In this vertical detection of the inhaler, only the roll \( \gamma \) and yaw \( \alpha \) are necessary, so Equation 5.13 is employed to compute the two angles, which is given by

\[
\alpha = \arctan \left( \frac{G_x}{G_y} \right)
\] (5.13)

\[
\gamma = \arctan \left( \frac{-G_z}{\cos(\gamma)} \right) = \arctan \left( \frac{-G_z}{\sqrt{G_y^2 + G_x^2}} \right)
\] (5.14)

The right of Figure 5.10 shows the yaw angle has a range from \(-180^\circ\) to \(180^\circ\). However, the arctangent function in Equation 5.14 has a limited output in the range \(\pm 90^\circ\). To solve this problem, the arctangent is replaced by \(\text{arctanent2} \), which can calculate the sign of input arguments to determine the quadrant in which the angle lies. The yaw angle is calculated by

\[
\alpha = \text{arctan2}(G_x, G_y)
\] (5.15)

After the above computation, the vertical orientation of the device is identified by the roll \( \gamma \) and yaw \( \alpha \).
5.6 Conclusion

This chapter presented a detailed explanation of signals analysis and relevant identification methods. The discussion began with acoustic signal analysis, considering the characteristics in the temporal and spectral domains. The next section introduced the acoustic features that were used to train the various models. To improve the performance of the trained model, data augmentation was applied to the sound. In terms of classification, HMM-GMM, CLDNNs and ResNet were discussed and used to classify breath sound events. The lower computation-cost model RF and SVM were implemented to detect the actuated inhaler event. In the motion signal section, software-based on QT was developed to analyse the motion signal. Finally was presented the development of a robust nested window method to detect inhaler shaking, and a method based on Euler geometry to identify the inhaler tilting.
Chapter 6

Experiments

6.1 Overview

Following the analysis and recognition of the signals, the first part of this chapter describes a series of experiments conducted to evaluate the system. The first experiment presents a robust mechanism to trigger the device. In the second experiment, the motion detection algorithms are evaluated under different activities. The following experiment tests four breath recognition models with the spectrogram and MFCC features. The performance of actuation detection is also discussed in this experiment.

As aforementioned in the literature review, healthcare professionals do not have an efficient and objective method to acquire their patients’ situation of treatment [6][151]. Poor inhaler techniques cause negative impacts, such as low adherence to medication, high risk of disease exacerbation, poor life quality and heavy economic burdens [6][7][10]. Therefore, the second part of this chapter describes the fourth experiment to track patients’ inhaler use and technique in real life. The feedback from this experiment was also used to explore the deviation between self- and electronic report.

6.2 Experiment 1

6.2.1 Overview

Appropriate triggering of the device is the premise of the success for the beginning of the measurement. This experiment aimed to validate and optimise the inhaler triggering mechanism. It was carried out in two phases: phase one employed the
second version of the device to track volunteers’ inhaler use. The second phase monitored volunteers by the optimised version, described below.

6.2.2 Methodology

Design
The phase one experiment tested the ideal design of the triggering mechanism in practical use. First, the volunteers were educated on how to use the inhaler and then divided into two groups A, B. The group A volunteers used the inhaler while they were sitting down in front of a desk. Group B was asked to use the inhaler while walking. Following the first stage, the second phase had the same content of the test, but participants were tracked by the optimised device.

Participants
A total of 6 candidates with ages between 25-35 participated in the experiment. Four candidates took part in the first phase and the others were assigned into the next split.

Test procedure
At first, participants were required to use the inhaler at least 5 times within a half-day and fill the record form (Table 6-1). The result of the first phase was used to optimise the hardware and improve the triggering mechanism. The new version of the device was employed in the second phase. This involved a slightly different test: two candidates did the same test at the same time.

Table 6-1 Experiment 1: inhaler triggered record form

<table>
<thead>
<tr>
<th>No.</th>
<th>Group:</th>
<th>Date:</th>
<th>When do you use it? (HH:MM:ss)</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.2.3 Analysis

Table 6-2 summarises the recorded results for the whole experiment. The phase one candidates used inhaler 62 times in total, but the devices stored 65 samples and only 58 records were triggered correctly. Compared with the sitting protocol, the continuously moving protocol caused a higher false-trigger rate (according to the detection algorithm of the second version device, introduced in Chapter 4). The device normally stays in standby mode until the inhaler is moved. The capacitive detection runs in a blocked loop and is stopped by a hand touch or timeout. This triggering method was used to minimize the device operation. This method worked well in group A; the standby inhaler was activated appropriately when a volunteer picked up the inhaler on the desk. However, the device showed poor performance if the candidates were walking. The device triggering mechanism employed the nested detection method. It was first triggered by motion and then by touch. Therefore, volunteers walking repeatedly waked up the device and caused it to start recording at the wrong time. Moreover, frequently triggering the device increased the power consumption.

To improve triggering accuracy, as described in chapter 4, a mechanical button was used to replace the electronic sensors. Once the button was pressed, a red flashing LED indicated the system was ready for the inhaler use. Although the new version added one more step before inhaler use, its operation was simple and clear. Phase two used the optimised version to track participants e and f. As the results showed, the new device would not be limited by a person’s activity. It succeeded in tracking all 35 inhaler usages, both sitting and walking.

<table>
<thead>
<tr>
<th>Group</th>
<th>Volunteers</th>
<th>Use times</th>
<th>Recorded samples</th>
<th>Correct recordings</th>
<th>Phases</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>One</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>One</td>
<td>Sitting</td>
</tr>
<tr>
<td>B</td>
<td>c</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td>One</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td>15</td>
<td>17</td>
<td>13</td>
<td>One</td>
<td>Walking</td>
</tr>
<tr>
<td>-</td>
<td>e</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>Two</td>
<td>Sitting</td>
</tr>
</tbody>
</table>
## 6.2.4 Conclusion

The experiment validated and analysed the two triggering methods to find an efficient and reliable way to activate the device. The findings demonstrated that the first mechanism, which featured by no-operation, was often falsely triggered via participants’ activities. Furthermore, its frequent activation reduced battery life. In the second phase, all participants’ inhaler use was tracked by the optimised triggering mechanism. The final device not only replaced the electronic trigger with a mechanical button, it also included an LED indicating device status. This LED feedback informed the user when the system had been activated. In short, the mechanical button reduced false system wake-up and an LED light increased the rate of successful activation.

### 6.3 Experiment 2

#### 6.3.1 Overview

Shaking the inhaler well before inhaling and holding the inhaler vertical during inhalation are critical steps corresponding to the delivery efficiency of the drugs. This experiment includes shake and hold device tests in different activities, such as sitting, walking and climbing stairs. It aimed to validate the stability and performance of motion detection algorithms.

#### 6.3.2 Methodology

**Design**

This experiment tested the reliability of motion algorithms. Each accelerometer data record was 20 seconds collected at 195 Hz sample rate (the final version of the device). Two groups of datasets, named ‘Shake’ and ‘Hold’, were used to test shake detection.
and tilt detection, respectively. The inhaler was used in various environments, yielding a total of 377 recordings (around 126 minutes), with participants performing four activities: sitting, walking, ascending and descending stairs.

**Test procedure**

The device worked in the ‘acc’ log mode that stored only accelerometer data into comma-separated values (CSV) file. Each file represented an activity and contained 20 seconds of time data in three axes. First, the device was placed flat on the table in order to calibrate three axes of the accelerometer; jitter noise was removed by a moving average filter (see Figure 6.1).

![Figure 6.1 Pre-processing the accelerometer outputs](image)

Following the calibration, a few samples were collected by holding the device upright to create the baseline. As chapter 5 describes, the coordination system was defined when the device was held vertically. Namely, in this reference, the Y-axis is vertical to the ground and the other axes are parallel to the earth plane. For the shake test, the 57 positive samples were collected by shaking the device during walking, sitting, going downstairs and going upstairs. The comparison samples were collected for the same
activities but without shaking. To simulate different conditions, there were a total of 133 negative samples collected with the device kept in the participant’s pocket. Collecting ‘Hold’ samples was performed for the same activities of the ‘Shake’ group. Then 62 positive samples were collected by holding the device upright with a slight tilt. The 125 negatives were collected by the device tilted greater than ±45° under the same conditions.

6.3.3 Analysis

Shake detection

The accelerometer samples from this experiment are shown in Figure 6.2. The last two rows show the shake waveform while the participant was sitting, walking, going downstairs and climbing stairs. The others were performing the same activities excluding shake motions. The accelerometer responses of normal activities ranged within ±2.5g, which shows a significant difference from the shake samples. The shake detection algorithm calculated an RMS subframe of acceleration for every 240 data points (around 1.2 seconds) with 50% overlap; the final discriminator outputs depended on whether there existed an RMS subframe exceeding the threshold. Namely, this detection algorithm discriminated whether an intense motion continued for more than 1.2 seconds; its outputs are illustrated in Figure 6.3. The test result shows this method is reliable and robust in these test samples (see Table 6-3).

Table 6-3 The test result of the shake detection algorithm

<table>
<thead>
<tr>
<th>Activity</th>
<th>Actual shake event (total samples)</th>
<th>Predicted shake event</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting</td>
<td>20(40)</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Walking</td>
<td>20(82)</td>
<td>20</td>
<td>100%</td>
</tr>
<tr>
<td>Upstairs</td>
<td>5(15)</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>Downstairs</td>
<td>12(53)</td>
<td>12</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>57(190)</td>
<td>57</td>
<td>100%</td>
</tr>
</tbody>
</table>
Figure 6.2 The accelerometer samples with the different activities.

Figure 6.3 The left plots show accelerometer signals with shake motions and the right are the outputs of the shake detection algorithm.
**Tilt detection**

These test records are from “Hold” group dataset, which contained positive and negative samples. The positive were collected from vertical holding during sitting, walking, upstairs and downstairs motion. The negative samples were recorded for the same activities, but the device was tilting or was upside down. Compared with the shake signal, the rotation angle does not change quickly while the user is using the inhaler. Therefore, the original acceleration data was smoothed by the moving average filter with a window of 195 points, which is equivalent to reducing the sampling rate of the rotation to 1 Hz. Figure 6.4 illustrates the device was held upright or upside down during walking or downstairs motion. This smoothing method minimises the jitter caused by activities. As the previous chapter discussed, one-unit of standard gravity (1g) is sensed by the Y-axis if the device is held upright. If the device was flipped upside down, the Y-axis output is reversed to negative with the same magnitude. In this experiment, the test device, for every recording, was kept at the same tilt angle for 20 seconds, which was the reason why the yaw and roll angles include a small fluctuation in the right of Figure 6.4.

As aforementioned, the equation of Yaw expresses the range of rotation angle from $-180^\circ$ to $180^\circ$. However, if the device was rotating clockwise or counter-clockwise over $180^\circ$ or $-180^\circ$, the sign of yaw angle will be reversed with the same magnitude. It causes the calculation of the yaw angle to become unstable and quickly flips over a short period of time (see Figure 6.5). In order to avoid this situation, the absolute value of Yaw was used in the orientation calculation. In Figure 6.6, the distribution of calculated orientation presents a clearly discriminated gap between positive and negative samples. The algorithm was used to detect the device tilting under various conditions, and the test results are shown in Table 6-4.
Figure 6.4 The acceleration signals (left) were averaged by the moving average method with a window of 195 values; The yaw and roll angles were used to describe the holding orientation (right).

Figure 6.5 The sign of the yaw angle is reversed if the rotation of the yaw extends over ±180°.
Figure 6.6 The device is considered as vertical if its yaw and roll angles are within \( \pm 35° \) (red area)

Table 6-4 The test result of the tilt algorithm

<table>
<thead>
<tr>
<th>Activity</th>
<th>Actual holding inhaler upright (total samples)</th>
<th>Predicted result</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting</td>
<td>26(66)</td>
<td>26</td>
<td>100%</td>
</tr>
<tr>
<td>Walking</td>
<td>23(91)</td>
<td>23</td>
<td>100%</td>
</tr>
<tr>
<td>Upstairs</td>
<td>7(10)</td>
<td>7</td>
<td>100%</td>
</tr>
<tr>
<td>Downstairs</td>
<td>6(20)</td>
<td>6</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>62(187)</td>
<td>62</td>
<td>100%</td>
</tr>
</tbody>
</table>

6.3.4 Conclusion

This experiment simulated the inhaler use under different activities, and then tested the reliability of the algorithms. The shake detection calculated the RMS of the combined axes every 1.2 seconds with 50% overlap, which successfully detected all shake motions and was not disturbed by other daily activities. The tilt detection used yaw and roll to measure the device tilt. In order to avoid the reversal of yaw, the yaw was given by its absolute value. According to the experimental results, both the shake
detection and tilt detection algorithm performed consistently and reliably during normal usage.

6.4 Experiment 3

6.4.1 Overview

The description of the third experiment is divided into two parts. The first part considers the performance of breath recognition models. It was a significant experiment, undertaken to explore the discriminative ability of the models applied to the same acoustic dataset. The actuation classification models are evaluated in the second part. A total of 130 test samples were used to estimate the performance of the two classifiers.

6.4.2 Methodology

Design

The experiment was designed to measure the performance of the classification models and show the comparison of the performances of models. The dataset contains breath samples collected from the three participants and the background sounds provided by Urban Sound and FSDnoisy18k datasets[1][2]. The breath sounds were split into two parts: training and test. During the training stage, a training breath sample was overlaid on a randomly selected background sound and the amplitude of the generated audio clip was reduced to -20 dBFS. In terms of model testing, the test dataset used the same method to estimate the performance of the models. The workflow for the training and testing phases is illustrated in Figure 6.7. Following the evaluation of breath recognition models, the inhaler actuation classifiers, random forest (RF) and support vector machine (SVM), were evaluated in an indoor environment.
Test procedure

This test aimed to evaluate all of the classification models discussed in the last chapter. For breath recognition models, the test dataset was generated at different levels of Signal-to-noise ratios (SNRs) to determine the performances of these models under various noisy situations. One of the generated samples is displayed in Figure 6.8. Three metrics (precision, recall and F1 score), are used to evaluate and compare the performance of the four models.

In terms of the actuation test, the parameter tuning of the random forest and SVM method are both processed by the cross-validated grid-search method [88], which search the appropriate parameters for the classifiers to maximise their performance.
Then, classifiers were tested on a total of 130 samples, which contained 90 actuated sounds and 40 irrelevant noise cases.

![Waveform plots for SNR levels of 5dB, 15dB, 25dB, and 35dB]

Figure 6.8 Overlapping a clean breath sound and a noisy sound at SNR = 5dB, 15dB, 25dB and 35dB

6.4.3 Analysis

This experiment used two types of features to train models, and the four models were used with the same dataset. The test result is summarised at Table 6-5. As a typical sequential model, HMM-GMM, trained by spectrogram features, had 90.1% accuracy at an SNR of 30 dB, but the weaker performance at SNRs of 15 dB and 0 dB, in which accuracy dropped to 79.4% and 59.6%, respectively. The deep learning models (ResNet18 and CLDNNs) showed better performance. The average accuracy of the single ResNet18 reached 90.4% and CLDNNs obtained an average accuracy of 91.2%. As chapter 5 discussed, the convolutional layers mainly extracted the spectral features, but the LSTM layers are good for the representation of temporal features. Therefore, the last model ‘ResNet18+CLDNNs’ concatenated feature mappings of ResNet18 and CLDNNs before the fully connected layers. However, this combination did not improve the performance of recognition and had an even lower average accuracy (91.1%) than CLDNNs.
In terms of MFCC features, the combination of ResNet18 and CLDNNs had the highest mean accuracy of 94.5%, and it even achieved an accuracy of 89% in noisy environments (SNR=0dB). CLDNNs and ResNet18 were ranked the second (93.5%) and third (91.5%) in average accuracy. Compared to HMM-GMM with the spectrogram feature, the HMM-GMM method returned a better performance with the MFCC feature. Its average accuracy increased to 88.8%. Furthermore, HMM-GMM had the highest accuracy of 98.9% for pure signals (SNR=30 dB), but it did not work well in noisy environments.

A classifier comparison between the RF and SVM is illustrated in Table 6-6. The two classifiers with optimised hyperparameters both achieved an accuracy above 95% for actuation identification. The SVM has higher precision than RF, indicating that the SVM has a strong discriminative ability for the actual positive samples. However, the recall of RF is 1.1% higher than that of SVM. In this case, the performance of RF and SVM has its own merits. Therefore, in order to balance the measure between precision and recall, the F1 score is introduced to evaluate the models. According to the F1 score, RF shows worse performance than SVM.

<table>
<thead>
<tr>
<th>Method</th>
<th>SNR = 0 Accuracy (%)</th>
<th>SNR = 15 Accuracy (%)</th>
<th>SNR = 30 Accuracy (%)</th>
<th>Mean accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spectrogram</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMM-GMM</td>
<td>65.9%</td>
<td>83.9%</td>
<td>84.9%</td>
<td>78.2%</td>
</tr>
<tr>
<td>ResNet18</td>
<td>85%</td>
<td>93%</td>
<td>93.1%</td>
<td>90.4%</td>
</tr>
<tr>
<td>CLDNNs</td>
<td>88%</td>
<td>92.9%</td>
<td>92.9%</td>
<td>91.2%</td>
</tr>
<tr>
<td>ResNet18 + CLDNNs</td>
<td>87.7%</td>
<td>92.9%</td>
<td>92.9%</td>
<td>91.1%</td>
</tr>
<tr>
<td><strong>MFCCs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMM-GMM</td>
<td>72.1%</td>
<td>95.4%</td>
<td>98.9%</td>
<td>88.8%</td>
</tr>
<tr>
<td>ResNet18</td>
<td>80.9%</td>
<td>96.1%</td>
<td>97.6%</td>
<td>91.5%</td>
</tr>
<tr>
<td>CLDNNs</td>
<td>87.7%</td>
<td>95.9%</td>
<td>96.7%</td>
<td>93.5%</td>
</tr>
<tr>
<td>ResNet18 + CLDNNs</td>
<td>89%</td>
<td>96.7%</td>
<td>97.7%</td>
<td>94.5%</td>
</tr>
</tbody>
</table>
### Table 6-6 Performance of the classifiers – actuation identification

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F1 (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>100%</td>
<td>97.8%</td>
<td>98.9%</td>
<td>98.5%</td>
</tr>
<tr>
<td>RF</td>
<td>96.7%</td>
<td>98.9%</td>
<td>97.8%</td>
<td>96.9%</td>
</tr>
</tbody>
</table>

#### 6.4.4 Conclusion

The acoustic classification model is a significant part of the whole system. This experiment evaluated the performance of different models trained by spectrograms and MFCCs, and it also explored the performance of the combined model. According to the comparison of the breath models, MFCCs provided more discriminative features than spectrograms of these audios. The MFCCs tests confirmed that the HMM-GMM had a good performance on the test dataset, but its accuracy dropped quickly as the SNR degraded. The ‘ResNet18+CLDNNs’ had the best performance in this experiment. This combined model not only worked well with pure signals but also had an acceptable accuracy in the noisy environment. As for the evaluation of actuation model, the RF and SVM classifiers both presented high accuracy in this test, but SVM showed higher F1 score than RF’s.

#### 6.5 Experiment 4

#### 6.5.1 Overview

Careful use of the pMDI is required to maximise the efficiency of drug delivery[30], so identifying critical errors in inhaler technique is a practical method to measure the adherence to the medication. The final experiment aims to investigate the difference between patients’ self-reports and electronic reports. The final version of the device was employed, over a three week period, to track the volunteers who had clinician-diagnosed asthma or COPD.
6.5.2 Methodology

Engagement process

From April 2017 to June 2018, a series of stakeholder meetings hosted by the British Lung Foundation North West Development team were used to understand real patients’ attitude regarding electronic devices monitoring their inhaler use. These meeting provided useful information and practical suggestions to optimise the design and functionality of the device.

Design

In this experiment, participants used electronic devices to assess their inhaler use and technique during daily life, for three weeks. At the same time, participants were asked to record their daily logs. The daily template is shown in Table 6-7. After three weeks, the records were collected and compared to the paper and electronic records. A typical waveform of an electronic recording is given in Figure 6.9.

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Shook the inhaler</th>
<th>Where did you use?</th>
<th>How good was your technique?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: 08/07 14:10</td>
<td>Yes No Don’t remember Indoor Outdoor</td>
<td>Very Poor Ok Good Very good</td>
<td></td>
</tr>
</tbody>
</table>
| ![Table 6-7 Daily log template](image)

...
Participants
Six participants with clinically diagnosed asthma or COPD for more than five years duration were drawn from the four advisory/support groups with ages greater than 55. Five participants used pMDI inhalers and one participant used a pMDI with a ‘spacer’.

Test procedure
The experiment was intended to monitor the participants for three weeks, from June 2018 to August 2018. Before the experiment, participants were instructed how to operate the device and were asked to complete the daily log as required. At the end of the experiment, all data were collected for further analysis.

Assessment standard
In this experiment, participants were asked to assess their pMDI technique using a five-level score in the paper records (see Table 6-7). The electronic records used the common assessment standard (Table 6-8) to assess participants’ inhaler technique. A
A formula was developed to obtain an independent score of inhaler technique; this used a sum of weighted factors, given by

\[ y_{\text{score}} = S \left( e_f \sum_{n=1}^{N-1} \omega_n e_n \right) \]

where \( y_{\text{score}} \) is the weighted score of the inhaler technique and its output score corresponds to the level of self-assessment (Very poor, Poor, OK, Good and Very Good).

Equation (6.1) outputs the weighted score of the inhaler technique and its output score corresponds to the level of self-assessment (Very poor, Poor, OK, Good and Very Good).

<table>
<thead>
<tr>
<th>STEPS</th>
<th>CHECKLIST</th>
<th>SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pMDI</td>
<td>pMDI with spacer</td>
</tr>
<tr>
<td>(1)</td>
<td>- Shake inhaler</td>
<td>- Shake inhaler</td>
</tr>
<tr>
<td></td>
<td>- Hold inhaler upright</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>- Exhalation (empty lungs)</td>
<td>- Exhalation (empty lungs)</td>
</tr>
<tr>
<td>(3)</td>
<td>- Actuation ( t_a ) (-0.5 s \leq a \leq 1.25 s)</td>
<td>- Inhalation ( t_{in} ) (( t_{in} \geq 2 s ))</td>
</tr>
<tr>
<td></td>
<td>- Inhalation ( t_{in} ) (( t_{in} \geq 2 s ))</td>
<td>- Breath times ( \geq 5 s ) (Tidal breathing)</td>
</tr>
<tr>
<td>(4)</td>
<td>- Hold breath ( t_h ) (( t_h \geq 10 s ))</td>
<td>- Hold breath ( t_h ) (( t_h \geq 10 s ))</td>
</tr>
<tr>
<td></td>
<td>- No cough</td>
<td>- No cough</td>
</tr>
<tr>
<td>(5)</td>
<td>- Actuation success</td>
<td>- Actuation success</td>
</tr>
</tbody>
</table>
6.5.3 **Analysis**

A total of 233 readings collected from the six participants over the three weeks. However, some participants occasionally forgot to complete the daily log, and one participant did not fill the log as the experiment required. Furthermore, participant A’s device battery drained quickly, which caused the device to only record the first 15 inhaler uses. In these cases, only 192 valid readings were used to compare with the electronic assessments (see Figure 6.10). A summary of the participants’ self-reports is displayed in Table 6-9. The device usually monitored and assessed all inhaler use, but all of the participants only assessed their first dose though they took two or more doses at once. To make the two record resources comparable, the original electronic assessments were replaced by the average score of its adjacent assessments.

According to the paper reports, participants took an average of 2.3 doses per day (max: 4 doses/day, min: 0 doses/day) and 100% shook their inhaler before inhaled drugs. In Figure 6.10 (a), in 136 of (70.8%) instances, participants assessed their techniques as ‘OK’ and thought 47 (24.5%) of instances were ‘Good’. However, electronic records showed different results. According to Figure 6.10 (b), ‘Poor’ use was indicated 67 times (34.9%), whilst the instances of ‘OK’ and ‘Good’ were also lower at 71 (37%) and 25 (13%), respectively. Furthermore, the cases of ‘Very poor’ use increased from 1 (0.5%) to 29 (15.1%).

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Disease</th>
<th>Age</th>
<th>Days</th>
<th>Device</th>
<th>Doses</th>
<th>Shaking (proportion)</th>
<th>Top self-assessment (proportion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Male</td>
<td>COPD</td>
<td>≥ 55</td>
<td>36</td>
<td>pMDI</td>
<td>19</td>
<td>Yes (100%)</td>
<td>Good (73.3%)</td>
</tr>
<tr>
<td>B</td>
<td>Female</td>
<td>COPD</td>
<td></td>
<td>22</td>
<td>pMDI</td>
<td>43</td>
<td>Yes (100%)</td>
<td>Good (64.3%)</td>
</tr>
<tr>
<td>C</td>
<td>Female</td>
<td>Asthma</td>
<td></td>
<td>25</td>
<td>pMDI</td>
<td>95</td>
<td>Yes (100%)</td>
<td>Ok (66.7%)</td>
</tr>
<tr>
<td>D</td>
<td>Female</td>
<td>COPD</td>
<td></td>
<td>25</td>
<td>pMDI</td>
<td>42</td>
<td>Yes (100%)</td>
<td>Ok (100%)</td>
</tr>
<tr>
<td>E</td>
<td>Male</td>
<td>COPD</td>
<td></td>
<td>21</td>
<td>pMDI</td>
<td>45</td>
<td>Yes (100%)</td>
<td>Ok (100%)</td>
</tr>
<tr>
<td>F</td>
<td>Female</td>
<td>COPD</td>
<td></td>
<td>23</td>
<td>pMDI with spacer</td>
<td>79</td>
<td>Yes (100%)</td>
<td>NA</td>
</tr>
</tbody>
</table>
Figure 6.10 (a) shows the participants’ self-assessments and (b) shows the electronic assessments.

Figure 6.11 summarizes the statistics regarding ‘critical errors’. Chart (a) illustrated statistics of participants’ inhalation time length. As the previous chapter discussed, slow-inhaling is beneficial since it maximises lung deposition [5][8], but, in this step, only participant E inhaled the drug appropriately. The actuation timing represents the time interval between the actuated inhaler and the start time of inhalation; the statistics are shown in Figure 6.11 (b). A study stated that pressing the inhaler canister around the beginning of the inhalation can be considered as good timing and even slightly delayed triggering increases the medical efficacy [154]. Thus, according to this study, a delayed actuation time within -0.5s to 1.25s is considered as a correct timing in this experiment. The timing assessment did not include participant F, who used pMDIs with spacers, because the actuation time had little impact on the inhaler with this modification. Participants A through E chose the correct time to actuate their inhalers and the average correct percentage reached 80.4%. The step of breath-holding is also a significant factor affecting drug efficacy [18]. According to Figure 6.11 (c), the length of time for holding breath varied in a wide range for half of the participants, and only participant F held their breath close to the recommended 10 seconds. Figure 6.11 (d) shows a significant difference between the individuals, which means participants’ performance varied widely for this step. In this experiment, one participant coughed several times when inhaling the drug. Coughing causes the
participant to hold their breath improperly and may force them to cough up the inhaled drug. However, many studies do not consider it as a factor affecting the efficiency of drug delivery. The bar chart at the bottom right of Figure 6.11 demonstrates the proportion of successful actuation. Four of the six participants sometimes failed in this step, but most of the instances were performed very well.

Figure 6.12 summarises all the participants’ inhaler technique assessments. Individuals showed good technique both in holding upright and shaking steps, with correct percentages over 90%. Actuation success and actuated properly were ranked third (89.4%) and fourth (80.4%) in the correct technique, respectively. Breath-holding shows the highest technique error in this experiment. Only 5.9% of cases held the breath for over 10 seconds. The other major sources of error included, in order, complete inhalation (12% correct) and exhalation (50.3% correct).

Figure 6.11 (a) Length of inhalation time; (b) actuation timing; (c) length of breath-holding time; (d) percentage of exhalation before inhalation; (e) percentage of cough in holding breath step; (f) percentage of successful actuation
Figure 6.12 The summary of average correct proportion of the recommended steps

6.5.4 Conclusion

The last experiment monitored six participants by inhaler usage and technique over three weeks, yielding significant data on pMDI use. The analysis of self- and electronic reports showed that participants overestimated their inhaler technique, even though all of them were experienced pMDI users. Further, participants also made mistakes with recording their self-reports. The electronic device provided an objective method to track and quantify engaged patients’ adherence to pMDI techniques. It can also assist healthcare professionals to understand a patient’s situation and increase the efficiency of the treatment. The device may also provide a digital platform for related professionals and researchers to enhance the volume of information from clinical trials.
Chapter 7
Discussion

Chronic respiratory diseases are the third leading cause of death worldwide [2], and increasing attention has been paid to improve the quality of long-term patient treatment in this area. Currently, inhaled therapy is the most efficient and effective method to deliver drugs to the site of the pathology [19][156]. As one of the most popular inhalers [18], pMDI requires a careful technique to maximise the efficacy of the medication. However, several studies indicate that technique errors exist widely among patients [23][157][158]. The ineffective communication from patient to physician, such as a patient’s self-report and healthcare professional consultation [151], may lead to a professional misunderstand a patient’s actual condition, even if they are aware of the seriousness of these errors.

To cope with this barrier, acoustic monitoring is an objective and efficient solution. In 2012, Holmes et al.’s study employed an Inhaler Compliance Assessment (INCA) device to record the sounds of ‘Diskus’ inhaler use and developed an automatic algorithm to identify participants’ inhalation [62]. Taylor and Holmes et al.’s study, in 2016, used the digital microphone to record audios and then identify the pMDI actuation [32]. Their following study used quadratic discriminant analysis (QDA) to detect the sound (inhalation, exhalation and actuation) recorded by INCA and achieved an accuracy of 88.2% [13]. As discussed in the literature review, the correct pMDI technique has eight steps and can be considered as two sensing types. The first type (acoustic events) included the breath sound and the actuation sound. Shaking the inhaler and holding the inhaler upright belonged to the second type (motion events). However, all of these prior studies focused on the first type detection and ignored the second one. Furthermore, slowly inhaling drugs and then holding the breath for more than 10 seconds are also significant steps to maximise the lung deposition [14][18][30], but these details were not discussed in these studies. The pMDI is the most cost-efficient...
product on the inhaler market, but its effects of medication depend largely on a patient’s pMDI technique. Thus, it is necessary to develop a specific device to monitor and assess patients’ pMDI use and technique.

Before any experiments were conducted, one evaluation system and three versions of the hardware prototypes were developed. During the initial stage, the basic system structure was based on an analysis of the recommended steps of the pMDI. Its feasibility was evaluated using the evaluation system. Following this system evaluation, the first prototype was designed, with a size of 32mm x 20mm. The board had an analogue microphone and an accelerometer to record audio and motion data, respectively. All of the power of the peripheral components were controlled by a load switch. Because of the high density of the first PCB, the analogue microphone of the first version device was subject to interference from the peripheral digital components. To avoid this interference, a digital MEMs microphone was employed to replace the analogue microphone in the second prototype. Furthermore, the digital MEMs microphone reduced the component count, hence the device size was reduced to 22 mm x 20.5 mm. The third prototype (final version) was developed after the first experiment. In this version, the touch sensor was discontinued and a mechanical button replaced the old trigger method to activate the device. To further increase the utilisation of the space, the PCB layout was adjusted and the size (19.5 mm x 19.5 mm) was approximately 15% smaller than the previous version. Moreover, an SPI flash chip was squeezed onto the board to provide extra memory storage. Although the final version of the device gave a reliable performance in the experiments, there are some limitations that need to be considered. In real-world, the mono microphone acquires the signal contaminated with environmental noise. The noise may be of two forms: stationary noise and non-stationary noise. The stationary noise does not change with respect to time, such as white noise. This type of noise can be filtered by traditional digital signal processing (DSP) methods. However, it has poor performance on non-stationary noise, whose statistics varying over time. Examples include chirp noise, traffic noise, human conversation and so forth. In such cases, suppression is often based on adaptive filtering using a dual-microphone design [159][160]. With the growing popularity of deep neural networks, non-stationary noise can also be
suppressed using a single recording source [161]. These studies suggest a useful
direction for further research, for the purposes of noise minimisation and audio signal
enhancement.

Prior to the acquisition of the acoustic and motion signal, appropriately triggering the
device is the most important step. Experiment 1 compared and discussed the two
trigger mechanisms. The first mechanism aimed to minimise the user’s operation,
using the accelerometer to detect inhaler motion. When the accelerometer is
triggered, a capacitive sensor starts to detect the inhaler and whether it is in the user’s
hand. This method triggers the device automatically; patients do not need to know
how to operate the device. However, the result of experiment 1 showed this
mechanism had a high false-trigger rate under normal usage, especially when
associated with activities such as walking, jogging and climbing stairs. Moreover, the
activities triggered the device frequently, causing increased power consumption. To
solve these problems, a mechanical button was used in a revised design to trigger the
system. A wake-up interrupt was generated when the button was pressed, and then
the device flashed red to indicate activation. All components of the trigger module
were passive, which very significantly reduced power consumptions for the trigger
step. According to the results from experiment 1, the second mechanism had high
accuracy for triggering the device appropriately and had good reliability under daily
usage. Furthermore, users found that it was easy and intuitive to operate. Despite the
robustness of the mechanical switch activation, it did have some limitations. Ideally,
the button is pressed before patients start to use the inhaler. However, the device
cannot record the whole progress of the inhaler use if patients activate the device late.
An example of late activation would be a patient triggering the device after they shook
the device and emptied their lungs of air. In this situation, the device would not
capture the inhaler use from the beginning to the end, so the patient’s inhaler
technique would be recorded as inadequate by the electronic assessment. However,
in relation to the required coordination for pMDI use, this assessment gap is easily and
effectively minimised by adequate training, because of the simple operation of the
device.
When the device is activated, its power management unit switches on the accelerometer and microphone to record data. As aforementioned, there are two types of data produced by the system: from the accelerometer and microphone. The accelerometer generates the acceleration signal, giving the current motion state of the device, and the microphone is responsible for recording the acoustic signal. Experiment 2 explored the stability of the motion detection algorithms; the recognition models for the acoustic signals were discussed in the following experiment. In experiment 2, shake detection and tilt detection were both tested in a similar environment under four activities, including sitting, walking, walking up and downstairs. In the second experiment, the shake detection algorithm showed very high discriminative ability between shaking and unrelated daily activities. The tilt detection algorithm used a Euler angle system to represent the inhaler rotating status, and returned robust performance to identify the rotation of the inhaler. As chapter 5 discussed, the rotation sequence of tilt detection is pitch $\beta$, roll $\gamma$ and yaw $\alpha$. If the second axis (roll) is rotated to $\pm 90^\circ$, it makes the orientation of pitch align with the orientation of roll. In this situation, the pitch is changing with the roll and cannot accommodate rotation about its actual axis. This phenomenon is called Gimbal Lock, which degenerates the three-dimensional rotation system into two-dimensional space. However, in terms of the algorithm, the tilt detection can still work under the occurrence of Gimbal Lock, because the tilt identification only considers the rotations of roll and yaw.

The purpose of experiment 3 was to evaluate the performance of acoustic recognition models. For breath recognition, four models (comprising HMM-GMM, ResNet18, CLDNNs and ResNet18+CLDNNs), were separately trained with the spectrogram and MFCCs features, and then were tested on the same test dataset. For comparison of two features, the average accuracy of the models trained on MFCCs was 4.4% higher than the models trained on the spectrogram. The combined model ‘ResNet18+CLDNNs’ achieved the highest average accuracy (94.5%). This combined model used ResNet18 and CLDNNs to encode the MFCCs feature and extract the temporal relationship within their successive frames. Furthermore, it gave good performance even with a SNR of 0 dB. As the experiment results showed, the HMM-
GMM model has different characteristics from the other models. It gave an acceptable performance on MFCCs features and even returned high accuracy for the pure audio clips. However, it had poor performance on with the spectrogram, which accuracy, compared with MFCCs features, dropped by 10.6%. Compared to the other models with the spectrogram, HMM-GMM only returned 78.2% average accuracy - about 12% lower than the other models. This difference in accuracy showed that the HMM-GMM has poor ability to extract features directly from the spectrogram. Although the deep learning models showed good performance with breath sound recognition at different SNR levels, it is necessary to further improve the system with new architectures. In terms of the data sources, this experiment was limited by the data variety. Ideally, a good dataset should include various ages, heights, weights and stages of respiratory diseases. Due to the limitation of the resources, this experiment only considered the feasibility of breath sound recognition based on the different methods. Moreover, how to suppress the impacts of wind noise on the breath recording process is an interesting topic to be discussed in future research, in particular, because wind noise shares similar features with breath sound. The actuation detection is designed to count dose. Ideally, it informs the patient, through the flashing LED, without any back-end processing on the mobile or cloud server, if there are only a few doses remaining in the inhaler. In terms of the recognition models, this target is easily achieved by the aforementioned models. However, these deep learning models are challenged by the limited computation resources of the low power-consumption microcontroller used in the device. Therefore, the two typical classifiers, random forest (RF) and support vector machine (SVM), were employed to solve this issue and were evaluated in the second part of the third experiments. First, the classifiers learned from extracted acoustic features, and then both were tested on one test dataset. Finally, the comparison of different metric scores showed that the SVM had higher overall performance than the RF. In practical actuation detection situations, it uses a sliding window to capture actuated events. However, this sliding procedure has high time-cost. How to quickly search for interesting acoustic activities is a challenge and needs to be considered in the next stage of the research.
The final experiment was an exploratory clinical study. It included the engagement process and the daily tracking data returned by actual patients. In the engagement process, there were a series of stakeholders meeting from April 2017 to June 2018. These meetings collected stakeholders’ perspectives of smart electronics, which then informed the design process. Furthermore, participants talked not only about the electronic devices but also provided feedback on the electronic assessment. The conversations were useful for further research to improve patients’ self-management.

In the patient study, six participants with asthma or COPD were tracked over three weeks, yielding sizeable datasets on their inhaler use. The experiment compared self- and electronic assessment records; the latter revealed that participants frequently overestimated the quality of their pMDI technique. Compared with a simple dose counter or an inhaler trainer, the device provided the electronic records that not only objectively assessed the participants’ technique but also identified critical errors when the participant used the inhaler in the real world. This additional information, captured by the device, can be used and hence for the purpose of patient training and hence optimise the dose taken. Moreover, this digital information is not only for the direct benefit of patients; it provides information on adherence to medication and is therefore useful for professionals and clinical researchers. Coughing during the inhalation stage was another finding in this experiment, which will result in the delivery of a much lower dosage than is technically estimated. Although this experiment revealed very important results and the potential for further research, there were some limitations associated with the study. First, the sample size was small, and the participants were drawn from one geographical area. This may diminish the generalisability of the data. Compared to prior studies of smart inhalers [13][151], this was a preliminary study, in which inhaler-experienced patients participated over a relatively short period. In terms of the assessment standard, there is no published work which explains the exact relationship between medical efficacy and each critical error. Therefore, the final experiment assumed that every step contributed the same weight to the final assessment score. However, each critical step may cause an unequal effect on the efficacy of drug delivery. Moreover, this study did not consider the cough factor into the assessment standard, although coughing behaviour may
potentially reduce the adherence of medication. These limitations require further work in the future.

Although there is still much work to be conducted in the future, the findings from this research are encouraging. They contribute significantly to understanding the reasons for low adherence to medication. Although the system and methods of analysis require refinement, this work has made a major contribution to this relatively new area of investigation and provides a solid foundation for further system development and clinical research. Moreover, this work is an important and timely example of the global transformation in healthcare. This transformation will assist patients in COPD self-management, improve the efficacy of treatment and hence enhance patient life quality.
Chapter 8

Conclusion

8.1 Overview

This thesis reported work on the development of a pMDI telemonitoring system and elucidated the differences between patients’ self-reports and objective measurements of compliance. For patients with asthma and COPD, adherence to medication was difficult to measure in the past, often because of the communication barrier between patient-doctor. Poor adherence to medication may cause disease exacerbation, increase the economic burden on families and societies, and even mislead a professional’s diagnosis of the patient’s condition. As for traditional methods, it is a huge challenge to objectively quantify adherence to inhaled therapy. If there is a feasible method to track patients’ daily inhaler use, what is the actual situation of the inhaler use outside the experimental environment? This research aimed to address these questions and to provide solid fundamental work for further research. Furthermore, the thesis also contributed to the transformation of digital medicine that is rapidly gaining traction across the world.

The thesis commenced with a literature review of asthma, COPD and inhaler techniques to provide a context to the research and better to understand the existing problems in this area. Following the literature review, descriptions were given of the various systems which were continuously developed and optimised, using results from the experiments and by attending to comments provided during the engagement process. The successful development of the system provided reliable hardware to track patients’ inhaler use and technique. After the steps of system development, a series of test experiments evaluated the performance of the hardware and software. Finally, an exploratory clinical study was conducted to understand participated patients’ pMDI use in their daily lives. These experiments generated very interesting results for further explorations in this area.
8.2 Summary of contributions and findings

The novel contributions and findings can be summarized with the following statements:

**Development of acoustic monitoring device for pMDI.**

As previously introduced, there no specific device existed for pMDI usage monitoring and evaluation. Compared with the dry powder inhaler, a shaking stage is important for the press meter dose inhaler, ensuring that the drug particles are thoroughly mixed with the pressurized propellant. The development of the device was characterised by four stages: system evaluation, the first prototype, the second prototype and the third prototype. From the first to the final prototype, the size of the PCB was reduced from 32mm x 20 mm to 19.5mm x 19.5mm; the recorded audio quality was increased significantly, and the final device only consumed around 15 µA in standby mode. The trigger mechanism of the device was also evaluated in experiment 1, which revealed that the trigger workflow was unreliable and power-hungry. A considerably improved one-button trigger method was implemented in the final prototype, which was immune to false activation. The successful development of the device provided a good platform for successive studies.

**Recognition of breath sound was implemented by HMM-GMM model, and the accuracy was further improved by ensemble deep learning models.**

The breath sound identification is an important part of the system. Initially, HMM-GMM was used to identify the different phases of breath; this model was trained on MFCC features and the test result published in ‘IEEE transactions on Instrument and Measurement’. However, the poor performance in the presence of ambient noise is a weakness of HMM-GMM. This was confirmed by the following test, experiment 3. In this third experiment, the HMM-GMM model was trained on larger sample size, and it gave good performance for the recognition of clean audio data, but its accuracy dropped dramatically if the input samples became noisy. To further improve the performance of the system, deep learning methods were then development. The deep learning architecture ResNet18 was modified to recognise the audio samples; tests
confirmed that it had better performance than HMM-GMM under noisy conditions. Another model, CLDNNs, was also used to discriminate the phases of the breath. This had the highest accuracy of all of the three single models. ResNet18 and CLDNNs exploit different characteristics of the audio features, and the work described in this thesis concatenated their advantages to further improve the task of recognition.

**Identifying the deviation between self- and objective records:** Patients overestimated their inhaler techniques and failing to hold their breath was their biggest problem. Cough is a potential critical error affecting the efficacy of drug delivery.

The final experiment was conducted to track patients’ inhaler use and technique in their daily lives. In this experiment, self-reports suggested 70.8% of inhaler use cases were acceptable and 24.5% of use instances were assessed as Good. However, the electronic assessment showed the good technique only occupied 13% of inhaler use cases and 34.9% of the inhaler technique assessments was poor; the proportion of acceptable assessments fell to 37%. The difference between the self- and electronic assessments revealed that participants were overly optimistic about their inhaler techniques. As for the details of the participants’ techniques, most of the participants failed to hold their breath for over 10 seconds. The inhalation technique was also a common source of error in this experiment. Only 12% of the instances satisfied the standard for inhalation technique. One participant coughed while inhaling the drugs, which was another interesting finding. The impact of the cough has been ignored in many studies but it potentially reduces the adherence to medication, since it prevents patients failing to hold their breath and may even cause a patient to cough up the inhaled drugs; this, in turn, may significantly reduce the deposition of the drug on the lung surface.

**Provided an objective method to measure adherence to medication, this method combined acoustic details and motional details**

As aforementioned, the system could track the patients’ pMDI use and technique objectively. Currently, other acoustic monitoring research for patients with asthma or COPD ignore motion details. However, the motion details, such as shaking the inhaler
and holding the inhaler upright, are critical steps for establishing adherence to inhaler technique. This thesis described the development of a system to record acoustic and motion information simultaneously. The second and third experiments discussed the identification algorithm or models for the two types of signal and assessed their feasibility and reliability. The acoustic and motion information provided a more comprehensive opportunity to understand the patients’ inhaler use and technique. This information can also be used to correct technique in future studies. Moreover, this system digitises the adherence to medication and enables professionals and researchers to acquire high quality, high volume information from their patients.

8.3 Further work

8.3.1 New clinical studies

Many clinical studies could be conducted to better understand the patients’ situations and to identify methods to improve adherence to medication. The final clinical experiment recruited six participants aged over 55; they were both from the same geographical area. Although the study focused on each individual’s personal inhaler use and technique, it is necessary for future studies to recruit more participants from different areas; this could improve the generalisability of the study and increase the statistical significance of the conclusions. As the previous chapters mentioned, electronic information feedback is also important for patients to improve their self-management. Studying the impact of electronic information feedback on patients’ inhaler technique improvement is also important to evaluate the effect of digital adherence to inhaler technique. In addition, the studies could explore new approaches to information delivery to enable patients and professionals to improve treatment quality.

8.3.2 How to correct patients’ inhaler technique

For the wider picture of the smart inhaler, tracking the patients’ inhaler use and technique is only a preliminary step. One of the primary targets is to correct technique
to improve life quality. Therefore, studying how to use the smart inhaler to notify and correct errors is a significant component of future research. The error notification methods can be classified as asynchronous and synchronous. The asynchronous approach is similar to the further clinical study mentioned in section 8.3.1. It tracks the inhaler technique over a period of time and then provides a summary report for patients. The synchronous method requires the smart device to have good real-time performance so that it can detect the occurrence of the critical errors at the point of use. Ideally, it could employ several feedback methods, such as vibration and flashing LEDs, to notify patients immediately if critical errors have occurred. There also are other potential interactive approaches to assist patients to improve their technique, such as the use of a small screen. These all represent interesting topics for further work. In the recent future, the device will be expected to identify usage problems for patients with poorly controlled asthma or COPD. For the ‘inhaler naive’ patients, the device will ensure their technique is optimised from the commencement of treatment.

8.3.3 Improved recognition algorithms

Experiment 3 confirmed that the ResNet18+CLDNNs deep learning model achieved an average accuracy of 94.5% on the limited number of test samples. In follow-up studies, collecting a larger number of training samples from various individuals would be essential to improve the model’s generalisability and accuracy. The design of an effective and efficient architecture based on a deep learning model for acoustic classification is a primary consideration for the future. Furthermore, the algorithm for sound enhancement/noise suppression is vital and a fertile area for future work, which aims to improve the perceptual quality of breath signals for robust breath recognition. As discussed in chapter 7, one of the enhancements that would benefit the system would be adaptive noise cancellation, but it needs the support of multiple-microphones [162]. Recently, a deep learning model was applied to suppress noise to improve speech quality [163]; this can also be used to enhance breath sound. The algorithms for sound enhancement/noise suppression could be implemented as part of future work.
8.3.4 A common device for various types of inhaler

This research developed a monitoring device attached to a pMDI inhaler but the long-term objective aims to develop a common device, fitting the major types of inhaler. Different enclosure adapters would ensure that the common device attached to various shapes of the inhaler. A mobile application would configure the device and then select the assessment standard that corresponded to the current inhaler.

8.3.5 Developing a mature system and creating a network community

Although the prototype hardware made significant contributions to this area of research, the hardware still has limitations, such as noise suppression and interactivity. In chapter 7, the dual-microphone design was discussed, and will be implemented in the next version of the hardware; this will enable a reduction of background noise and enhance audio quality. In this research, hardware development mainly focused on feasibility and reliability, rather than emphasizing connectivity and interactivity. For the final version of the device, the recorded data was downloaded via an SD reader that was inconvenient for patients and professionals. Instead, wireless connectivity, such as Wi-Fi and Bluetooth Low Energy (BLE), is a necessary future objective. It not only simplifies the step of downloading files but also makes the device interactive for users. The development of an online platform, such as cloud services, websites and mobile applications, are also important aspects of further work. Once the platform is established, it will facilitate the growth of online patient communities. Patients will get the opportunity to share their experiences, gain encouragement from others and even make friends. Healthcare professionals could participate in these communities to help their patients. However, the internet community is a double-edged sword. Many factors, such as privacy, internet harassment and online security, must be carefully considered in any such further development.
8.4 Closing remarks

In 2016, information about smart inhalers was almost nonexistent. However, with the rapid growth of smart devices, smart inhalers have gained traction and are a subject of major research globally. This thesis presented novel contributions and findings to this field, yet it was very much preliminary work regarding the emerging picture. As a part of the digital medical transformation, the work provided a good starting point to electronically measure adherence to inhaler technique. Future experiments and work will focus not only on the development of the entire system but also on how to use these resources to actually improve the patients’ life quality and to increase professionals’ work efficiency.
References


2013.


Appendix A

Journal Publications

A.1 An Inhaler Tracking System Based on Acoustic Analysis
Hardware and Software (2019)
An Inhaler Tracking System Based on Acoustic Analysis: Hardware and Software

Wenyang Xie®, Student Member, IEEE, Patrick Gaydecki, Senior Member, IEEE, and An Louise Caress

Abstract—In treating asthma and chronic obstructive pulmonary disease (COPD), acquisition of authentic and effective feedback from patients on regimen adherence is difficult. Face-to-face and oral reporting methods do not satisfy current intelligent medication best practices. This paper presents a system to track and analyze daily inhaler usage. A portable electronic device that attaches to the inhaler uses an accelerometer and capacitive sensors to detect users’ motion and an embedded digital microphone to capture sounds while the inhaler is in use. In terms of analysis, sound features are extracted, and breath phases are identified by employing a hidden Markov model with a Gaussian mixture model. A feature template is also constructed and used to search for and identify “canister pressed” events. The system provides objective feedback, quantifying asthma, and COPD patients’ adherence to medication regimens. Although interest in asthma adherence to medication regimens is growing, there is still a relative paucity of research and, indeed, compliance devices in this area; the tracking system can help doctors better understand the patient’s condition and choose an appropriate treatment plan. At the same time, patients can also improve their self-management by system feedback.

Index Terms—Acoustic monitoring, breath phase identification, hidden Markov model—Gaussian mixture model (HMM-GMM), inhaler techniques, random forest (RF), support vector machine (SVM).

I. INTRODUCTION

The number of patients with asthma and chronic obstructive pulmonary disease (COPD) is gradually increasing. A recent projection predicts COPD will be the third leading cause of death in much of the world by 2020 [1]. Typical treatment for asthma and COPD is inhaled medications, but this requires good coordination of exhalation and inhalation, especially when using the metered dose inhaler (MDI) technique [2]. Appropriate use of the MDI technique can deliver the correct dose of medicine into patients’ lungs; however, one study showed that only 14% of patients with asthma used their MDI correctly, while accurate use for COPD patients was 15% [3]. To achieve therapeutic benefits, it is critical that patients use inhaler techniques correctly and that doctors and health providers understand patients’ conditions and the necessary treatment. Currently, the ability to assess how regularly and effectively patients use their inhaler on a daily basis is limited. Such assessment of patients’ inhaler technique is usually done face to face by a health provider, which is, in practice, highly inefficient. Some inhalers are designed to provide time of use as well as dose count, but few can provide any indication of whether doses were inhaled correctly or at all. Few devices allow accurate and objective assessment of how patients use the full range of inhalers on a daily basis during routine care [3]–[5]. This low rate of accurate use and the difficulty of confirming adherence implies that patients are often not compliant with health providers’ or doctors’ recommendations. Incorrect use of inhaler techniques is a significant factor in prescribed treatment [6]. Moreover, poor inhaler use also unnecessarily increases medical budgets and pressure on healthcare organizations, while patients suffer from inefficacy of treatment and other potential problems. Currently, dose counting is used to track patients’ inhaler use in clinical trials [7], but only a limited number of inhalers have in-built dose counters (usually they are expensive; these are not inhalers that would routinely be prescribed in practice). Furthermore, tracking daily inhaler use is still difficult due to the lack of a reliable method [4]. Thus, there is a considerably space for improvement in assessing whether patients are taking the right number of doses. It is clearly important, given the rate of nonadherence in asthma and COPD patients and the known adverse, potentially life-threatening, consequences of lack of adherence.

An algorithm to identify breath sounds is also significant to the implementation of adherence analysis. One study used power spectral density to discriminate between inhalation and exhalation, with an accuracy of 92.8% in a relatively silent environment [8]. Discrimination of inhalation and exhalation is based on a single segment of breath sound. However, a serial raw-audio signal not only contains multiple segments of interest but also contains unexpected information, such as voice, music, and noise. It is thus necessary to demarcate breath sounds within a raw audio signal. One study detected inhalation sounds based on Mel-frequency cepstral coefficients (MFCCs), with an accuracy of 89% [9]. Ruisky and Lavner [10] constructed an efficient algorithm to label breath sounds in audio signals, with a correct identification rate of 98%. Two studies used similar methods, template matching, to detect breath sounds. The template
was constructed by using singular value vectors, which used singular-value decomposition to degrade the matrix of MFCC, to represent the essential characteristics, and to increase efficiency [9], [10]. Various methods are used for classification in analysis of acoustic sounds, such as artificial neural networks, k-nearest neighbors, the Gaussian mixture model (GMM), the hidden Markov model (HMM), and time delay neural network [11], [12]. For example, GMM was employed to classify acoustic signal of moving targets, its classification accuracy was up to 92.15% in mono input [13]. In practice, the duration of inhalation and exhalation is unpredictable. The variable length of each segment cannot be represented by fixed-size matrices. Therefore, classification methods are required to process the variably sized sequence.

This paper presents a system including hardware and software that tracks inhaler usage based on acoustic monitoring. The electronic device has low power consumption, is miniaturized, flexible, and can be retroactively fitted to standard inhalers. The software uses efficient, embedded algorithms to detect and recognize inhaler usage. The original sounds are truncated based on sound activity detection (SAD), with the aim of removing unnecessary silent parts. HMM-GMM, which is one of the most popular solutions in sequential prediction [14], is used to create an acoustic model and identify breath phases. The GMM is used to model the feature distribution and can be thought of as a single state HMM. Usually, the GMM contains several mixtures to fit the input signal and then its mixture distribution is used as the emission probability of an HMM. The GMM is particularly appropriate in this context due to the variable nature of the signal being classified. Dose counting is also based on audio analysis, with basic audio features classified by a random forest (RF) and support vector machine (SVM) approaches.

II. SYSTEM DESIGN

To measure the rates of accurate inhaler use, this system is designed to monitor daily inhaler use and analyze the collected data. It provides assessments that show whether patients are in compliance with the recommended usage steps. The system has two main parts. The hardware implements automatic capturing and recording of breath sounds, while the software identifies breath phase and analyzes adherence on a PC. The recommended steps of inhaler use (see Table I) were developed according to Asthma U.K. [15].

Three of the steps in the recommended instructions for MDI use are most significant and require good coordination of breath while pressing the canister. Therefore, the designed system focuses on tracking and timing of the breath phase. A system block diagram is shown in Fig. 1.

The exhale slowly phase has the lowest accuracy rate for all recommended steps. If this stage of the procedure is disregarded, accurate identification rises to 96.9% [16], which means that the exhalation step accounts for a large proportion of the failure to adhere to the proper technique. Therefore, the monitoring device must record the entire process of inhaler use. Tracking should be triggered, and the record-function should be run before the first exhalation. This triggering mechanism is implemented by a combination of detection of motion and touch. Furthermore, it also activates the system from standby status, which is a mode used to extend battery duration, reduce frequency of recharging, and minimize inconvenience of daily use. The triggered device then records the whole breathing progress and limits the frequency spectrum within a specific range via digital filters. Finally, raw audio data are labeled with a time stamp and saved to storage memory. Adherence is analyzed on a PC, as shown in the right part of Fig. 1. It contains four parts, including data preprocessing (such as normalization, hamming framing, and

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>RECOMMENDED INHALER USE STEPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDI Step by step</td>
<td>Phases</td>
</tr>
<tr>
<td>1. Shake the inhaler well before use</td>
<td>-</td>
</tr>
<tr>
<td>2. Remove the cap</td>
<td>-</td>
</tr>
<tr>
<td>3. Empty lungs of air before inhalation</td>
<td>Exhalation</td>
</tr>
<tr>
<td>4. Place inhaler in mouth between teeth and close mouth around it.</td>
<td>-</td>
</tr>
<tr>
<td>5. Inhale slowly, at the same time pressing the top of the inhaler once while continuing to breathe in slowly until a full breath is taken.</td>
<td>Inhalation &amp; Spray</td>
</tr>
<tr>
<td>6. Remove inhaler from mouth, hold breath around 10 seconds and breathe out</td>
<td>Exhalation</td>
</tr>
</tbody>
</table>
feature extraction, model training, construction of the "pressing event" template, and fusion of the results. All sound data are split into segments. Each truncated part is then converted to a feature matrix of MFCCs and normalized on a common metric for further training. As for the canister pressed event detection, a fixed sliding window scans the whole recording and extracts audio features, such as MFCC, zero-crossing rate, and spectral centroid. In the final stage, the trained classifier outputs the predicted result.

A. Hardware Architecture

A diagram of the hardware is shown in Fig. 2. Its purpose is to capture the user's motion, record breath sounds, and store data. The electronics feature low power consumption since the microcontroller is sleeping in standby mode most of the time. For reduction of static power consumption, this mode also turns off most of the peripheral components via a load switch chip. The cooperative triggering of the capacitive sensor and three-axis accelerometer activates the microcontroller when the device is picked up and prepared for use. The activated microcontroller then switches on all peripheral loads, such as LED indicator, SD card, and MEMS microphone. According to empirical research on usage duration, the average recording is 25 s and the external memory stores this time-stamped audio. Finally, the microcontroller turns off the unnecessary peripherals and returns to standby mode until the next activation (program structure is shown in Fig. 3). The size comparison given in Fig. 4 shows that the electronics are very small and can attach to various existing inhalers with specific enclosures.

B. Algorithm Development

The inhalation and exhalation phases are characterized by significantly different spectral features, as shown by the example in Fig. 5. The spectrum of the inhalation phase is typically broadband in nature, extending from 20 to 8000 Hz. In contrast, the energy of the exhalation signal is maximal around 500 Hz and gradually falls away from this point.

Therefore, a method was developed for the automatic detection of breath sound in speech and identification of the breath phases. The algorithm consists of two main parts. The first is the detection of the breath phases, which is used to train an acoustic model via supervised learning. First, during preprocessing, the raw signal is limited to filter out unexpected noise within the range 20 Hz–8 kHz. SAD is employed for slicing the temporal sequence while retaining the appropriate segments for the subsequent computation of the MFCC. In the next step, the labeled training samples are
split into overlapping frames. A single frame is computed as an MFCC feature vector, normalized, and the processed matrix trained via the GMM-HMM method; this yields an acoustic model. The test samples are used for performance validation. In addition, to quickly detect when the canister is pressed, a template is built. The final output is the merged results of the two parts, calculating the time intervals between identified locations. The whole process is shown in Fig. 6.

1) Breath Sound Preprocessing:

a) Bandpass filter: The raw signal passes through a Butterworth bandpass filter, which has a passband between 20 Hz and 8 kHz, to remove noise and enhance the accuracy. Outside of this band, the analysis revealed that there is almost no contribution from the breath sounds. Attenuation of the out-of-band signal, which derived from a number of ambient sources (often percussive in nature) was, therefore, essential in maximizing the signal-to-noise ratio.

b) Sound activity detection: The implementation of SAD reduces the computation load and increases the efficiency of detection. It is based on the crest factor, expressed in decibels as

\[
Crest_{factor}_{dB} = 20 \times \log_{10} \left( \frac{f_{peak}(x)}{f_{RMS}(x)} \right)
\]

\[
f_{RMS}(x) = \sqrt{\frac{\sum_{n=0}^{N-1} x_n^2}{N}}
\]

(1)

The crest factor reflects the ratio of the peak and mean levels and indicates the activity of sound when that ratio exceeds a given threshold. Fig. 7 illustrates the relationship between the active sounds and the crest factor. Each margin of detection window is 10% of the raw sound, to maximize the retention of useful information.

c) MFCC extraction: The feature extraction block separately divides the effective segments into overlapping frames; the window length and overlap between windows are 25 and 10 ms, respectively. Each of the MFCC features is computed and generates a vector. This process is shown in Fig. 8.

d) Standard normalization: To avoid mismatched metrics in the training phase, it is necessary to map all data on the same scale of statistics. The function is given by

\[
f_{normal}(x) = \frac{x - \mu}{\sigma}
\]

(2)

where \(\mu\) is the mean of the signal and \(\sigma\) is the standard deviation.

2) Breath Sound Training: A multivariate Gaussian model was employed to express the probability distribution of the feature matrix. This is given by [17]

\[
p(x) = \sum_{k=1}^{K} \pi_k N(x; \mu_k, \Sigma_k)
\]

\[
= \sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}} \quad (\pi_k > 0)
\]

(3)

where mixture weights \(\pi\), covariance matrix \(\Sigma\), and mean \(\mu\) are estimated via the expectation–maximization algorithm, which aims to find the most appropriate parameters for a given fit.

The computed GMM is considered as a single and special state of HMM [18]. The emission probability of the latent state, in fact, is calculated by the previous step as

\[
b_j(x) = \sum_{k=1}^{K} \pi_k N(x; \mu_{j,k}, \Sigma_{j,k}) \quad (\pi_k > 0).
\]

(4)

These observations \(X\) based on the output of states (such as inhalation, exhalation, and nonbreath) correspond to the computed GMM distributions in the previous step and are shown in Fig. 9.

\(A_{ij}\) is transition probability. According to the Bayesian theorem, the posterior probability is expressed as follows:

\[
p(S_j | X, \theta) = \frac{p(X | S_j, \theta) p(S_j | \theta)}{p(X | \theta)}
\]

\[
\theta = \{A, \mu_j, \Sigma_j\}.
\]

(5)

Using the conditional independence property [18], the expression takes the form

\[
p(S_j | X, \theta) = \frac{a_j(\theta) \beta_j(n)}{p(X | \theta)}
\]

\[
a_j(\theta) = p(x_1, \ldots, x_n, S_j | \theta),
\]

\[
\beta_j(n) = p(x_{n+1}, \ldots, x_N | S_j, \theta).
\]

(6)

Assuming the observed sequence \(X\) is known, the Baum–Welch algorithm can be employed to evaluate (6) to obtain HMM parameters.

3) Detection of Spraying: Detection of effective usage (dose counting) is indispensable to recording the frequency and duration of medical usage of MDIs. It also reflects the patients' dependence on the prescription, i.e., whether patients are decreasing or increasing their use. Therefore, detection must be effective and efficient. In terms of MDI, the acoustic signal arises from the ejection of an aerosol cloud, by a propellant, from the metering valve [19], which then expands in an expansion chamber and is sprayed out by an actuator nozzle. Analysis of sample recordings shows a typical pressing feature indicated by a peak in the frequency domain within the range of 1600–1800 Hz. Fig. 10 shows that the majority of the energy is aggregated on this subband and the rest is widely spread on the whole band. Furthermore, the graph of autocorrelation, shown in Fig. 10, displays periodic characteristics. Furthermore, it is hard to extract a fixed central frequency to identify the sound, because of changes in the container pressure and pressing jitter. Fig. 11 shows the major frequency
statistics of 46 samples; the average frequency is 1658 Hz, and the standard deviation is 37.7. Therefore, to avoid unnecessary computation, a band gate is set within 1658 Hz ± 10% to block any unsatisfactory signal, before the feature extraction process. Capturing the major frequency of the sound allows partial identification of the unknown signal, but only considering this will result in low specificity. Additional features are considered to improve performance and accuracy of detection.

a) Extraction of acoustic features:

1) MFCC is an appropriate feature to reflect hearing perception. Equation (7) normalizes each coefficient and averages them

\[
M_{\text{norm}} = \frac{1}{N} \sum_{j=1}^{N} \frac{M_j - \text{mean}(M_j)}{\text{var}(M_j)}
\]  

(7)
where $M_i$ is the $i$th coefficient in the MFCC matrix and $M_j$ is the $j$th subframe.

2) The zero-crossing rate is an important feature in this detection. It is expressed by

$$f_{zc} = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \text{sgn}(x(n)) - \text{sgn}(x(n-1)) \right].$$  \hspace{1cm} (8)

3) The spectral centroid indicates the locations where most of the energy is concentrated. The brightness (in a timbral sense) of a sound depends on this. It is expressed as the spectral-weighted sum divided by its unweighted sum [20]

$$f_{st}(n) = \left( \sum_{k=0}^{\frac{N}{2}-1} k|F(k, n)| \right) / \left( \sum_{k=0}^{\frac{N}{2}-1} |F(k, n)| \right)$$  \hspace{1cm} (9)

where $N$ is the fast Fourier length of each temporal block.

4) The spectral spread describes the flatness of the spectrum; it is an indicator of the noisiness [20]. This feature is given by

$$f_{ss}(n) = \left( \frac{\sum_{k=0}^{\frac{N}{2}-1} (k - f_{st}(n)) \cdot k|F(k, n)|}{\sum_{k=0}^{\frac{N}{2}-1} |F(k, n)|} \right)^{\frac{1}{2}}.$$  \hspace{1cm} (10)

5) Energy entropy is the measurement of the distribution of energy. It is based on Shannon information entropy. Incriment of energy entropy means the sound contains more uncertainty of energy and vice versa. It is given by [21]

$$H(x) = \sum_{i=1}^{N} p_i \log_2(p_i)$$ \hspace{1cm} (11)

where $p_i$ is the probability of obtaining useful information. According to this concept, the information probability can be substituted by the ratios of subblocks 

energy to the total energy. Therefore, energy entropy is defined as

$$H(n) = \sum_{i=0}^{M} \frac{|f_i(n)|^2}{\sum_{i=0}^{M} |f_i(n)|^2} \log_2 \left( \frac{|f_i(n)|^2}{\sum_{i=0}^{M} |f_i(n)|^2} \right).$$ \hspace{1cm} (12)

b) Classification method: The classification step uses two popular classifiers on a small data set. RF is a collection of “weak” classifiers (decision tree) which are combined and contribute to a strong classifier [22]. It stochastically selects subset of features and generates several mutually independent trees. The output prediction is voted by the “weak” classifiers [23]. The second classifier is SVM, which is a discriminative method that is implemented in regression and classification. Its classification strategy is based upon choosing an appropriate hyperplane that contains the maximum decision boundaries [18], [24], [25]. Recently, SVM has gained popularity as a classification method in areas of medical diagnosis and detection [27]-[30]. The hypothesis parameters of classifiers are tuned via a grid search function [31]. Fig. 12 shows the visual result for the continuous pressing test. The first row of Fig. 12 shows the SVM prediction and the bottom row shows the results based on RF classification. In Fig. 12, SVM and RF both require approximately 0.012 s for prediction of a segment 15 s in length, thus showing similar performance with this small number of features.

4) Detection Output: The final output combines all the detected events. Fig. 13 displays a waveform of actual inhaler use, in which the different steps are automatically labeled by the detection algorithms described in this paper. The transparent rectangles mark the locations of the breath timeline, while the dark square is the detected spray event and the time intervals within the steps are shown by hatched rectangles. This format provides useful information, such as the sequence of use steps, the duration of each step, and the number of doses. For example, Fig. 13 shows that the user held their breath for only a short time of a few seconds, which may cause the medicine to be immediately exhaled [32].

III. PERFORMANCE EVALUATION

A. Device-Triggered Evaluation

A study was designed to evaluate the reliability of the electronics and the collection of information, corresponding to the ease of use of the smart inhaler. The study took place over 2 days and included 4 volunteers. Before the study, the volunteers were given minimal training on the correct use of the device, and then they were asked to count the number of times they used the monitoring device and record it as a written record. Table II shows that a total of 62 uses were reported by volunteers, but due to false triggering, the devices recorded 65 audio traces containing three unrelated records. After filtering out the incorrectly triggered recordings, the accuracy increased to 93.5%.

The electronics developed and used to acquire volunteers’ data is shown fit to an inhaler in Fig. 14. At the time of writing, this inhaler is the most widely used design. As our system evolves and becomes more reliable, it is planned to trial it on a number of different inhaler models. Since the
accuracy and reliability of the system are based upon an automated algorithm that identifies the various phases of the signal, we do not anticipate that the change of inhaler will affects the system’s overall performance.

B. Breath Phase Detection

In this test, healthy volunteers were divided by gender into two groups; in each group, one volunteer was trained in or had previous experience using the inhaler. Group 1 recorded a total of 60 breath sounds each, and group 2 recorded 20 uses for each volunteer. The test was carried out in an indoor environment. Table III shows that inhaler use of volunteers A, C, D, and F, who had related experience, was correctly recognized at a rate of over 90%. However, for volunteers B and E, who were both inexperienced participants, the average accuracy fell to 88%. As a result of unskilled operation, the breathing sounds are quieter and inhalation is slow. This results in lower
accuracy of detection for volunteers B and D. Although more robust noise rejection algorithms are still being developed, preliminary tests to establish performance in the presence of noise have been assessed, as evidenced in Table III. The original high-quality signals were degraded by adding randomly selected noise data, which included traffic and urban street noise. As shown in the final column, with an SNR of 25 dB, the accuracy of the system drops to a mean of 79.77%, in comparison to a mean of 92.83% under ideal conditions. The reduction of performance in the presence of noise is inevitable but will be ameliorated by compensation algorithms currently under development.

C. Spray Identification

A total of 130 samples were collected in an indoor environment, among which, 90 contained spray sounds and while the other 40 were irrelevant noises. Table IV shows a comparison of the performance of the two classifiers. After optimizing the hyperparameters for the two classifiers, RF results in higher recall than SVM, indicating that RF has strong discriminative ability for the relevant sample. However, the precision of RF is slightly lower than that of SVM. In this case, it is difficult to judge the performance of the two classifiers in this data set. Therefore, a combinatorial average of the precision and recall was used as an indicator—F1 score. According to the F1 score, SVM shows better performance than RF, and its accuracy reaches 96.1%.

IV. CONCLUSION

Tracking and quantifying the adherence of patients to prescribed medication regimes is an important factor in increasing the efficacy of treatment, especially with the rapid development of medicine technology. This paper presents a monitoring system that includes both hardware and software that provides a practical method for estimating the adherence of pMDI users to medication regimens. Although other methods are possible, based on artificial intelligence or machine learning, the algorithm presented here was efficient and robust, even in the presence of some noise. However, the system also needs improvement in the future work. Detection is compromised in very noisy surroundings, such as on a busy street or in shopping malls; its accuracy degrades in such environments. Therefore, a noise estimation model for such common situations and breath sound enhancements are the next focus of research. Furthermore, the entire training data set was collected in a quiet environment. In terms of signal analysis, a pure signal can reflect the characteristics of the target. However, this may cause the classifier to become incapable of generalizing, which can be considered as an example of an overfitting problem. Therefore, the collection of positive samples with different backgrounds is also a focus of future research. Initial tests suggest that the system has significant potential for the subjective monitoring of patients, aid clinicians in the analysis of compliance, and will ultimately improve medical self-management for users of inhalers.

REFERENCES

Assessing the quality of inhaler technique – patient report versus acoustic monitoring: an exploratory study

Wenyang Xie, Patrick Gaydecki, Jean Hennings, Ann-Louise Caress

ABSTRACT

Background:

Hundreds of millions of people suffer from asthma and COPD across the world, but less of them use inhaler appropriately, causing inefficacy of treatment, worsening of their condition, and increasing healthcare costs in the millions every year. This study was conducted to investigate the deviation of pMDI usage and technique, in real life, between patients’ self-reports and electronic reports.

Methods:

Six participants with clinician-diagnosed asthma or COPD participated in this study and yielding 233 readings (containing a total of 323 doses taken by the participants). The adherence to pMDI use was tracked and measured by a consistent system; the comparison of subjective self-assessment and the objective assessment in real life were analysed in this study.

Results:

Participants thought their techniques were acceptable (OK) in 70.8% of inhaler use cases; they assessed 24.5% of use times as using Good technique. However, in the electronic assessments, the acceptable technique (OK) was reduced to 37%; only 13% of assessments were good, and 34.9% were Poor. Concerning details of participants’ inhaler techniques, only 5.9% of inhaler use cases show that participants hold their breath longer than recommended instruction. Poor inhalation technique was ranked the second error, with an average of 88% of the total instances using the improper technique; only 50.3% of inhaler uses were completely exhaled before actuation.

Conclusions:

Compared with objective assessments, participants overestimated their pMDI inhaler techniques. The electronic monitoring provided a feasible method for measuring pMDI technique. This detailed information could assist healthcare professionals and patients in enhancing self-management.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) and asthma are major airway diseases, and, together, they accounted for 6.6% of global deaths in 2017. According to the global burden of disease (GBD) dataset (1990–2017), chronic respiratory diseases have become the third cause of death worldwide. Inhaled therapy is the mainstay of treatment for asthma and COPD, since treatment can be delivered to the site of the pathology quickly and efficaciously. Pressurised metered-dose inhalers (pMDIs) and dry powder inhalers (DPIs) make up the majority of the inhaler market. Effective treatments of asthma and COPD require long-term adherence to medication. Poor adherence is frequently related to high risk of disease exacerbation, hospitalisation and even mortality. Patients and their families suffer from economic burdens and poor quality of life. The diseases also cause heavy economic burdens...
for countries. For example, in the EU, the average annual direct and indirec
t cost for asthma and COPD was €82.3 billion in 2011.13 In the UK, the cost of direct healthcare for asthma was £964.9 million in 2012.14 A study projected that, from 2011 to 2030, direct costs of COPD in
England and Scotland would increase by approximately 53.3% and 30%, respectively.14

Patients often fail to use pMDIs with an appropriate technique, despite education.1,13,14 Luczak-
Wozniak et al.’s study stated that 98% of the patients with asthma and COPD were confident
about their inhalation technique, but at least one error occurred while most sampled patients
were using their inhalers.15 In Arora et al.’s experiment, 94% of patients used pMDI with
inappropriate techniques, and around 46% of them did not hold their breath, or only held
their breath for a short time, after inhaling.15 Choosing the proper timing between actuation
and inspiration is a challenge, and 45% of patients could not actuate the inhaler with the right
timing, which led to a loss of lung deposition.17 Moreover, several studies have indicated
that slow inhalation makes for higher lung deposition, but many patients inhaled drugs too
fast.16-19 Furthermore, low peak inspiratory flow rate (PIFR) <30L/min with delayed actuation
has been shown to contribute to reduced dose loss in the lungs.16

Several approaches are used to measure adherence to medication.7 Self-reports, telephone
interventions and regular interviews are common evaluation methods, but they are inefficient
and subjective.7,20-22 Direct methods, such as blood and urine analysis, are objective and more
accurate. However, they are usually invasive and costly and require professional skills.23
Recently, Hassal et al. assessed adherence to medication by analysing inhaled drug residuals
in hair samples, but this method’s accuracy is limited.9 Moreover, indirect methods cannot track
patients continuously, which means patterns of adherence are still unknown.13 An ideal
assessment of adherence approaches should be cost-effective, practical and easily interpreted.24

Monitoring adherence by the electronic device has begun to develop over the last ten years.
Although some methods record usage of metered-dose and time,25-27 useful pieces of
information are missed, such as critical errors and PIFR. To solve the problems, acoustic
analysis is employed in adherence measurements. Inhalation, exhalation, actuation and PIFR
show specific temporal and spectral characteristics, which can be classified by signal
processing and machine learning methods.28-29 Some studies have used the INHALer
Compliance Assessment (INCA) device to monitor patient use of DPIs.28-31 However, few
studies develop a specific electronic device to assess patients’ pMDI techniques.

Therefore, the objectives of this study were to develop an electronic device and to assess the
real-life inhaler adherence of patients with asthma and COPD.

METHODS

Ethics approval

The study was conducted according to the ethical principles of the Declaration of Helsinki,
and informed consent was obtained from all participants. The ethics approval was granted
by North West - Greater Manchester East Research Ethics Committee (REC reference no.: 18/NW/0205, IRAS project ID: 231016).

Engagement process

From April 2017 to June 2018, we utilised a series of stakeholder meetings hosted by the
British Lung Foundation North West Development team, made up of four patient support or
advisory groups in Greater Manchester and Blackpool. These meetings were used to acquire patients’ perspectives about electronic monitoring devices. The collected information fed into the device design and functionality.

Electronic design

Critical errors in inhaler technique can reduce the efficacy of inhaled drugs37, so assessing adherence by identifying errors is a practical approach. Therefore, we developed an electronic device to capture the performance of key steps in patients’ use of pMDIs. The steps for correct inhalation are shown in Table 1.36-38, These steps fell into two types sensing types—motion events and acoustic events. Motion events (step 2 and step 4) were recorded by the acceleration of the accelerometer and acoustic events were tracked by microphones. This electronic device was based on an electronic inhaler tracking system37, but we used mechanical buttons to activate it instead of touch sensing, which decreased false-triggers. The device recorded sensor data using a secure digital (SD) card while patients were using their pMDIs. The working flow is shown in Figure 1, and the interpreted details for each sensor’s processed data are illustrated in the grey blocks.

PIFR estimation

The time-domain signal and power spectrum density have both been used to estimate PIFR36,39. These methods have been stated for specific devices and do not provide a normalised benchmark. However, lung volume is constant for an individual, so the PIFR increment decreases as the duration of inhalation increases. This relationship is also shown in inspiratory flow studies31,40. Therefore, we used the duration of inhalation to estimate the level of PIFR.

Shake detection

The device’s accelerometer was used for this purpose; shaking events were detected by the rapid shaking detection (RSD) algorithm40.

Tilt sensing

Detecting whether the patient holds the inhaler upright is equal to validating the inhaler’s tilt angle. We used accelerometer data for a vertical inhaler as a coordination system and identified abnormal events by their deviation from that reference42.

Study design

For three weeks, from June 2018 to August 2018, we conducted an electronic monitoring assessment concerning participants’ inhaler use in their daily life. Before the study, we instructed participants on operating the device (see Figure 2). Participants were asked to complete daily logs (Table 2). After three weeks, we collected the questionnaires and compared the self- and electronic records (Figure 3).

Participants

The participants comprised six volunteers, drawn from the four advisory/support groups involved in the engagement work. All had clinician-diagnosed asthma or COPD for more than five years duration. The participants used their pMDI with the electronic monitoring device attached to their inhaler as shown in Figure 2. One of the participants used a ‘spacer’ with their pMDI. The participants’ inhaler use and techniques were tracked by the device for over three weeks, and 233 recordings were collected from them.
**Assessment standard**

Inhaler usage was captured via paper and electronic records. In the usage log, participants were asked to make self-assessments of their inhaler use frequency and technique (see Table 2). Electronic records were interpreted at five levels to objectively record patients’ inhaler techniques. The levels corresponded to the levels of self-assessment. The assessment standard\(^{10,43,44}\) is shown in Table 3, and we assessed the patients’ technique using the formula:

\[
y_{score} = S\left(\eta \sum_{n=1}^{N} \omega_n e_n \right)
\]

where \(e_n\) is the score of actuation success; \(e_m\) represents the score of each step except actuation success (\(e_f\)); \(S\) is scalar weight and the sum of weights \((\omega_n)\) is equal to one.

According to equation (1), we quantified and linearly mapped acoustic data on a number line \(y_{score}\). This score corresponded to levels of self-assessment (Very good, Poor, OK, Good and Very Good).

**Data analysis**

Data analysis was conducted in two stages. First, we compared paper and electronic records of dose frequency over three weeks. Some participants failed to record correct doses; who recorded once but took twice doses. Where this occurred, to make the two record sources comparable, we averaged scores of adjacent doses taken. Second, we summarised individuals’ inhaler techniques as numerical data. Box and bar charts\(^{45}\) were employed to illustrate the numerical distribution of critical steps for participants.

**RESULTS**

We analysed data collected from six patients over three weeks, yielding 233 readings. However, participants occasionally forgot to record their self-assessment and one participant did not fill this column as we required. In total, therefore, only 192 valid self-assessments were compared with the electronic scores in Figure 4. Table 4 presents a summary of data from participants’ paper records. The number of electronic records was larger than the number of self-reports because participants only assessed their first use, even if they took two doses at once, whilst the electronic device tracked and assessed all doses. To render the two types of records comparable, we averaged adjacent scores when participants inhaled multiple times. Participants reported taking an average of 2.3 doses per day (max: 4 doses/day, min: 0 doses/day) and always shook their inhalers before inhaling their medication. In 136 (70.8%) of instances, participants thought their inhaler usage was OK and in 47 (24.5%) instances they assessed it as Good. However, electronic records suggested that most patients over-assessed the quality of their inhaler technique (Fig 4b). According to the electronic device, 67 (34.9%) of uses were ‘Poor’, whilst the percentage of ‘OK’ used was reduced to 71 (37%) and the percentage of ‘Good’ also fell down to 25 (13%). Moreover, there were even 29 cases (15.1%) of ‘Very Poor’ use (compared with just one in the self-reports).

Figure 5 shows the statistics regarding ‘critical errors’. Improper inhalation technique has the second error rate (Figure 5 a). Using pMDIs requires slow inhalation to maximise lung deposition\(^{15,46}\), but participants A through D and F inhaled intensively, their average time of inhalation being only around 0.8 seconds. In this step, only participant E inhaled drugs properly most of the times. Figure 5 (b) shows the interval between the beginning of inhalation and inhaler actuation, which we called ‘actuation timing’. Triggering actuation
around the beginning of inhalation can be considered as correct timing, and even slightly delayed actuation benefits the increment of lung deposition\textsuperscript{15}, so we set correct timing as falling within -0.5s to 1.25s. Participant F was not assessed for this step, because the timing has little effect on the use of pMDIs with spacers. Other participants could choose the right timing to actuate their inhaler, and the average correct proportion of these reached 80.4\% (see Figure 5b). Breath-holding is also a critical step affecting drug delivery\textsuperscript{17}. Figure 5 (c) illustrates that only three of the six participants did or could infrequently hold their breath for over ten seconds, but the length of time varied widely. Importantly, only one of the six participants (Participant F) held their breath for the recommended 10 seconds. Exhalation before device actuation and inhalation of the medication is important for the efficacy of drug delivery. As Figure 5 (d) demonstrates, participants’ performance of this step varied widely and, notably, individuals were either very good or very poor at this step.

Many studies have not considered cough as a critical error. However, when we analysed the breath-holding step, we found that participant A coughed several times while inhaling (see Figure 5 [e]). This was because this participant could not hold their breath properly, forcing them to cough up the inhaled medication, thus reducing its efficacy.

The final aspect we considered was whether participants successfully actuated the inhaler, or not — clearly a critical aspect of the whole process. As Fig 5(f) demonstrates, this step was generally performed well, but nonetheless, four of the six participants sometimes had errors in this aspect.

Figure 6 shows the average of each correct step for each participant. Holding breath presented the highest error. Only 5.9\% of the instances held their breath for 10 seconds or more. Inhalation technique and completely breathing out was ranked second (correct: 12.0\%) and third (correct: 50.3\%) in technique errors. The correct percentage of other steps were all over 80\%.

**DISCUSSION**

Objectively tracking real-world pMDI usage and technique is difficult, because feedback is collected from patients’ self-reports or face-to-face visits with healthcare professionals. The former may overestimate patients’ inhaler use and furthermore may not allow the full range of steps critical to successful inhaler technique to be captured adequately. The latter presents an artificial situation may only occur infrequently and cannot capture everyday inhaler usage.

There has been considerable growth in inhaler devices which can capture some information about inhaler use. However, in many devices, this is limited to dose counting, which provides information about how often the inhaler was taken, but not how well.

In a small number of recent studies, electronic devices have been employed to monitor both patients’ inhaler usage and their technique\textsuperscript{60,69-72,82-84}. However, few studies have tracked technique for patients who use pMDIs. This is important, because these devices are particularly challenging to use, but are still very commonly prescribed, in part due to their low cost\textsuperscript{85-87}.

**Principal outcomes**

In this work, we first optimised an electronic device\textsuperscript{87}, based on preferences from patients who attended stakeholder meetings. Our device was, in consequence, more robust and more easily operated. Second, we tracked six participants’ pMDI usage and techniques over three
weeks, providing a sizeable body of data on inhaler use. Comparisons between paper and electronic records showed that participants—all of whom were very experienced and well-motivated users of their inhalers—over-estimated the quality of their inhaler technique. Additionally, participants also sometimes recorded the wrong number of doses taken.

The electronic device not only provided objective assessments but also quantified critical errors in inhaler technique. Having this type of data is important in assisting healthcare professionals and patients themselves to increase the efficacy of treatment by minimising critical errors [4,5,9]. The device may also provide a way to digitalise adherence to medications, which will enhance the volume and quality of information available to healthcare professionals and could also be of value to researchers, for example in medication or device trials.

**Strengths and weaknesses in relation to other studies**

pMDI inhalers have one of the largest market shares, but few studies have tracked pMDI users’ real-world technique. Our previous work has demonstrated the accuracy of our data capture algorithm [9]. Currently, each device costs approximately 25 pounds to produce and this could be even cheaper in the commercial manufacture, but also that a proper cost-effectiveness study is needed. The work reported here suggests that the device presents a feasible solution for tracking patients’ pMDI usage and technique.

Our study has some advantages over other comparable work. Holmes et al. [9] used the INCA recording device to assess DPI techniques; Taylor et al. [8] used the INCA attached to a pMDI to detect actuation and inhalation, utilising audio recorded from healthy volunteers and hospitalised patients. Additionally, Taylor et al. [8] introduced historical acoustic monitoring systems for inhalers to measure adherence. However, these studies did not include comparisons between subjective and objective assessments. Moreover, they were not as comprehensive as ours, because they used acoustic data to assess pMDI technique, but did not assess the importance of other non-acoustic errors. For example, compared with DPIs, a significant step for pMDIs is thoroughly shaking the inhaler, which makes drug particles spread evenly in the propellant [10,12].

In other pMDI studies [6,7,19], data were collected in the laboratory or were even simulated, neither of which can reflect real patterns for patients. By contrast, our study monitored patients’ inhaler use in real-world, everyday usage situations.

Our study had some limitations. The number of patients was small—although this was to some extent offset by the volume of data captured from each individual. Participants were recruited from a single geographic area and from self-selecting groups, which may reduce the generalisability of our findings. Although of scale to other comparable work, this was a preliminary study, which followed volunteer participants, all of whom were experienced inhaler users, over a relatively short period. Although our data are encouraging, future larger-scale studies are required. There is a lack of evidence explaining the relative contribution of each ‘critical error’. In the electronic assessment, we, therefore, assumed each ‘critical error’ was equally important to the final score. Intuitively, critical steps may contribute different outcomes to the delivery of the drug. This is an area which requires further work in the future.

**CONCLUSION**
pMDIs remain common in the treatment of respiratory conditions like asthma and COPD. This study has made a valuable contribution to the limited data available about how individuals use pMDIs in real-world situations. We have demonstrated that a novel electronic device has the potential to provide detailed information about pMDI usage and technique. A better understanding of not only how often, but how well individuals use pMDIs could assist healthcare professionals and patients themselves in improving the efficacy of inhaled treatment. W.X. takes responsibility for the content of the manuscript, including the data and analysis.

ACKNOWLEDGEMENTS

Author contributions

W.X. planned and structured the study, developed the system, performed data analysis and drafted the manuscript. P.G. and A.L.C. supervised the study, reviewed the results and revised the manuscript. J.H. organised stakeholder meetings, collected samples and revised the manuscript.

Financial/nonfinancial disclosures

None declared.

Role of sponsors

The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

Other contributions

We thank all the study participants, stakeholder meeting team and Greater Manchester AHSN for providing supports to make our study a success.

REFERENCES

1. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization [Internet]. Seattle, WA: IHME, University of Washington, 2018 [cited 2019 Feb 26]. 10.1002/[link-is-in-library]/article.20190226.1.g
11. Gulliford A. Most cost-effective treatments for lung disease identified. BMJ [Internet]. 2017 Oct 3;359:j4970. 10.1136/bmj.j4970


37. Asthma UK. How to use a pMDI inhaler: Asthma UK [cited 2019 Mar 6].


42. Burchfield TR, Venkatram S. Accelerometer-based human abdominal movement detection in wireless sensor
Table 1 Correct inhaler technique – pMDI
(1) Take the mouthpiece cap off
(2) Shake the inhaler well
(3) Keep the inhaler upright
(4) Breath out slowly
(5) Place mouthpiece between lips and over tongue
(6) Actuate the inhaler while breathing slowly and deeply and continue to slowly breath until lungs feel full
(7) Hold breath 10 seconds
(8) Exhalation

Table 2 Daily questionnaire template

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Shook the inhaler</th>
<th>Where did you use?</th>
<th>How good was your technique?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: 08/07 14:30</td>
<td>Yes</td>
<td>Indoor</td>
<td>Very good</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>Don’t remember</td>
<td></td>
<td>Poor</td>
</tr>
</tbody>
</table>

Table 3 Electronic records assessment standard

<table>
<thead>
<tr>
<th>Steps</th>
<th>Checklist</th>
<th>pMDI</th>
<th>pMDI with spacer</th>
<th>Score</th>
</tr>
</thead>
</table>
| (1)   | - Shake inhaler
       |        | - Shake inhaler  | 1     |
|       | - Hold inhaler upright |      |                  | 0     |
| (2)   | - Exhalation (empty lungs) | - Exhalation (empty lungs) | 1     |
| (3)   | - Actuation $t_a (-0.5 s \leq a \leq 1.25 s)$
       |        | - Inhalation $t_{in} (t_{in} \geq 2 s)$
       |        | Or - Breath times $\geq 5 s$ (Tidal breathing) | 1     |
| (4)   | - Hold breath $t_h (t_h \geq 10 s)$
       |        | - No cough
       |        | Or - No cough (Tidal breathing) | 1     |
| (5)   | - Actuation success | - Actuation success | 1     |

Table 4 Participants' daily records

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Age</th>
<th>Days</th>
<th>Device</th>
<th>Doses</th>
<th>Sharing (proportion)</th>
<th>Top self-assessment (proportion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Male</td>
<td></td>
<td>36 days</td>
<td>pMDI</td>
<td>19</td>
<td>Yes (100%)</td>
<td>Good (73.3%)</td>
</tr>
<tr>
<td>B</td>
<td>Female</td>
<td></td>
<td>22 days</td>
<td>pMDI</td>
<td>43</td>
<td>Yes (100%)</td>
<td>Good (64.3%)</td>
</tr>
<tr>
<td>C</td>
<td>Female</td>
<td>≥ 55</td>
<td>25 days</td>
<td>pMDI</td>
<td>95</td>
<td>Yes (100%)</td>
<td>Ok (66.7%)</td>
</tr>
<tr>
<td>D</td>
<td>Female</td>
<td></td>
<td>25 days</td>
<td>pMDI</td>
<td>42</td>
<td>Yes (100%)</td>
<td>Ok (100%)</td>
</tr>
<tr>
<td>E</td>
<td>Male</td>
<td></td>
<td>21 days</td>
<td>pMDI</td>
<td>45</td>
<td>Yes (100%)</td>
<td>Ok (100%)</td>
</tr>
<tr>
<td>F</td>
<td>Female</td>
<td></td>
<td>23 days</td>
<td>pMDI with spacer</td>
<td>79</td>
<td>Yes (100%)</td>
<td>NA</td>
</tr>
</tbody>
</table>
Figure 1 | Flow diagram of the electronic

Figure 2 | Experimental monitoring device
Figure 3 | The one of acoustic record from an individual

Figure 4 | Comparison of self-assessment between (a) patients’ and (b) electronic records
Figure 5 | (a) Comparison of length of inhalation time; (b) timing of actuation; (c) length of hold breath time; (d) percentage of exhalation before inhaled drug; (e) percentage of cough during holding breath; (f) percentage of successful actuation

Figure 6 | Average percentages of correct techniques