2021

Cough Monitoring Through Audio Analysis

Thomas Snella
Technological University Dublin, snellathomas59@hotmail.com

Recommended Citation
Snella, T. (2021). Cough Monitoring Through Audio Analysis. Technological University Dublin. DOI: 10.21427/3JC0-FR92

COUGH MONITORING THROUGH AUDIO ANALYSIS

A Dissertation Presented

by

THOMAS SNELLA

Submitted to the Graduate Research School of the

Technological University Dublin in partial fulfillment

of the requirements for the degree of

MASTER OF PHILOSOPHY

August 2021

Supervisors

Dr. David Dorran   Prof. Ivana Dusparic   Dr Damon Berry

ABSTRACT

The detection of cough events in audio recordings requires the analysis of a significant amount of data as cough is typically monitored continuously over several hours to capture naturally occurring cough events. The recorded data is mostly composed of undesired sound events such as silence, background noise, and speech. To reduce computational costs and to address the ethical concerns raised from the collection of audio data in public environments, the data requires pre-processing prior to any further analysis.

Current cough detection algorithms typically use pre-processing methods to remove undesired audio segments from the collected data but do not preserve the privacy of individuals being recorded while monitoring respiratory events. This study reveals the need for an automatic pre-processing method that removes sensitive data from the recording prior to any further analysis to ensure privacy preservation of individuals.

Specific characteristics of cough sounds can be used to discard sensitive data from audio recordings at a pre-processing stage, improving privacy preservation, and decreasing ethical concerns when dealing with cough monitoring through audio analysis.

We propose a pre-processing algorithm that increases privacy preservation and significantly decreases the amount of data to be analysed, by separating cough segments from other non-cough segments, including speech, in audio recordings. Our method verifies the presence of signal energy in both lower and higher frequency regions and discards segments whose energy concentrates only on one of them. The method is iteratively applied on the same data to increase the percentage of data reduction and privacy preservation.

We evaluated the performance of our algorithm using several hours of audio recordings with manually pre-annotated cough and speech events. Our results showed that 5 iterations of the proposed method can discard up to 88.94% of the speech content present in the recordings, allowing for a strong privacy preservation while considerably reducing the amount of data to be further analysed by 91.79%.

The data reduction and privacy preservation achievements of the proposed pre-processing algorithm offers the possibility to use larger datasets captured in public environments and would beneficiate all cough detection algorithms by preserving the privacy of subjects and by-stander conversations recorded during cough monitoring.

DECLARATION

I certify that this thesis which I now submit for examination for the award of Master of Philosophy, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. This thesis was prepared according to the regulations for graduate study by research of the Technological University Dublin and has not been submitted in whole or in part for another award in any other third level institution. The work reported on in this thesis conforms to the principles and requirements of the TU Dublin's guidelines for ethics in research. TU Dublin has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature Thomas Snella Date 17/12/2021
Candidate

ACKNOWLEDGMENTS

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (D-REAL) under Grant No. 18/CRT/6224.

The completion of this study could not have been possible without the support and guidance of Dr. David Dorran, the main supervisor.

I would also like to thank:

Prof. Ivana Dusparic and Dr. Damon Berry, for their availability throughout the completion of the study.

Dr. Jane Courtney, for her help in creating data for the study.

My family, for their trust and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 OVERVIEW AND MOTIVATION

Cough is a common symptom observed in many diseases for which patients seek medical attention [1, 2]. According to Schappert [3], the most frequently mentioned symptoms during medical visits are related to the respiratory system, and cough is the most frequently mentioned reason having to do with illness or injury. Indeed, Cough is a symptom of over a hundred diseases [4, 5] including chronic bronchitis, acute tracheitis, pneumonia, lung abscess, tuberculosis, lung cancer, and pulmonary oedema [2, 6], and there are more than 50 medical complications associated with coughing [7].

The assessment of cough has greatly contributed in the field of cough pharmacology [8]. The most widely used approaches to assessing cough include methods for measuring cough-specific quality of life, subjective severity, cough frequency, intensity, and sensitivity of the underlying cough reflex [9-13]. However, subjective reporting of cough frequency and intensity is not reliable [14-17], and the measurement of the frequency at which cough events are occurring is often preferred to track respiratory ailments [4, 8, 10, 18-23]. Cough frequency is increasingly recognized as a measurable parameter of respiratory disease [24], and automatic cough detectors must be compared to the gold standard: manual counting of cough sounds by an expert [18, 25-27].

Over the last three decades, many automatic cough detection algorithms were developed because the manual analysis of cough sounds by a listener is time consuming and impractical in ambulatory conditions [8, 25, 28, 29].

Most of the current cough detection algorithms use audio analysis to monitor cough frequency; however, other techniques have been tested and compared, such as electrocardiography, electromyography, nasal thermocouple sensors, chest belts, airflow signals, and accelerometry [18, 23, 30-36].

When dealing with cough monitoring through audio analysis, it is important to understand what a cough is and what differentiates it from other sounds. A cough is composed of three distinct phases which allow in defending the lower airways: an inspiratory phase followed by a forced expiratory effort initially against a closed glottis (compressive phase), followed by active glottal opening and rapid expiratory flow [8, 37-40].

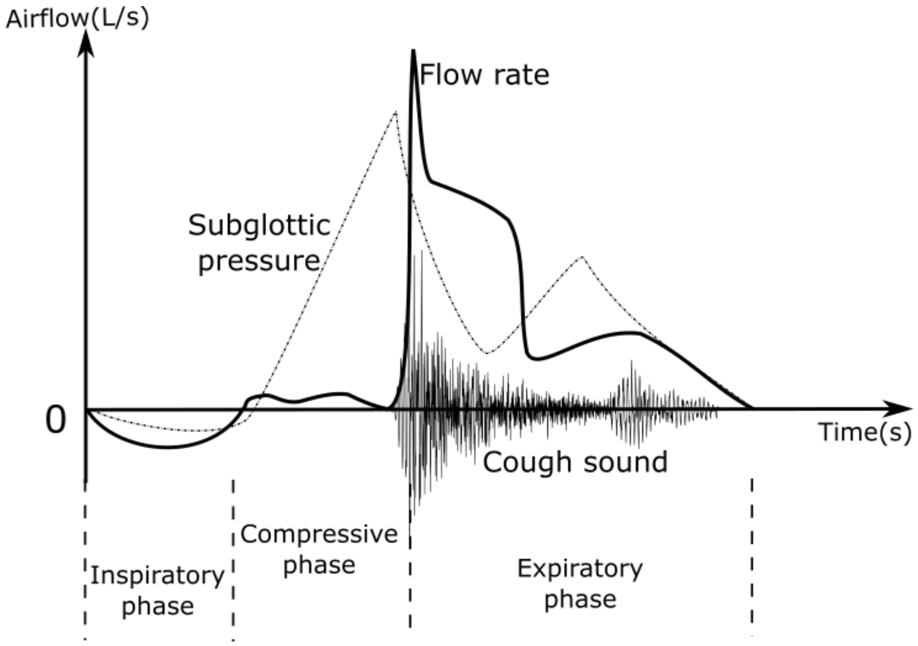


Figure 1: From McCool [37], Schematic diagram depicting changes in flow and subglottic pressure during the inspiratory, compressive, and expiratory phases of cough.

It is the expulsive phase that creates the sound of a cough (Figure 2). The sound is also composed of three phases: an explosive phase, an intermediate phase and a voiced phase [8, 18, 41]. The cough sound originates in a sudden air expulsion from the airways and is so characteristic that it is easily identified from other sounds by the human ear [4]. However, it becomes harder for a machine to identify cough sounds from similar sounds such as speech, laughing, sneezing, throat clearing and other ambient sounds [42].

Other airway defensive reflexes, such as the expiration reflex, can be mistaken for coughs but have different properties. The expiration reflex will prevent aspiration of material into the lungs while cough will clear the airways from debris. Fontana [8] explains that the inspiratory phase of the cough is what differentiates a cough from an expiration reflex.

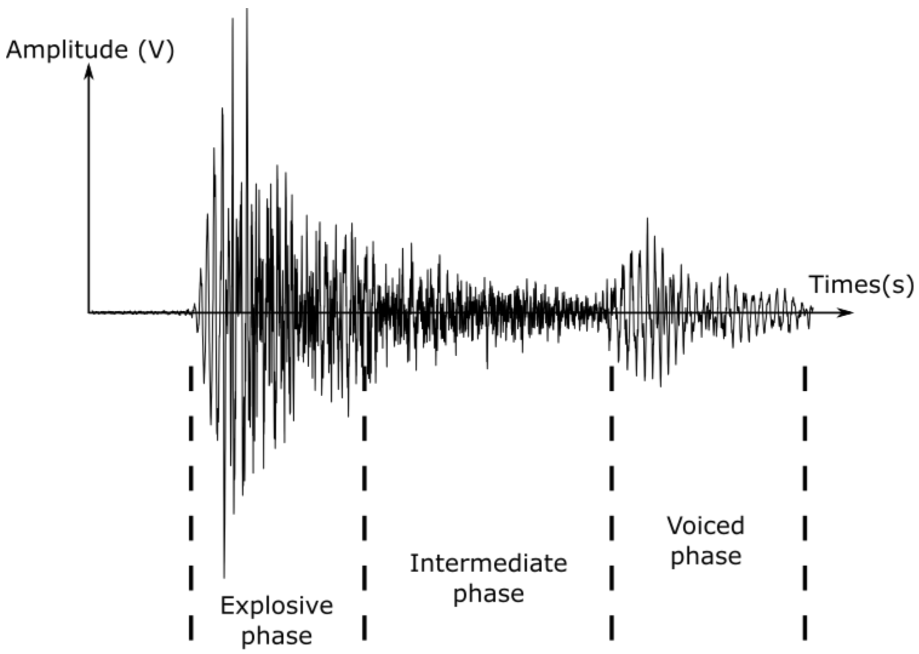


Figure 2: From Hall et al. [18], The component phases of the cough sound: opening of the vocal cords (first phase), air flow through the open larynx (second phase), and re-apposition of the cords (third, voiced phase—not always present).

Cough monitoring methods can be divided into three main categories: Automatic cough detection and segmentation, automatic classification of coughs that are already detected, and diagnosis of an illness based on the cough sound and type [43].

The typical workflow for cough detection algorithms is composed of the following three steps [43, 44]:

- Sound event detection, a pre-processing stage to remove silence within the signal.
- Feature extraction, the most useful features are used as inputs for a model classifier.
- Classification, sound events are classified into cough and non-cough events by a trained classifier

The pre-processing stage of cough detection algorithms is typically used to reduce the amount of data to further be analysed. More rarely, pre-processing has been used in attempts to discard sensitive data from audio recordings [45-48].

1.2 Thesis Statement

Current cough detection algorithms do not preserve the privacy of individuals being recorded while monitoring respiratory events. There is a need for an automatic pre-processing method that removes sensitive data from the recording prior to any further analysis to ensure privacy preservation of individuals. This motivation introduces our thesis statement:

Specific characteristics of cough sounds can be used to discard sensitive data from audio recordings at a pre-processing stage, improving privacy preservation, and decreasing ethical concerns when dealing with cough monitoring through audio analysis.

1.3 Aims And Objectives Of The Present Work

The area of cough detection through audio analysis could beneficiate from improvements related to privacy preservation, the identification of new cough event features, and the limitation in the algorithms' performance evaluation.

1.3.1 PRIVACY PRESERVATION

Ethical concerns raised from the collection of audio data in public environments need to be addressed by researchers in their study. From the research conducted, only a few researchers mentioned methods used to preserve privacy. These methods typically involve the removal of undesired audio segments from the collected data. The data is modified using subsampling, on-event recording, or alteration of audio; however, there is typically a loss of audio quality or number of cough events in the remaining data. A pre-processing algorithm that can remove undesired audio events, including speech, would beneficiate all cough detection algorithms by preserving the privacy of subjects and by-stander conversations recorded during cough monitoring.

1.3.2 IDENTIFICATION OF NEW COUGH EVENT FEATURES

From the research conducted, it appears that researchers have a tendency to use the "state-of-the-art" feature extraction method at the time of their research. It started with methods like time-domain analysis and methods similar to the ones used for speech recognition. More recently, the progress made with Neural Networks made it a commonly used method for cough feature extraction.

These methods have proven to be effective, however the number of false positive and false negative reported in studies shows that there is room for improvement and features that represent cough events more accurately need to be found.

Researchers have been studying the combination of different features to create more accurate cough models. However, further research is required to find the ideal set of features with aim to find the state-of-the-art method for cough monitoring through audio analysis.

1.3.3 LARGE DATASET FOR A CONSTANT EVALUATION ACROSS ALL COUGH DETECTION ALGORITHM.

Our study shows that the datasets used for the evaluation of existing algorithms differ from one study to another. The quality, size, and type of data composing the datasets have a significant impact on the evaluation of the algorithm. This diversity in the evaluation of existing algorithm does not allow for the identification of the state-of-the-art cough detection algorithm. Therefore, a new evaluation method with constant data and metrics is needed. We believe a large dataset for couch detection algorithm evaluation purpose should be created and shared publicly with the scientific community. This dataset should represent all recording conditions and should be composed of audio segments recorded with a high diversity in the type of microphones, study subjects, type of coughs events, type of non-cough events, quality, and background noises.

Furthermore, the evaluation of cough detection algorithms needs to be done using universal metrics and evaluation criteria. The state-of-the-art algorithm for cough detection can then be identified from a fair evaluation of all algorithms based on results obtained using a unique dataset and constant evaluation criteria.

1.4 OBJECTIVES OF PRESENT WORK

From the identified possible improvements in the field of cough detection through audio analysis, we focused on the development of a pre-processing technique that can considerably reduce the amount of data to analyse, while preserving the privacy of the persons being recorded. We think that contributing to the scientific community with a privacy preservation algorithm would benefit all existing and future cough detection algorithms as it could be used at a preprocessing stage, not requiring any alteration of existing algorithms.

1.5 RESEARCH CONTRIBUTIONS

The main contributions of this thesis are:

- **Privacy preservation pre-processing algorithm for cough detection**: This thesis contains a pre-processing algorithm that detect and preserves cough events while removing enough speech to make the resulting audio file unintelligible. This new pre-processing method contributes to the field of cough detection through audio analysis where the collection of audio data in public environments raises ethical concerns.

- **Data reduction pre-processing algorithm for cough detection**: By removing sensitive data from the audio recordings in addition to other non-cough sound events, the proposed algorithm achieves high performance at reducing the amount of data to be analysed in the next stages of the cough detection algorithm. All cough detection algorithms can benefit from this new algorithm, decreasing their required computational cost.

This thesis will provide, in chapter 2, an overview of existing cough detection algorithms from the recording of audio data and its implications to the classification of sound events into cough and non-cough events. In chapter 3, we will discuss the state-of-the-art of existing cough detection algorithms by comparing their published results and we will cover the limitation of existing algorithm. This will lead to a detailed explanation of the proposed new privacy preserving pre-processing algorithm in chapter 4. We will conclude with potential future work in the area of cough monitoring with audio analysis in chapter 5.

Chapter 2

LITERATURE REVIEW

Over the last decades, a small portion of pulmonary research has been focused on the systematic detection of cough sounds and considerable progress has been made in the development of cough assessment tools [45]. This chapter gives an overview of related work and background information, covering the typical cough monitoring steps when dealing with cough monitoring through audio analysis, from cough recordings to cough sounds classification.

2.1 Cough Recordings And Ethical Implications

The collection of audio data raises audio privacy concerns since private conversations can be recorded alongside sounds of interest. Many papers cover the topic of preserving privacy during audio data collection in a way that speech is altered or eliminated from the original recording before analysis [46-48]. However, the captured data needs to be stored prior to using these techniques and sensitive data can be extracted by the researcher. Ethical principles must be respected to ensure the privacy preservation of research subjects.

2.1.1 Ethical Considerations

Smith [49] details principles for research ethics which relate to audio data collection. She explains that during the consent process, the subject must be informed with clarity on the purpose of the research, the rights to withdraw from the research, the limits of confidentiality, and relevant risks and benefits. Subjects must be given information about how the data collected will be used and shared with the community.

Ethics principles can easily be applied when collecting audio data in a controlled environment since only research subjects are being recorded. Consent forms can be signed by each subject and the data collected can be used and shared without causing any harm. One negative consequence of performing audio data collection in a controlled environment is the lack of ambient background noise. Ideally, audio data would be collected in public spaces where the recording of coughs of numerous different subjects would allow for better cough models and increase the performance of the algorithm. However, it becomes impossible to obtain the consent of each person being recorded during the data collection.

To overcome this issue, real-time pre-processing could remove sensitive data from a recording as it is being recorded. Audio would then be altered before being available for any further analysis, respecting privacy and limiting the need for consent. However, pre-processing the data brings another issue as cough events can be deleted alongside the undesired sensitive data.

2.1.2 PRIVACY PRESERVATION

Larson et al. [47] attempt to meet all the requirements for objective cough monitoring as outlined by the medical community. These goals include accuracy, low false positives, mobility, compactness, privacy preservation, unobtrusiveness. Their system uses Principal Component Analysis (PCA) of audio spectrogram for prevention of speech reconstruction. By reducing dimensions of the data with PCA, sound becomes unintelligible and only sound events classified as cough are reconstructed for analysis by a medical professional. The fidelity of reconstruction is highly dependent of the number of components used in the PCA and they found that 25 components produce a good fidelity cough sound while simultaneously disguising most of the spoken words. However, the fidelity of cough sound reconstruction is limited in [47] when a higher privacy preservation is desired, inducing a forced choice between cough quality and privacy.

Sub-sampling offers an automatic privacy protection and subjects being recorded do not need to worry about identification or sensitive speech being compromised. Sound sub-sampling was used by Kumar et al. [48] to make it difficult to retrieve speech information. Their technique consists in recording for a short period of time (1 second) every few seconds allowing for short sounds to be extracted. Cough sounds, having an average duration of 500 ms, can still be recorded while only part of the speech will be recorded making it difficult to be reconstructed. However, the process of sub-sampling original recordings used in [48] may also remove potential cough events from the recordings and therefore corrupt the data when the purpose is to accurately monitor cough event occurrences for one particular patient.

Nguyen and Luo [50] introduce a non-intrusive cough detection technique using a smartwatch. They assume that when someone coughs, there is a prior reflex of moving rapidly the hand towards the mouth. They detect similar movements using the smartwatch's accelerometer allowing the sound recorder to only be activated at this specific moment and for a few seconds. This technique used by Nguyen and Luo increases privacy preservation while considerably reducing the non-relevant data collected. Their technique is promising however there is significant reliance on the fact that the subject's mouth will always be covered while coughing and always using the same hand. Many cough events can still be missed and undesired sound events will potentially be recorded for any quick movement of the hand when wearing the smartwatch. The technique proposed in [50] is furthermore flawed with the recent COVID-19 pandemic and the instruction of the WHO to avoid covering the mouth with the hand while coughing.

2.2 PRE-PROCESSING OF AUDIO SIGNAL

Pre-processing techniques are usually used to reduce the amount of data in a recording prior to conducting any analysis.

In previous literature, most of the data removed is of low intensity, such as silence, or of very high frequency, such as background noise. Ideally, most of the speech present in a recording should be removed at the pre-processing stage for privacy preservation while all cough events should be left intact. However, speech is often considered as a sound event similar to the cough sound; therefore, most pre-processing techniques do not remove the speech content in the recordings prior to analysis to ensure that cough events do not get deleted alongside.

Removing silent segments from the audio recording by using manual editing, as in [51-52], ensures the preservation of cough events; however, manual editing is laborious and the task is typically automated.

Shi et al. [44] discuss simple threshold methods that use time-domain features to remove silence from the data, such as the energy entropy method used in [28,44,53-57] consists of framing the signal, calculating each frame energy and keeping only frames whose energy is above a defined threshold. Matos et al. [58] apply a dynamic energy threshold to the signal before their HMM classification, the threshold is set at 5 dB above the local neighborhood signal level and allow to eliminate silent segments of the recording. To reduce the chances of inadvertently discarding cough events from the recordings, they keep a 10-second segment of data for each sound event detected.

Energy threshold techniques are the most commonly used but a multitude of other methods have been tested in previous literature. [15,43,59] use the standard deviation of the signal to highlight the components of the data that contain the most variance, such that only these components are used for further analysis. Larson et al. [55] developed an event detection logic that only triggers when there is a rapid increase in acoustic energy relative to the noise floor. Ye et al. [60] employ probabilistic latent component analysis (PLCA) to perform time-varying noise separation focusing on some frequency bands, enhancing reliability of audio spectra.

These methods have proven to be effective in the removal of silent segments while keeping most of the coughs present in the data. However, other sounds are also kept after the pre-processing stage, raising audio privacy concerns, and limiting the effectiveness of the data reduction process.

The importance of privacy preservation is being acknowledged more frequently, making it essential to use the pre-processing stage of the algorithm not only for data reduction but also to remove speech from the audio recordings.

2.3 FEATURE EXTRACTION

The identification of unique cough features is typically key to automatic cough detection and, over the last decade, numerous approaches have been adopted. Drugman et al. [61] explain that most of the features used for cough detection fall under three categories: features describing the spectral content, measure of noise, and prosody-related features.

This section will cover the most widely used feature extraction techniques for each category and some of the less common techniques. Their performance will be compared in chapter 3.

2.3.1 FREQUENCY BASED FEATURES

The frequency-domain has been widely exploited for cough detection. Previous studies have shown that while the cough sound can vary significantly from one subject to another, it shares similar features in the frequency-domain.

For example, it has been found that cough energy scatters through a wider frequency range than speech [62,63]. While speech energy concentrates on the low frequency region below 2000 Hz, most of a cough energy lies within the mid-low frequency range (300 Hz to 8 kHz) [50,62-64]. In further research, Kosasih et al. [65] extract important cough features from frequencies up to 90 kHz. However, this requires capturing the data at a very high sampling rate, therefore, increasing greatly the computation cost.

A variety of methods based on frequency components analysis of the signal allow for the extraction of cough features. In early work, Barry et al. [15] use Linear Predictive Coding (LPC) to get their features from the signal spectral envelope. The features obtained from the Fourier transform illustration of the logarithmic magnitude spectrum of the LPC are cepstral coefficients and contain useful information to identify cough sounds [66]. Similarly, Mel Frequency Cepstral Coefficients (MFCCs) are commonly used features for representing spectral pattern in speech recognition and several attempts for cough detection using MFCCs have been made [29,42-44,46,50-52,54-57,59,62,67-74]. This approach mimics the human auditory system by using filters with frequency bands equally spaced on the Mel scale. Matos et al. [42] created the Leicester Cough Monitor (LCM), a detection algorithm based on statistical models of the time-spectral characteristics of cough sounds. They use the MFCC parameterisation to describe the properties of each frame of their data in the cepstral domain keeping the 13 first coefficients and their first and second order derivatives to form the feature vector. To overcome the wide range of variances in the different order MFCCs, they apply cepstral liftering to detect quieter cough sounds. In a similar manner, Tracey et al. [54] use MFCCs as the primary features for acoustic analysis after detecting potential cough events from rapid increase in acoustic energy relative to the noise floor.

Many other extraction methods originate in the adaptation of the MFCCs [60,63,71]. For example, Ye et al. [60] follow the Mel filter bank idea by designing a uniformly spaced triangle filter bank to describe sound events. Usually, the bandwidth of the filters in the Mel filter bank gradually increases, reducing the number of filters in the higher frequency region and emphasising the lower frequency region. Instead, the triangle filter bank of Ye et al. captures "richer temporal-spectral" features by keeping a constant bandwidth. In addition, they perform eigen-decomposition based on filtered audio spectrogram to further characterise significant patterns in sound events.

Liu et al. [75] evaluate and compare Gammatone Cepstral Coefficients (GTCCs) with MFCCs. In this approach, an Equivalent Rectangular Bandwidth (ERB) scale is used instead of the MFCCs Mel scale, resulting in a smoother filter bank and a more accurate model of the human auditory system than with the triangle filters employed in MFCCs. However, You et al. [63] mention that, like noise, cough energy scatters across the entire spectral area while MFCCs and GTCCs emphasise the lower frequency region and may be unsuitable for cough detection. Similarly to [60], You et al. [63] propose a subband technique that emphasises the local frequency band from the full-band spectrum. The subbands are generated based on GTCCs with a smooth filter bank and features can be extracted from each subband signal by "any kind of common feature extraction method".

Shin et al. [62] also found MFCCs unsuitable for cough detection and introduced Energy Cepstral Coefficients (ECC) with a filter bank based on ERB to obtain the spectral pattern of a sound signal.

Miranda et al. [71] compare MFCCs with Short-time Fourier Transform (STFT, previously used in [56,76]) and Mel-scaled Filter Banks (MFB) by evaluating them with deep architecture networks. Their research shows that MFCCs first and second derivatives do not improve the performance and can be omitted, reducing the number of features. Furthermore, Miranda et al. [71] mention that "less engineered" features obtained from STFT and MFB provide better cough detection accuracies, showing that MFCCs may not be the ideal feature extraction method with deep architectures. Nevertheless, a feature extraction method can be combined with another to increase system performance. Nguyen and Luo [50] complement MFCCs features with Zero Crossing Rate (ZCR) and Chroma Feature Analysis. ZCR can be used as a measure of the rate of change in the frequency content and can be useful to either detect noise or cough as they are non-stationary sounds [43,46,50-52,56,73,77].

Cough is a non-stationary signal and varies significantly in the time-domain and frequency-domain [65]. This characteristic can be used to discard non-cough sounds from a signal as previously done by Barry et al. [15] with their Hull Automatic Cough Counter (HACC). They calculate the standard deviation of the signal to identify potential cough events, reducing the amount of data to be analysed. Kosasih et al. [65] explain that, ideally, a method that captures the time and frequency changes simultaneously would be more suited for cough analysis and techniques such as STFT or wavelet could be used for this purpose. Their study argues that more detail can be extracted from wavelet representation of the signal compared to time-domain or frequency-domain alone.

The wavelet representation is used to extract cough features in several papers [65,78-80]. Dat et al. [79] apply wavelet to characterise non-stationary sound event spectrograms. They introduce the Spectrogram Image Wavelet Representation (SIWR) to extract useful information from the 2D time-frequency representation of the sound signal. This idea comes from the fact that humans can easily locate the characteristic elements in a spectrogram, and it is possible to visually see the sound event among background noise. Following the same approach of casting the cough detection task as a visual recognition task, Convolutional Neural Networks (CNN) have been used to identify cough features from two dimensional sprectro-temporal images [44,53,68,81-83]. Amoh and Odame [53] mention that one issue with this technique is the need for pre-segmentation to obtain fixed size input images and post-processing is also required to align the predictions with the audio signal. Their second approach uses Recurrent Neural Networks (RNN) with sequence-to-sequence labeling for capturing temporal and spectral dependencies between initial burst, middle phase, and final burst of a cough. Recent improvements of RNN make it more efficient for speech recognition and machine translation like tasks [53].

2.3.2 OTHER FEATURES

In some research, the frequency-domain is not explicitly used to extract cough features. Murata et al. [84] investigate the characteristics of cough sounds acoustically with time-domain analysis. Monge-Alvarez et al. [85] propose the use of local Hu moments as a robust feature set for automatic cough detection in smartphone-acquired audio signals.

However, cough features are typically extracted from the frequency-domain and complemented with other relevant features to improve cough detection algorithms, such as the Harmonic to Noise Ratio (HNR) used to compare a signal to the level of background noise, the Cepstral Peak Prominence which is correlated with the amount of breath sound in the voice, or the Spectral Flatness which measure the noisiness of a spectrum [61].

Nguyen and Luo [50] use Zero Crossing Rate (ZCR) to complement their set of MFCC features. The ZCR is used to detect higher frequency content in a signal such as noise, speech or cough events. They also use the Chroma Feature Analysis (CFA) to project the signal onto 12 distinct pitch classes.

Drugman et al. [61] mention in a study that HNR appears in the features that convey the greatest relative intrinsic information when compared with other features in a set of 105 features.

Prosody-related features include the pitch, loudness, timber, and length of sounds in a signal. These features are usually used in the music field for tasks such as note recognition; However, they can be applied to cough detection algorithm to get a broader set of features and potentially increase the features combined relevance.

2.3.3 NUMBER OF FEATURES

Researchers have been studying the combination of different features to create more accurate cough models. It is not uncommon that the latest algorithms use hundreds of features in their attempt for the creation of robust cough detection algorithms [52,56,57,61,67,73,74,86,87]. For example, Brown et al. [86] extract a total of 733 features before using PCA to reduce the feature vector.

Other researchers studied the usefulness of gathering such a large number of features. Ye et al. [88] adopt subspace analysis to describe acoustic signals as it has a lower feature dimension, making processing acoustic subspace more efficient compared to dealing with raw feature vectors. Miranda et al. [71] performed experiments to evaluate if MFCC derivatives provide improvements to the algorithm performance as commonly assumed.

From [71], it appears that using first and second derivatives of MFCCs as features does not improve classification performance in most cases. Furthermore, in [61], Drugman et al. explain that each feature conveys information, but this information is only of interest if it is not already conveyed by another feature. This redundancy between features and their relative joint information is the key to the ideal number of features. An excessively large number of features can greatly affect the computation cost of the algorithm while performing as effectively as an algorithm using less features but with high relative joint information. Drugman et al. [61] calculated redundancy and relative joint information for 105 features in their cough detection algorithm. It appeared that only 20 features were enough to convey most of the information and that increasing the number of features was not worth the computation cost.

It is however disputable that the 105 features selected had too much redundancy and that with a larger number of features, the minimum number of features useful in the conveying of information would also increase. Further research is required to find the ideal set of features for the field of cough monitoring through audio analysis.

2.4 SOUND EVENTS CLASSIFICATION

The classification stage of cough detection algorithms is typically performed with the use of machine learning techniques. This section covers the most common machine learning methods implemented for cough monitoring and the results of those algorithms are compared in chapter 3.

2.4.1 COMMON CLASSIFICATION ALGORITHMS

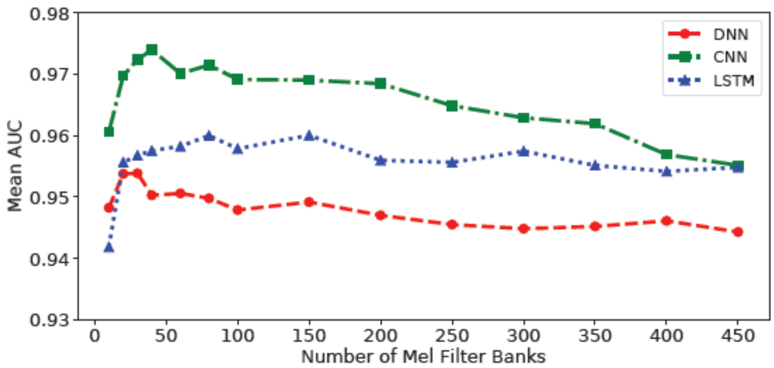
Hidden Markov Models are often used for cough detection as they can characterise the spectral properties of a time-varying pattern [42,44,58,62,69,70,72,89]. Matos et al. [42] propose an HMM recognition algorithm that follows a keyword-spotting approach. They created cough and filler models to train the HMM. The recognition process works by finding the sequence of models that fits an unknown input frame sequence with the highest probability.

Dat et al. [79] model the sound spectrogram image in wavelet representations using Generalised Gaussian Distribution (GGD) modelling. For classification algorithm optimisation, they use a Generalized Gaussian Distribution Kullback-Leiber kernel Support Vector Machines (SVM) to embed the given probabilistic distance into a quadratic programming machine. SVM is another common approach for cough sound classification [44,50,54,63,75,79,86]. You et al. [63] use SVM as classifier in their ensemble approach. This ensemble method aims to combine the outputs from multiple classifiers for a better accuracy. However, they mention that HMM may be more effective with larger dataset allowing more complex models.

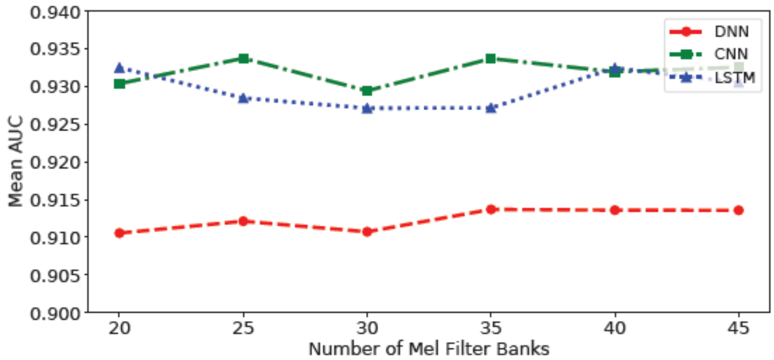
With the development of Neural Networks, Artificial/Probabilistic Neural Network [15,50,52,59,62,67,73,74,90], Convolutional Neural Network [71,83], and Deep Neural Network [53,57,69,71] became the most popular methods for cough events classification.

The HACC developed by Barry et al. [15] uses Probabilistic Neural Network (PNN) after calculating characteristic spectral coefficients of sound events. Their PNN uses a Bayesian classifier approach and is trained to recognise the feature vectors of cough and non-cough models with the aim to correctly classify future sound events.

Shin et al. [62] use a two-stage classification algorithm. In the first step, they classify sound events into noise or cough/speech with an Artificial Neural Network (ANN). Then, in the second step, the output of the ANN is combined with a filtered envelope of the signal to form the input sequence for the HMM that deals with the temporal variation of the sound signal.



(a) MFB



(b) MFCC

Figure 3: Cough classification performance as a function of the filter bank dimension for (a) MFB features and (b) MFCC features. From Miranda et al. [71].

Miranda et al. [71] make a comparison of three different type of neural network for cough detection: a convolutional neural network, a Deep Neural Network (DNN), and a long-short term model (LSTM) which is an artificial Recurrent Neural Network (RNN) architecture. Figure 3 shows the mean Area Under the ROC Curve (AUC) for each type of neural network, representing the dependence of performance on the number of filters used during feature extraction stage. The AUC provides an aggregate measure of performance and its score ranges from 0 to 1, where '0' means that an algorithm's predictions are 100% wrong and '1' means that the predictions are 100% correct. Their research shows that the CNN, having the highest mean AUC, performs slightly better than the DNN and LSTM for cough detection.

2.4.2 ALTERNATIVE CLASSIFICATION ALGORITHMS

In some cases, other methods were selected by researchers after being compared with the more traditional techniques. Larson E. et al. [47] trained a random forest (RF) classifier for the classification of cough sound events. They chose a RF classifier as it is less sensitive to parameter variation than SVM and NN while reaching equivalent performances. Similarly, Larson S. et al. [55] selected a Sequential Minimal Optimisation (SMO) approach after comparing the results with NN and SVM approaches. This choice was made as the performance of the various methods was similar, while the SMO approach is easier to implement.

Ye et al. [60] conduct multi-class sound classification through exploiting class conditional distributions based on extracted acoustic subspaces. They use Kernel Fisher Discriminant Analysis to map the data into kernel discriminant feature space for classification.

Nguyen and Luo [50] propose a confidence level on prediction with a prediction set output instead of a simple prediction. Each sound event is compared to a bank of models and the prediction set is built as a function of the probability for a new sound event to belong to each model class. If the probability of one sound event belonging to one specific model is higher than the defined threshold, the prediction is returned with a single label. Otherwise, a set of prediction is returned, requiring manual annotation of the sets or a second stage classification. This method reduces the number of false positive in the classification process; however, it requires a second stage classification to return only single label predictions.

2.4.3 Cough Type Classification

This work focuses on the study of existing cough detection algorithms; however, it is important to note there are many machine learning algorithms that attempt to classify a detected cough event as a function of its type. The typical goal of these algorithms is to detect pulmonary diseases at an early stage. This requires cough events to be accurately detected and segmented prior to type classification analysis. The most common cough type classifications differentiate between wet and dry coughs [91-93], recognise spontaneous from voluntary coughs [94], or more recently detect COVID-19 characteristics [68].

Chapter 3

DISCUSSION OF CURRENT STATE-OF-THE-ART

3.1 A Performance Comparison

In this chapter, we make an attempt to compare existing cough detection algorithms with the aim to determine the state-of-the-art. However, it is important to note that the comparison of different automatic cough detection methods is significantly limited as studies use different datasets with main differences in the following: recording conditions, type and position of microphones, study subjects and types of non-cough sounds included in the recording [18].

Furthermore, there is no universally defined unit of cough [95] and it is unclear if the expiration reflex (which sounds like a cough without the inspiration phase) should be counted as a cough sound in cough detection algorithms [96].

There are several metrics to quantify cough and algorithms performance [55]. Cough can be quantified in coughs/hour or in terms of cough episodes or epochs. The definition of a cough epoch can also differ from one study to another [95,97].

It is also argued that the detection of cough epochs is as clinically meaningful as the detection of single cough events [41,95,97]; the performance evaluation is however affected. Teyhouee et al. [89] shows in their results a difference of up to 5% and 10% in sensitivity and specificity respectively, when identifying single cough events versus cough epochs with their algorithm.

Researchers often evaluate their algorithms in terms of specificity, sensitivity, and accuracy [44]:

$$Specificity = \frac{TN}{(TN+FP)} \quad (1)$$

$$Sensitivity = \frac{TP}{(FN+TP)} \quad (2)$$

$$Accuracy = \frac{TP+TN}{(TP+FP+TN+FN)} \quad (3)$$

where:

TP = True positive: Coughs correctly identified as coughs,
FP = False positive: Non-coughs incorrectly identified as coughs,
TN = True negative: Non-coughs correctly identified as non-coughs,
FN = False negative: Coughs incorrectly identified as non-coughs,

Once again, those evaluation criteria are calculated differently in the cough literature. For example, Matos et al. [42,58] use the "Birring specificity" metric to calculate their true negative number. They first detect sound events, then report a classification stage specificity by sorting their detected events into TP, FP, TN, and FN. It is argued that this method does not reflect the performance of the overall system [55] [6] . Vizel et al. [6] calculate the number of true negatives as a function of the number of 1-second intervals in which no cough is detected in both the automatic detection and manual cough annotation.

The variance in the definition of evaluation terms and calculation significantly impact the consistency of performance results in cough studies.

From the study undertaken, it is clear that universal metrics are needed to find the state-of-the-art cough detection algorithm. In recent research, Bilen et al. [98] propose a robust evaluation technique of sound event detection by re-defining TPs and FPs through the combination of several criteria. Their method can be adapted to varying needs, including cough detection, by adjustment of evaluation parameters.

Table 1: Performance of existing cough detection algorithms in terms of sensitivity and specificity

Author	Date	Feature Extraction	Classification	Sensitivity	Specificity
Barry et al.	2006	PCA	NN	0.8	0.96
Murata et al.	2006	Time-domain Anal.	Discriminant Function	0.902	0.965
Matos et al.	2006	MFCC	HMM	0.82	NC
Knocikova et al.	2008	Wavelet	Discriminant Function	0.85-0.9	NC
Shin et al.	2009	MFCC	NN/HMM	0.913	0.953
Vizel et al.	2010	Time-freq Domain	Pattern Matching	0.96	0.94
Larson E et al.	2011	PCA	RF	0.92	0.995
Tracey et al.	2011	MFCC	SVM	0.81	NC
Drugman et al.	2011	105 Handcrated Feat.	SVM	0.819	0.996
Larson S et al.	2012	MFCC	SMO	0.755	0.996
Drugman et al.	2012	222 Handcrated Feat.	NN	0.947	0.95
Martinek et al.	2013	MFCC	NN	0.86	0.91
Swankar et al.	2013	201 Handcrated Feat.	NN	0.934	0.945
Liu et al.	2014	MFCC	DNN/HMM	0.901	0.866
Sterling et al.	2014	MFCC	HMM	0.782	NC
Liu et al.	2015	MFCC	HMM/GMM	0.836	0.909
Ferdousi et al.	2015	MFCC/ZCR ...	NN/SVM/Bayesian	0.875	0.909
Amrulloh et al.	2015	MFCC/ZCR ...	NN	0.93	0.98
Kosasih et al.	2015	Wavelet	Logistic Regression	0.94	0.88
Amoh et al.	2016	CNN/RNN	CNN/RNN	0.877	0.927
Pramono et al.	2016	MFCC/ZCR	Logistic Regression	0.923	0.9
Liaqat et al.	2017	MFCC/ZCR	RF	0.841	0.8
Rocha et al.	2017	MFCC/STFT/ZCR ...	Not Mentioned	0.934	0.834
Di Perna et al.	2017	MFCC	Binary Classifier	0.86	0.8
You et al.	2017	acoustic subspace	SVM	0.871	0.879
Klco et al.	2017	Not Mentioned	Octonionic NN	0.82	0.96
Nguyen et al.	2018	MFCC/ZCR/CFA	NN/SVM/RF	0.987	NC
Windmond et al.	2018	MFCC/ZCR	RF	0.819	NC
Kadambi et al.	2018	168 Handcrated Feat.	DNN	0.937	0.976
Monge-Alvarez et al.	2018	Hu Moments	K-Nearest Neighbour	0.885	0.998
Barata et al.	2019	CNN	Ensemble CNN	0.917	0.901
Kvapilova et al.	2019	CNN	CNN	0.90-0.995	0.75-0.999
Teyhouee et al.	2019	Not Mentioned	HMM	0.87	0.9
Bales et al.	2020	CNN	CNN	0.919	0.862

Table 1 shows a selection of existing algorithms, those that reported the sensitivity and specificity of their algorithm. On average, the sensitivity and specificity reported are 87.67% and 92.18% respectively. Surprisingly, there is no flagrant variance in the reported score of sensitivity and specificity for cough detection algorithms from 2006 to 2020. Algorithms from 2006 report similar score than the more recent ones. This emphasises the need for a fair comparison of existing cough detection algorithms using constant metrics and a unique dataset.

As can be seen from Table 1, MFCCs are the most frequently mentioned features in cough detection literature; however, it is argued that MFFCs are poor features for privacy preservation as they reveal not only speech, but also inflection, and prosody [47,99]. Researchers attempt to increase the performance of algorithms and solve privacy related issues by experimenting with handcrafted features to identify better feature combinations.

In the classification stage of cough detection algorithms, NN, HMM, and SVM have proven to be effective but NN is used more frequently as shown in Table 1.

It is important to note that the progress in technology has made the CNN technique more popular for feature extraction and classification in the most recent studies. However, the reported results do not show any improvement when compared with other methods.

Amoh and Odame [53] compared two deep learning methods (CNN/RNN) with more conventional cough detection algorithms. They state that deeply learned features are more effective than hand-crated ones for cough detection.

A considerable advantage of using a DNN approach is that it learns to extract features while training, therefore removing the need for a feature extraction step before the classification of cough events. However, audio recordings used in cough monitoring analysis typically contain only a few seconds of cough events for each hour of recording. Liu et al. argue that DNN-only approaches would hardly capture features of cough if trained on the whole dataset, and they state that DNN is not suitable for the feature extraction step as it would occupy a lot of unnecessary computational resources. Instead, Liu et Al propose a two-step algorithm where features are extracted using a common MFCC approach to perform keyword detection. Followed by a second step where a combined DNN-HMM method is used for classification.

Similarly, Kadambi et al. propose a cough detection algorithm with a deep neural network (DNN) trained using MFCCs and other features to discriminate cough sounds from background noise. A total of 168 features are used as inputs to the DNN.

Table 2: Cough detection algorithms with best reported performance in terms of sensitivity and specificity

Author	Date	Feature Extraction	Classification	Sensitivity	Specificity
Larson et al.	2011	PCA	RF	0.92	0.995
Kadambi et al.	2018	168 Handcrated Feat.	DNN	0.937	0.976
Amrulloh et al.	2015	MFCC/ZCR ...	NN	0.93	0.98
Vizel et al.	2010	Time-freq Domain	Pattern Matching	0.96	0.94
Drugman et al.	2012	222 Handcrated Feat.	NN	0.947	0.95

There is no particular pattern characterising the top 5 cough detection algorithms shown in Table 2 as different methods were used in each study. Furthermore, it is not guaranteed that the same algorithms would populate the top 5 using the same evaluation setup and comparison of existing algorithms. For example, Drugman et al. [67] is in the current top 5; however, they use contact microphones which facilitates the detection of cough events in noisy environment.

Issues related to the use of different material, dataset, and metrics make it difficult to compare existing algorithms from reported results, but other issues affect the performance comparison. Some papers present results that appear to be erroneous. For example, Nguyen and Luo [50] which use their algorithm to return a prediction set of labels for each sound event detected. They state in their performance analysis that for 95% confidence level with Conformal Prediction, the 1-NN-Euclidean algorithm returned almost every label in the prediction set for the new samples. This means that the number of FP is expected to be significantly high and therefore the accuracy of the algorithm should be low as per equation (3). If all labels are returned in the prediction set out of 50 labels, one label would be correctly identified (TP) and the 49 others would be FP, resulting in an algorithm accuracy of 2% (As no label is returned "negative", TN = FN = 0). However, they show an accuracy of 95.25% in their results, which would correspond to the sensitivity score of the algorithm as it does not include the number of false positive. It can be argued that Nguyen and Luo [50] use their own definition of TP, FP, TN, and FN leading to these unexpected results, such as not counting incorrectly returned labels in the prediction set as FPs.

3.2 STATE-OF-THE-ART

From the research conducted, it is clear that the state-of-the-art of cough detection algorithms cannot be identified from the reported results in each study. This is due to the difference in recording conditions, type of microphones, study subjects, and type of non-cough sound events included in the dataset, as well as the use of different cough counting metrics and definition for evaluation criteria. For a fair evaluation and comparison of the different proposed cough detection methods, a large public dataset, universal metrics, and evaluation criteria definitions are needed.

Chapter 4

METHODOLOGY

In previous literature, the pre-processing methods used in cough detection algorithms have proven to be effective in the removal of silent segments while keeping most of the coughs present in the data. However, other sounds are also kept after the pre-processing stage, raising audio privacy concerns, and limiting the effectiveness of the data reduction process.

In this chapter, we propose a simple but effective pre-processing method that increases privacy preservation and considerably reduces the amount of data to be analysed while keeping 99.02% of the cough samples manually pre-annotated in the recordings. We use multiple-iteration pre-processing to further increase privacy preservation and the data reduction percentage by 20%. The amount of speech in the signal after a 5-iteration pre-processing is reduced by 88.94%. The remaining speech content is unintelligible and composed of higher energy syllables. This study compares the results of our algorithm with the performance of a more conventional pre-processing technique that also uses an energy threshold to discard non-relevant data.

4.1 RESEARCH DESIGN

Our method verifies the presence of signal energy in both lower and higher frequency regions and discards segments whose energy concentrates only on one of them. The method is iteratively applied on the same data to increase the percentage of data reduction and privacy preservation.

The principle behind our approach comes from the fact that cough energy occupies the entire spectrum area while other sounds, such as speech, are typically only present in certain frequency regions [50,62,63]. Figure 4 shows that while most cough energy is below 5 kHz energy is still present in much higher frequency regions whereas speech energy mostly lies below 2.5 kHz. Kosasih et al. [65] extracted important cough features from frequencies up to 90 kHz. However, this requires capturing the data at a very high sampling rate, therefore, increasing greatly the computation cost.

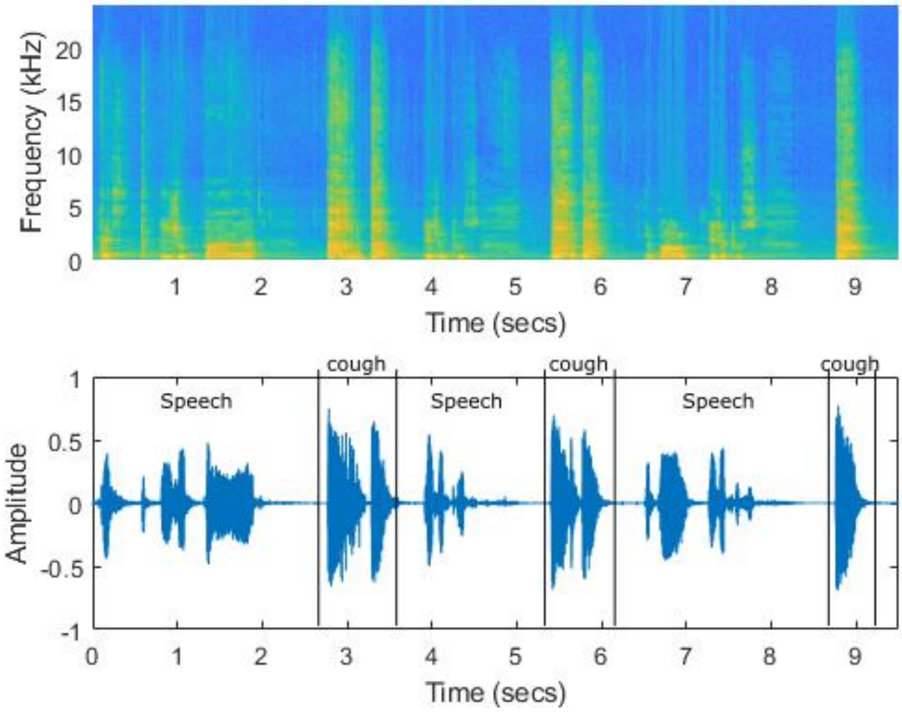


Figure 4: Spectrogram (top) and time-domain (bottom) representation of an audio signal composed of speech and cough events.

4.1.1 DATA REDUCTION SCORING SYSTEM

One aim of the research is to remove the most possible undesired data while preserving all cough events present in the original audio recording. While designing our pre-processing method, we created a scoring system to identify the optimal settings of key parameters for the algorithm. The data reduction performance of our pre-processing method is evaluated as a function of two measurements:

The percentage of data discarded. Calculated from the ratio of the duration of the remaining data over the total duration of the original audio file.

The percentage of cough events preserved. Calculated from the percentage of samples annotated as cough events preserved after pre-processing.

However, the audio content to be analysed post pre-processing are the cough events. Therefore, cough preservation must be prioritised over data reduction when calculating the performance of the algorithm. We decided to create a scoring system combining both parameters into a score *S* which represents the sum of both weighted parameters.

As cough preservation is of significantly higher importance than data reduction, the percentage of cough preservation is given a weight of 0.8 while the percentage of data removed is given a weight of 0.2.

The score *S* is calculated with the formula*:*

$$S = 0.8P_{cough} + 0.2P_{reduction} \tag{4}$$

where P_{cough} represents the percentage of cough preserved in the signal and $P_{reduction}$ is the percentage of data discarded after pre-processing.

This score is used in the calculation of key parameters for the thresholds of our algorithm.

4.1.2 SIGNAL FILTERING

In the first step of our pre-processing method, two signals are created by applying a high-pass 10th order Butterworth filter with a cut-off frequency of 4 kHz and a low-pass 10th order Butterworth filter with a cut-off frequency of 400 Hz to an original audio signal. The cut-off frequencies were determined from experiments on cough events recorded at various sampling rate. We used cough events sampled at 44.1 kHz from the ESC dataset [100], cough events sampled at 16 kHz from the AMI corpus [101], and coughs events sampled at 44.1 kHz from a person medically diagnosed with chronic cough. Energy is clearly present in the region close to half the sampling frequency for a majority of cough events; however, a few cough events have a lower energy presence in the higher frequency region. The cut-off frequency of the high-pass filter was gradually decreased to find its optimal value so that most of the speech is filtered out while lower intensity coughs can still be detected. Best results are obtained with a cut-off frequency of 400 Hz for the low-pass filter and a lower cut-off frequency of 4 kHz for the high-pass filter.

The two signals created represent the high frequency content and low frequency content of the original signal. Figure 5 confirms the presence of cough signal energy in both high and low frequency regions while, clearly, most of the speech energy only appears in the lower region.

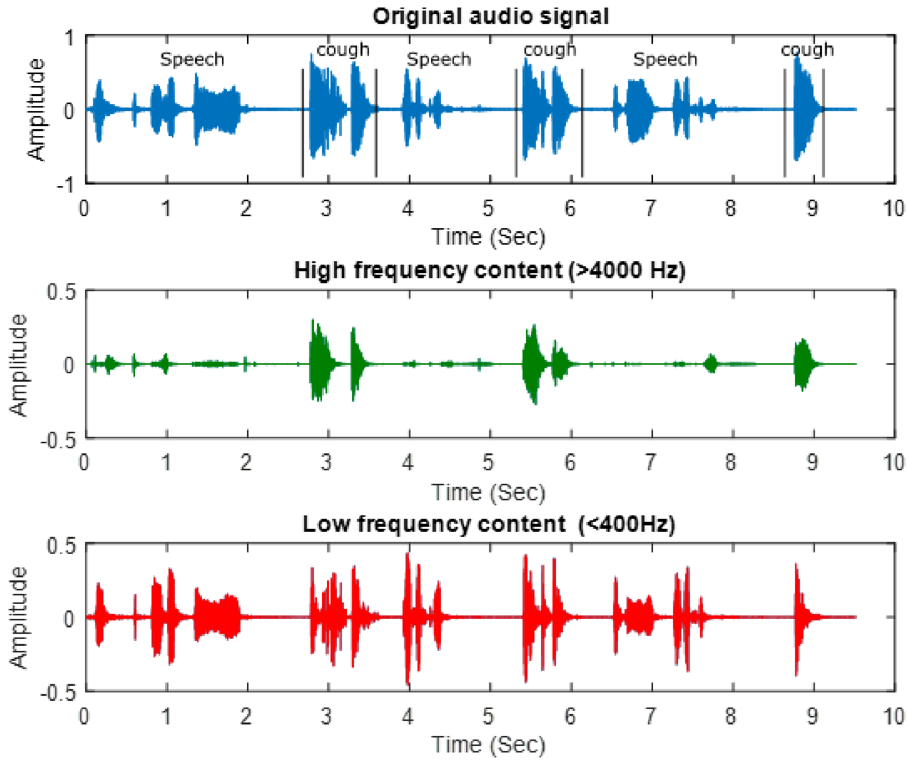


Figure 5: High frequency content (centre) and low frequency content (bottom) of the original audio signal (top).

4.1.3 Low energy segments removal

The second step consists in removing low energy segments from both filtered signals. We use a frame processing threshold technique to detect segments with significant energy in the higher and lower frequency regions. The signals are segmented into 50 ms non-overlapping frames and the energy of each frame is calculated and compared with a threshold value.

The energy of the i^{th} audio frame, E_i, is calculated with the formula:

$$E_i = \frac{\sum_{n=0}^{N-1} x_i(n)^2}{N} \qquad (5)$$

where $x_i(n)$ represents the n^{th} sample of the i^{th} frame x and N is the number of samples per frame.

4.1.3.1 THRESHOLD DEFINITION

The energy of each frame is compared with a threshold value to identify and discard low intensity frames. This value is unique for each signal and corresponds to a percentage of the mean energy of the entire raw signal.

The energy threshold value, T, is defined by:

$$T = \alpha \frac{\sum_{i=0}^{M-1} E_i}{M} \qquad (6)$$

where E_i represents the energy of the i^{th} frame of the signal, α is the threshold percentage parameter and M is the total number of frames in the signal.

The threshold percentage parameter α has a significant impact on the data reduction percentage and the preservation of the cough events present in the recordings. To determine the optimal value for α, we conducted a two-part experiment using 1.5 hour of the AMI meeting recordings containing speech, silence and 13 cough events of different intensity.

For a fair evaluation, the audio recordings must contain a combination of silent segments and multiple sound events such as speech, cough sounds and background noise. The majority of the data previously used to evaluate our filter cut-off frequencies contain only cough events and silent segments. Therefore, to not bias the results, the recordings from the chronic cough patient, and the ESC dataset are removed from the data set and only the recordings from the AMI corpus are used in the evaluation data.

4.1.3.2 ALPHA PARAMETER

In the first part of the experiment, we determine the optimal value of the α parameter for each signal by measuring the percentage of data reduction and the percentage of cough preservation after one iteration of the algorithm. The score S for each value of α is calculated as per equation (4).

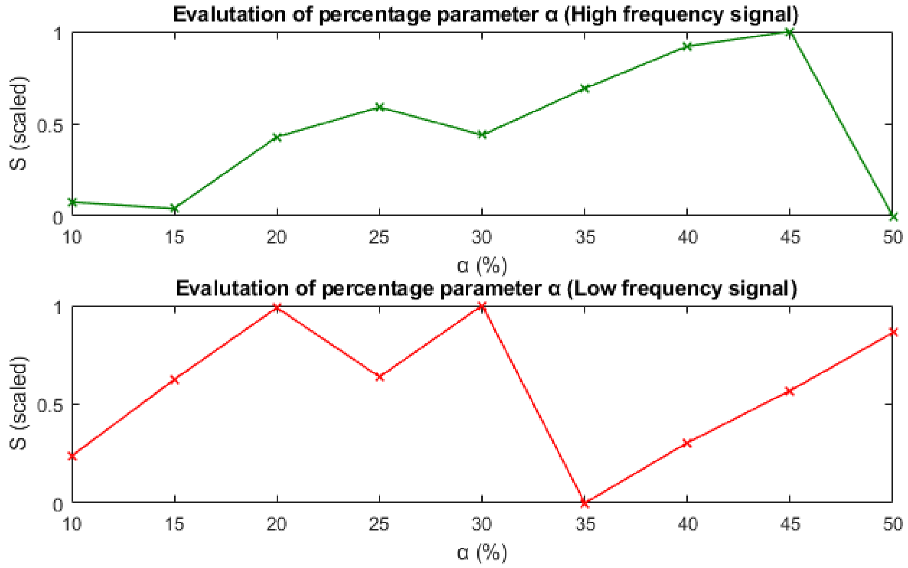


Figure 6: Evaluation of the threshold percentage parameter α for the high frequency signal (Top) and the low frequency signal (Bottom).

The value of α with the highest normalised score S is selected as the initial α value in each signal. From Figure 6, the initial α is 45% for the high frequency signal and 30% for the low frequency signal.

When iteratively applying the algorithm to the same signal, the mean energy of the signal is increasing with every iteration since low energy content is discarded. With a constant value for α, the threshold value is increasing with the mean energy of the signal.

4.1.3.3 BETA PARAMETER

In the second part of our experiment, we implement a parameter β which provides a progressive threshold compensating for the increase in the mean energy of the signal by decreasing the percentage parameter α after each iteration.

The progressive energy threshold value, T_p, is defined by:

$$T_p = \left(\alpha - \beta(r-1)\right)\frac{\sum_{i=0}^{M-1} E_i}{M} \quad ,r \geq 1 \qquad (7)$$

Where E_i represents the energy of the i^{th} frame of the signal, α is the threshold percentage, β is the percentage decrease of α after each iteration, r is the pre-processing iteration number, and M is the total number of frames in the signal

The optimal value of β is determined with an experiment. The overall percentage of data reduction and the percentage of cough events preserved after pre-processing is measured and the score S is calculated as per equation (4).

Figure 7 shows how the overall normalised score S is changing when varying β in both the high frequency and the low frequency signals. The overall performance of the algorithm increases when setting β to 4% for the low frequency signal threshold while no improvement is observed when varying β for the high frequency signal threshold.

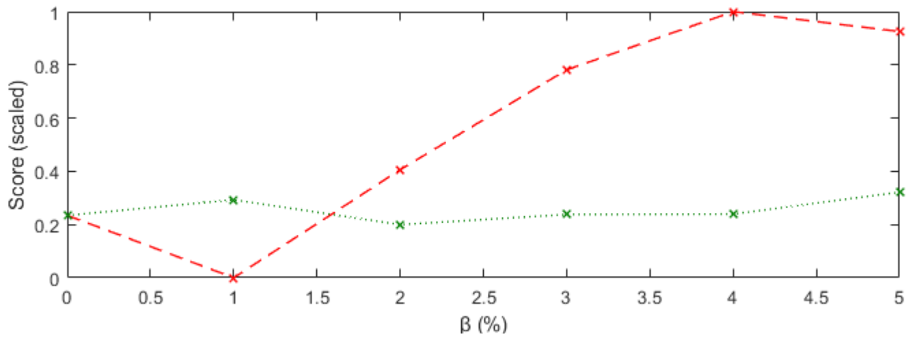


Figure 7: Evaluation of β for the high frequency signal (Dotted line) and the low frequency signal (Dashed line).

Therefore, our algorithm threshold values are calculated as per equation (7) using the following parameters:

- High frequency signal: $\alpha = 45\%$, $\beta = 0\%$,

- Low frequency signal: $\alpha = 30\%$, $\beta = 4\%$.

When comparing each frame energy to the threshold values, the frame is annotated as being a potential cough event if its energy is above the defined threshold, such as:

$$x_i = \begin{cases} 1, & if\ E_i \geq T_p \\ 0, & otherwise \end{cases} \quad (8)$$

where $x_i = 1$ means all samples from the i^{th} frame are assigned the value of 1 (potential cough)

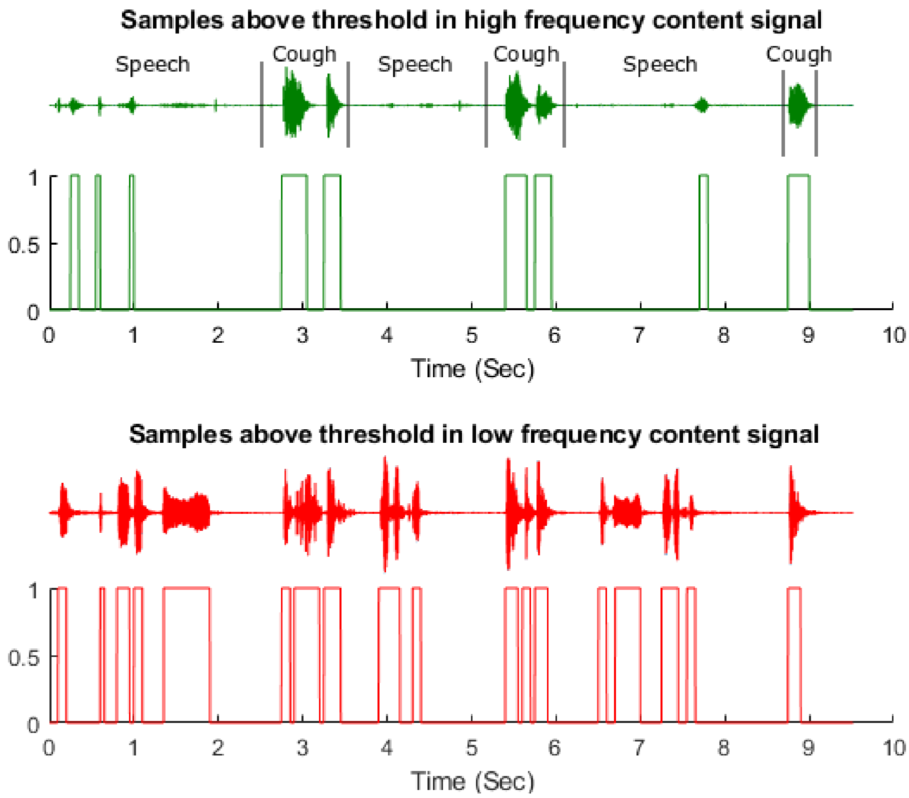


Figure 8: Thresholded high frequency content signal (top) and low frequency content signal (bottom). All cough events appear in both thresholded signals while only parts of speech are preserved in the high frequency content signal.

4.1.4 THRESHOLDED SIGNALS COMBINATION AND SMOOTHING

The next stage in our pre-processing technique is to combine the thresholded outputs from both signals (Figure 8) using a logical AND conjunction.

Figure 9 shows that the high and low frequency content in speech does not systematically occur simultaneously. The logical AND conjunction allows for cough detection while discarding all sounds with only low or high frequency content, increasing the data reduction percentage.

The raw logical AND conjunction output needs to be smoothed to ensure the cough events are preserved integrally. Therefore, all samples around a positive output (30 ms before to 300 ms after) are also annotated as potential cough events. This 330 ms window is typically appropriate to catch the entirety of a cough event while limiting possible speech intelligibility issues in case of false positive. It can be clearly seen in Figure 9 that cough events are preserved entirely after smoothing while all speech is removed from the audio signal.

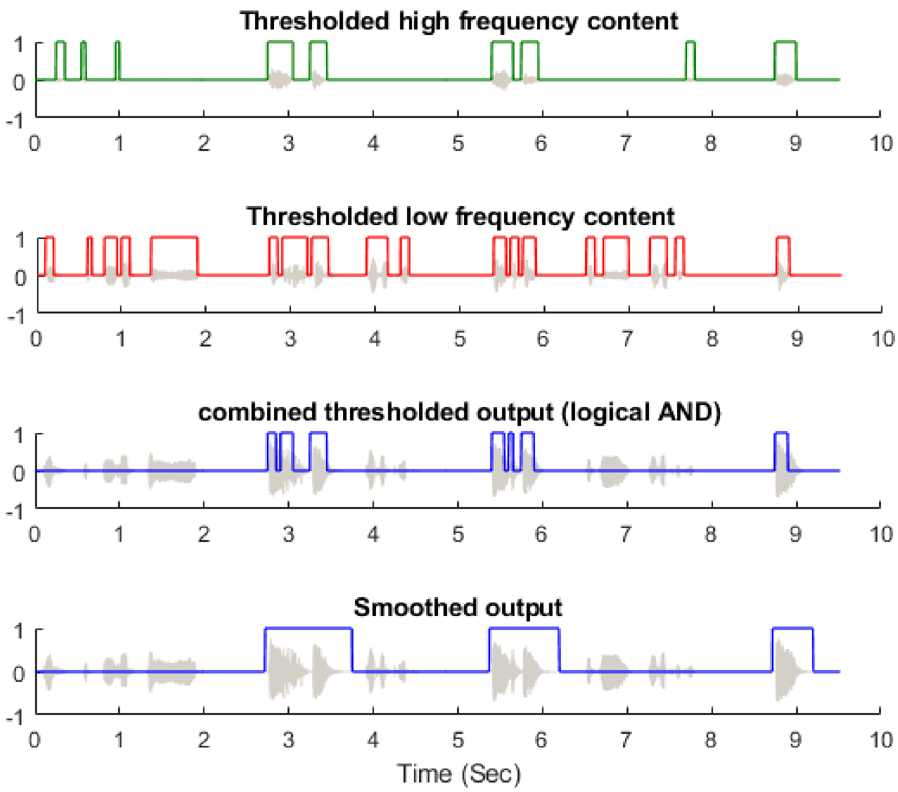


Figure 9: Outputs after logical AND conjunction and smoothing

4.1.5 PRE-PROCESSED AUDIO SIGNAL

In the final step of our pre-processing technique, we create a new audio signal composed of only the samples detected as a potential cough event in the smoothed output, keeping a time stamp track of each event in the original signal. For multi-iteration pre-processing, the algorithm follows the same steps using the new audio signal instead of the original signal at each iteration.

The data is then ready for feature extraction and classification, limiting computation time and the amount of speech that can potentially be classified as a cough event by the classification algorithm.

4.2 EVALUATION DATA

To evaluate our algorithm, we use a subset of the Augmented Multi-party Interaction (AMI) corpus [101], an annotated multi-modal data set consisting of 100 hours of meeting recordings. The AMI corpus is publicly available under the Creative Commons Attribution 4.0 license agreement and has been annotated with the start and end times of cough events in the recordings; however, many of the annotated cough events were respiratory sounds such as throat clearing, moaning, sigh, and other similar sounds.

For a reliable evaluation of our algorithm, the AMI meetings containing at least 10 cough annotations were selected for a second re-annotation. All non-cough respiratory sound annotations were discarded, and the AMI meetings were preserved if they contained at least 2 coughing events after the re-annotation, leaving 16 meeting recordings in our evaluation data. Each AMI meeting is identifiable with a unique ID and is publicly available for download online. The meeting recordings were grouped by type of meeting and capture location to create five audio files of different length and containing cough events of different intensity.

Table 3: AMI meeting IDs which constitute the five audio files of our dataset.

File 1	Natural meetings captured in Edinburgh
EN2001a, EN2001e, EN2006a, EN2006b.	
File 2	Scenario meetings captured in Edinburgh
ES2002a, ES2002b, ES2002d, ES2005b, ES2008b, ES2012b, ES2013c, ES2016b.	
File 3	Scenario meeting captured in Switzerland
IB4004.	
File 4	Natural meetings captured in Switzerland
IN1002, IN1012.	
File 5	Scenario meeting captured in the Netherlands
TS3006c.	

Table 4 shows each file duration and their new cough events distribution after re-annotation. Our subset contains 113 coughs in more than 12 hours of recordings. The word count reported in the AMI corpus annotations will be used to evaluate the privacy preservation performance of our algorithm. A total of 144 583 words are annotated in our data set.

Table 4: Data set files description

File Number	File Duration	New Cough Count	Speech Word Count
1	4h24m57s	35	52,358
2	4h55m31s	44	52,922
3	0h39m53s	15	9,134
4	1h33m01s	13	21,234
5	0h43m04s	6	8,755
Total	12h16m26s	113	144,583

Wearable microphones are often used for cough monitoring; however, it can be intrusive to continuously carry a recording device. An ideal way of monitoring the occurrence of cough events is to capture environmental sounds. The AMI meetings recordings sampled at 16 kHz were captured with multiple microphones including one headset and one lapel microphones per participant, and two omni-directional microphone arrays placed at the centre of the room.

In our study, we evaluate our algorithm using both the mix headset recordings, which combines all the individual headset files in each meeting, and the microphone (Array1-01) from the first omni-directional microphone array.

These two sets of recordings will help measure the performance of the algorithm under two conditions: Cough monitoring with wearable technology and environmental sensing cough monitoring.

4.3 RESULTS

4.3.1 EVALUATION CRITERIA

We evaluated our method based on three criteria:

Data reduction - Percentage of data removed from the original audio signal.

Cough preservation - Percentage of samples annotated as cough events preserved after pre-processing.

Privacy preservation - Percentage of samples annotated as speech discarded after pre-processing.

4.3.2 COMPARISON WITH A REGULAR METHOD

Our algorithm was tested on each of the five files constituting our data set and the results were compared with a regular pre-processing method that also uses a simple energy threshold approach. The energy threshold of the regular method was set to 40% of the signal mean energy to match with the values used in our algorithm.

The results for data reduction and cough preservation presented in Table 5 are obtained with the mix headset capture version of the data set while Table 6 shows the results obtained with one microphone (Array1-01) of the omni-directional microphone array. Both tables report the percentage of data discarded from the original audio file and the percentage of cough preservation calculated by using the validated cough count from Table 4.

The scaled averages in Table 5 and Table 6 represent the overall percentage of data removed from the original audio signal and is calculated by combining the data reduction percentage with the duration of each file in Table 4. The results for both audio file versions are combined in Table 7 to obtain the overall performance of the pre-processing method. Our algorithm reduces the data set duration from 12h16m26s to 3h29m43s, discarding 71.52% of the data. The regular pre-processing method reaches 37.26% data reduction, leaving the new signal with a duration of 7h42m03s. Both methods preserve over 99% of the annotated cough events. This shows that our algorithm is on average twice as effective as a regular pre-processing method at reducing the amount of data to be analysed.

Table 5: Performance comparison (Mix headset)

	Our Algorithm		Regular Method	
File Number	**Data Reduction**	**Cough Preservation**	**Data Reduction**	**Cough Preservation**
1	72.10%	100.00%	48.21%	100.00%
2	80.43%	99.96%	53.19%	100.00%
3	69.88%	100.00%	26.75%	100.00%
4	71.74%	99.56%	27.90%	91.58%
5	73.98%	100.00%	42.91%	100.00%
Scaled Average	75.39%	99.93%	46.17%	99.03%

Table 6: Performance comparison (Mic. array)

	Our Algorithm		Regular Method	
File Number	**Data Reduction**	**Cough Preservation**	**Data Reduction**	**Cough Preservation**
1	70.64%	100.00%	31.77%	100.00%
2	68.10%	100.00%	26.49%	100.00%
3	61.87%	100.00%	21.08%	100.00%
4	62.09%	83.55%	28.23%	99.98%
5	63.60%	100.00%	27.08%	100.00%
Scaled Average	67.66%	98.11%	28.35%	99.99%

Table 7: Combined performance comparison

	Our Algorithm		Regular Method	
File Number	**Data Reduction**	**Cough Preservation**	**Data Reduction**	**Cough Preservation**
Mix headset	75.39%	99.93%	46.17%	99.03%
Array-01	67.66%	98.11%	28.35%	99.99%
Overall	71.52%	99.02%	37.26%	99.51%

4.3.3 MULTIPLE ITERATION PRE-PROCESSING

It is possible to increase the percentage of data removed from the original audio by performing the pre-processing stage on the same data multiple times successively. For each iteration, our algorithm follows the steps described in the methodology; however, to prevent the removal of lower intensity coughs from the recordings, a progressive energy threshold is calculated for the low frequency signal as described by equation (7).

We performed five iterations of our pre-processing on both file versions (mix headset and microphone array). The results in Table 8 and Table 9 show the detailed performance of our algorithm at each of the five iterations for both file versions. The results are combined in Table 10 to obtain the overall performance of the algorithm over a five-iteration pre-processing.

The percentage of data removed from the original file and the percentage of cough preservation is reported for each file and for each iteration. From the overall results shown in Table 10 and using the score formula as per equation (4), we can identify that the algorithm reaches higher performance with 2 and 3 iterations.

Table 8: Multiple pre-processing stage (Mix headset)

		Iteration				
		1	2	3	4	5
File 1	**Data Reduction**	70.64%	85.76%	90.49%	92.31%	93.22%
	Cough Preservation	100%	99.94%	99.09%	99.03%	99.03%
File 2	**Data Reduction**	68.10%	82.68%	87.44%	89.42%	90.46%
	Cough Preservation	100%	100%	99.99%	98.68%	98.60%
File 3	**Data Reduction**	61.87%	77.63%	81.96%	83.88%	84.85%
	Cough Preservation	100%	100%	99.91%	99.91%	99.91%
File 4	**Data Reduction**	62.09%	76.23%	80.27%	81.37%	81.60%
	Cough Preservation	83.55%	80.12%	75.07%	74.74%	74.09%
File 5	**Data Reduction**	63.60%	80.02%	85.76%	88.04%	89.12%
	Cough Preservation	100%	100%	99.55%	99.55%	99.55%
Scaled Average	**Data Reduction**	67.66%	82.54%	87.23%	89.06%	89.95%
	Cough Preservation	98.11%	97.69%	96.81%	96.24%	96.14%

Table 9: Multiple pre-processing stage (Mic. array)

		Iteration				
		1	2	3	4	5
File 1	**Data Reduction**	72.10%	84.97%	88.84%	90.42%	91.34%
	Cough Preservation	100%	99.98%	99.42%	99.42%	99.19%
File 2	**Data Reduction**	80.43%	92.06%	95.31%	96.58%	97.09%
	Cough Preservation	99.96%	93.97%	89.34%	88.93%	87.71%
File 3	**Data Reduction**	69.88%	83.39%	86.91%	88.41%	89.12%
	Cough Preservation	100%	100%	100%	89.91%	89.91%
File 4	**Data Reduction**	71.74%	85.33%	89.59%	91.44%	92.37%
	Cough Preservation	99.56%	95.88%	95.22%	94.42%	94.09%
File 5	**Data Reduction**	73.98%	85.38%	88.60%	90.09%	90.80%
	Cough Preservation	100%	100%	100%	100%	100%
Scaled Average	**Data Reduction**	75.39%	87.80%	91.41%	92.89%	93.63%
	Cough Preservation	99.93%	97.17%	95.12%	93.53%	92.94%

Table 10: Overall performance with multi pre-processing stages

		Iteration				
		1	2	3	4	5
Mix headset	Data Reduction	75.39%	87.80%	91.41%	92.89%	93.63%
	Cough Preservation	99.93%	97.17%	95.12%	93.53%	92.94%
Mic. Array-01	Data Reduction	67.66%	82.54%	87.23%	89.06%	89.95%
	Cough Preservation	98.11%	97.69%	96.81%	96.24%	96.14%
Scaled Average	Data Reduction	71.52%	85.17%	89.32%	90.98%	91.79%
	Cough Preservation	99.02%	97.43%	95.96%	94.89%	94.54%

Table 10 shows that the percentage of data reduction can be increased by around 20% with five iterations; however, the improvement in data reduction from one iteration to another decreases progressively. The overall data reduction percentage increases, on average, by 13.65% with the second iteration then by 4.15%, 1.65% and 0.81% with respectively the third, fourth and fifth iterations. The number of iterations has an impact on the number of cough events preserved in the recordings as, on average, the cough preservation percentage drops by 1% with every iteration.

4.3.4 PRIVACY PRESERVATION EVALUATION

To evaluate privacy preservation, we use the start and end times of each annotated word in the data. Table 11 and Table 12 show the percentage of speech samples discarded after each stage of a five-iteration pre-processing. The scaled average is calculated by combining the percentage of speech discarded with the number of words in each file in Table 4.

Table 11: Speech removal percentage (Mix headset)

	Iteration				
	1	2	3	4	5
File 1	61.85%	78.49%	83.61%	85.72%	86.97%
File 2	75.02%	89.76%	93.97%	95.60%	96.28%
File 3	59.19%	76.96%	81.62%	83.57%	84.44%
File 4	65.15%	80.56%	85.71%	88.04%	89.26%
File 5	66.07%	80.40%	84.40%	86.23%	87.09%
Scaled Average	67.24%	82.94%	87.63%	89.56%	90.56%

Table 12: Speech removal percentage (Mic. array)

	Iteration				
	1	**2**	**3**	**4**	**5**
File 1	61.80%	79.95%	85.91%	88.34%	89.59%
File 2	62.13%	78.88%	84.53%	86.93%	88.18%
File 3	48.90%	68.75%	74.31%	76.84%	78.23%
File 4	56.21%	72.38%	79.59%	83.02%	84.61%
File 5	55.60%	73.45%	80.49%	83.44%	84.88%
Scaled Average	59.89%	77.33%	83.40%	86.00%	87.33%

Tables 11 and 12 show that while our algorithm is efficient in discarding speech from the audio file with one iteration, the percentage of speech removed after 5 iterations is significantly increased.

The overall percentage of speech discarded after pre-processing is calculated in Table 13. Our algorithm discards, on average, 63.57% of the speech content in the data with one iteration. When a five-iteration pre-processing is applied, 88.94% of the speech is discarded.

Table 13: Overall speech removal percentage

	Iteration				
	1	**2**	**3**	**4**	**5**
Mix headset	67.24%	82.94%	87.63%	89.56%	90.56%
Mic. Array	59.89%	77.33%	83.40%	86.00%	87.33%
Scaled Average	63.57%	80.13%	85.52%	87.79%	88.94%

While some information, such as the sex of the speaker, can still be guessed from the remaining 11.06% of speech in the audio, the content of the speech cannot be recovered, and the privacy of the speaker is considerably preserved with our algorithm.

4.3.5 RATING OF INTELLIGIBILITY

An objective rating of intelligibility can be made by speech transmission index. The speech transmission index is used to measure speech transmission quality and can be linked to subjective intelligibility tests such as the percentage of correctly identified words [102]. It is a 0 to 1 index, where a value of 1 means the speech remains perfectly intelligible and the closer the value approaches 0, the more information is lost. The speech transmission index is used to rate speech intelligibility from bad to excellent on a five-point scale.

Table 14: Intelligibility rating scale

Index	0 - 0.3	0.3 - 0.45	0.45 - 0.6	0.6 - 0.75	0.75 - 1
Rating	Bad	Poor	Fair	Good	Excellent

Our algorithm preserves 36.43% of the words present in the signal with one iteration and from 11.06% to 19.87% with multiple iterations. From the relation between speech intelligibility and speech transmission index values [102], our algorithm achieves a speech transmission index ranging from 0.2 to 0.3, which is rated as "bad intelligibility" according to [103], the lowest level of intelligibility on the five-point scale.

4.4 DISCUSSION

When dealing with cough monitoring through audio analysis, the audio data is typically recorded with wearable or ambient microphones. Our data set is constituted of meeting recordings captured with headset microphones and two microphone arrays. We evaluate our algorithm on two versions of this data set: a mix headset version, which combines all headset recordings in one meeting, and one microphone from a microphone array.

The criteria of evaluation are the percentage of data reduction, the percentage of cough events preservation and the percentage of speech removed from the data after pre-processing.

A higher data reduction percentage is reached with the mix headset version of our data set. The algorithm removes 75.39% of the data with the mix headset files and 67.66% with the microphone array files. When a five-iteration pre-processing is applied, a data reduction of 93.63% and 89.95% is achieved. The difference in the data reduction percentage between the mix headset files and the microphone array files can be explained by the fact that the energy difference in sounds like background noise, speech, and other human sounds is greater in the mix headset files than in the microphone array files. With less variance in the signal energy level, the number of sound events below the threshold decreases and the data reduction percentage is impacted. The position of the capturing device can also impact the cough preservation percentage as a coughing subject might not be facing the microphone reducing the energy level of some coughing events.

An increase in the risk of discarding lower intensity coughs is observed when performing multiple-iteration pre-processing. Our results showed that the percentage of cough events preserved with a five-iteration pre-processing drops by 1% with every iteration.

The recording of audio data raises ethical issues in relation to privacy. Regular silence removal pre-processing techniques delete only silence and lower energy sound events from the data, often leaving speech intact and clearly intelligible. Our algorithm addresses this issue by discarding all speech segments that does not contain significant energy in both high and low frequency regions. Privacy preservation is increased by removing part of speech from the data, distorting the meaning of the speech segments left after pre-processing and limiting the possibility that speech maybe listened to by a reviewer. Furthermore, multiple-iteration pre-processing can increase privacy by removing up to 88.94% of speech from the original recording at the cost of increasing the risk of discarding lower energy cough events.

Chapter 5

FUTURE WORK AND CONCLUSION

The aim of this study is to improve pre-processing techniques traditionally used in cough detection algorithms. We propose an effective pre-processing method that increases privacy preservation by removing parts of speech from the data in addition to silence and other low energy sound events. Our algorithm is tested on two audio data sets constituted of meetings captured by wearable microphones and an ambient microphone. A data reduction percentage of 71.52% is reached and 99.02% of the cough events were preserved after pre-processing. This performance makes our algorithm two times as effective as a regular simple energy threshold pre-processing method. Furthermore, by pre-processing the same data multiple times with our algorithm, the data reduction can be increased by 20%, bringing the data reduction percentage to 91.79%. Performing five iterations of our pre-processing method greatly increases privacy preservation by discarding 88.94% of the speech from the data; however, there is an added risk of inadvertently discarding lower energy cough events.

We currently use an energy entropy technique to compare each frame energy to the mean energy of a signal. In future work, we will investigate the implementation of frequency-domain features instead of time-domain features in the second stage of our algorithm. Techniques such as spectral flux, which measure the variance of the power spectrum in a signal, could provide a better identification of cough events and discard more non-relevant data; therefore, it could potentially increase the overall performance of our pre-processing method.

Chapter 6

REFERENCES

[1] F. De Blasio et al., “Cough management: a practical approach,” *Cough,* vol. 7:7, 2011.

[2] G. Ainslie, “Assesment of cough,” *CME,* vol. 27, no. 2, pp. 68-71, Feb 2009.

[3] S. Schappert, “National Ambulatory Medical Care Survey: 1991 summary,” in *Vital and health statistics,* Vols. Series 13: Data From the National Health Survey No, 116, 1994, pp. 1-110.

[4] J. Korpás, J. Sadloňová and M. Vrabec, “Analysis of the Cough Sound: an Overview,” *Pulmonary Pharmacology,* vol. 9, pp. 261-268, 1996.

[5] D. E. Leith, J. P. Butler, S. L. Sneddon and J. D. Brain, “Cough,” in *Handbook of physiology, the respiratory system,* vol. 3 Part 1, 1986, pp. 316-336.

[6] E. Vizel et al., “Validation of an ambulatory cough detection and counting application using voluntary cough under different conditions,” *Cough,* vol. 6, no. 3, 2010.

[7] R. S. Irwin, “Complications of cough: ACCP evidence-based clinical practice guidelines,” *Chest,* Vols. 129 (1 Suppl):54S-58S, Jan 2006.

[8] G. A. Fontana, “Before we get started: what is a cough?,” *Lung,* vol. 186, no. Suppl 1:S3-6, 2008.

[9] P. S. Cho, S. S. Birring, H. V. Fletcher and R. D. Turner, "Methods of Cough Assessment," *The Journal of Allergy and Clinical Immunology: In Practice,* vol. 7, no. 6, pp. 1715-1723, 2019.

[10] S. S. Birring and A. Spinou, "How best to measure cough clinically," *Current Opinion in Pharmacology,* vol. 22, pp. 37-40, 2015.

[11] A. A. Raj and S. S. Birring, "Clinical assessment of chronic cough severity," *Pulmonary Pharmacology & Therapeutics,* vol. 20, no. 4, pp. 334-337, 2007.

[12] S. Leconte, D. Ferrant, V. Dory and J. Degryse, "Validated Methods of Cough Assessment: A Systematic Review of the Literature," *Respiration,* vol. 81, pp. 161-174, 2011.

[13] K. K. Lee et al., "Sound: a non-invasive measure of cough intensity," *BMJ Open Respiratory Research,* vol. 4, no. 1, 2017.

[14] I. M. Paul, K. Wai, S. J. Jewell, M. L. Shaffer and V. V. Varadan, "Evaluation of a new self-contained, ambulatory,objective cough monitor," *Cough,* vol. 2:7, 2006.

[15] S. J. Barry, A. D. Dane, A. H. Morice and A. D. Walmsley, "The automatic recognition and counting of cough," *Cough,* vol. 2:8, 2006.

[16] M. Krajnik, I. Damps-Konstanska, L. Gorska and E. Jassem, "A portable automatic cough analyser in the ambulatory assessment of cough," *BioMedical Engineering OnLine,* vol. 9, 2010.

[17] G. H. Botha et al., "Detection of tuberculosis by automatic cough sound analysis," *Physiological Measurement,* vol. 39, no. 4, 2018.

[18] J. I. Hall, M. Lozano, L. Estrada-Petrocelli, S. Birring and R. Turner, "The present and future of cough counting tools," *Journal of Thoracic Disease,* vol. 12, no. 9, Sep 2020.

[19] K. M. Schmit et al., "Evaluating Cough Assessment Tools A Systematic Review," *Chest,* vol. 144, no. 6, pp. 1819-1826, 2013.

[20] S. T. Kulnik, N. M. Williams, L. Kalra, J. Moxham and S. S. Birring, "Cough frequency monitors: can they discriminate patient from environmental coughs?," *Journal of Thoracic Disease,* vol. 8(11), pp. 3152-3159, 2016.

[21] C. R. Woolf and A. Rosenberg, "Objective assessment of cough suppressants under," *Thorax*, vol. 19, pp. 125-130, 1964.

[22] M. Grobman, T. Lever and C. R. Reinero, "Discrimination between respiratory and non-respiratory sound waveforms in dogs using acoustic wave recordings: An objective metric of cough," *The Veterinary Journal,* vol. 253, 2019.

[23] J. Y. Hsu, R. A. Stone, R. B. Logan-Sinclair, M. Worsdell, C. M. Busst and K. F. Chung, "Coughing frequency in patients with persistent cough: assessment using a 24 hour ambulatory recorder," *European Respiratory Journal,* vol. 7, no. 7, pp. 1246-1253, 1994.

[24] H. Sumner et al., "Predictors of Objective Cough Frequency in Chronic Obstructive Pulmonary Disease," *American Journal of Respiratory and Critical Care Medicine* , vol. 187, no. 9, 2013.

[25] J. A. Smith, J. E. Earis and A. A. Woodcock, "Establishing a gold standard for manual cough counting: video versus digital audio recordings," *Cough,* vol. 2, 2006.

[26] R. Turner and G. H. Bothamley, “How to count coughs? Counting by ear, the effect of visual data and the evaluation of an automated cough monitor,” *Respiratory Medicine Journal,* vol. 108, no. 12, pp. 1808-1815, 2014.

[27] N. Yousaf, W. Monteiro, S. Matos, S. S. Birring and I. D. Pavord, “Cough frequency in health and disease,” *European Respiratory Journal,* vol. 41, pp. 241-243, 2013.

[28] A. Barton, P. Gaydecki, K. Holt and J. A. Smith, “Data reduction for cough studies using distribution of audio frequency content,” *Cough,* vol. 8, 2012.

[29] L. Di Perna et al., “An automated and unobtrusive system for cough detection,” *2017 IEEE Life Sciences Conference (LSC),* pp. 190-193, 2017.

[30] M. Młyńczak, K. Pariaszewska and G. Cybulski, “Automatic cough episode detection using a vibroacoustic sensor,” in *Engineering in medicine and biology society (EMBC), 2015 37th annual international conference of the IEEE,* 2015.

[31] H. Mohammadi, A. Samadani, C. Steele and T. Chau, “Automatic discrimination between cough and non-cough accelerometry signal artefacts,” *Biomedical Signal Processing and Control,* vol. 52, pp. 394-402, 2019.

[32] A. B. Chang, R. G. Newman, P. D. Phelan and C. F. Robertson, “A new use for an old Holter monitor: an ambulatory cough meter,” *European Respiratory Journal,* vol. 10, pp. 1637-1639, 1997.

[33] T. Drugman et al., “Objective Study of Sensor Relevance for Automatic Cough Detection,” *IEEE Journal of Biomedical and Health Informatics,* vol. 17(3), pp. 699-707, 2013.

[34] S. A. Mahmoudi et al., “Sensor-based System for Automatic CoughDetection and Classification,” in *8th ICT Innovations Conference 2016. ELEMENT 2015 - Enhanced Living EnvironMENTs*, Macedonia, 2015.

[35] M. Soliński, M. Łepek and Ł. Kołtowski, “Automatic cough detection based on airflow signals for portable spirometry system,” *Informatics in Medicine Unlocked,* vol. 18, 2020.

[36] M. A. Coyle et al., “Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease,” *Cough,* vol. 1, no. 3, 2005.

[37] F. D. McCool, “Global Physiology and Pathophysiology of Cough,” *Supplement diagnosis and management of cough: accp evidence-based clinical practice guidelines,* vol. 129, no. 1, pp. 48S-53S, 2006.

[38] M. Polverino, F. Polverino, M. Fasolino, F. Andò, A. Alfieri and F. De Blasio, “Anatomy and neuro-pathophysiology of the cough reflex arc,” *Multidisciplinary Respiratory Medicine,* vol. 7(1): 5, 2012.

[39] P. Piirilä and A. R. Sovijärvi, “Objective assessment of cough,” *European Respiratory Journa,* vol. 8, pp. 1949-1956, 1995.

[40] A. H. Morice, “Rebuttal: Cough Is an Expiratory Sound,” *Lung,* vol. 186, pp. 7-9, 2008.

[41] A. H. Morice et al., “ERS guidelines on the assessment of cough,” *European Respiratory Journal,* vol. 29, pp. 1256-1276, 2007.

[42] S. Matos, S. S. Birring, I. D. Pavord and D. H. Evans, "Detection of Cough Signals in Continuous Audio Recordings Using Hidden Markov Models," *IEEE Transactions On Biomedical Engineering,* vol. 53, no. 6, pp. 1078-1083, 2006.

[43] R. X. Pramono, S. A. Imitaz and E. Rodriguez-Villegas, "A Cough-Based Algorithm for Automatic Diagnosis of Pertussis," *PLOS ONE,* vol. 11(9), 2016.

[44] Y. Shi, H. Liu, Y. Wang, M. Cai and W. Xu, "Theory and Application of Audio-Based Assessment of Cough," *Journal of Sensors,* vol. 2018, 2018.

[45] A. Spinou and S. S. Birring, "An update on measurement and monitoring of cough: what are the important study endpoints?," *Journal of Thoracic Disease (Chronic Cough) ,* vol. 6, no. Suppl 7, 2014.

[46] D. Liaqat, E. Nemati, M. Rahman and J. Kuang, "A method for preserving privacy during audio recordings by filtering speech," in *2017 IEEE Life Sciences Conference (LSC),* Sydney, 2017.

[47] E. C. Larson, T. Lee, S. Liu, M. Rosenfeld and S. N. Patel, "Accurate and Privacy Preserving Cough Sensing using a Low-Cost Microphone," in *UbiComp'11,* Beijing, 2011.

[48] S. Kumar, L. T. Nguyen, M. Zeng, K. Liu and J. Zhang, "Sound Shredding: Privacy Preserved Audio Sensing," in *HotMobile'15,* Santa Fe, 2015.

[49] D. Smith, "Five principles for research ethics," *Monitor on Psychology,* vol. 34, no. 1, p. 56, 2007.

[50] K. A. Nguyen and Z. Luo, "Cover Your Cough: Detection of Respiratory Events with Confidence Using a Smartwatch," *Proceedings of Machine Learning Research,* vol. 91, no. 1, pp. 1-18, 2018.

[51] A. Windmon, M. Minakshi, S. Chellappan, P. R. Athilingam, M. Jahanson and B. A. Jenkins, "On Detecting Chronic Obstructive Pulmonary Disease (COPD) Cough using Audio Signals Recorded from Smart-Phones," in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, Funchal, 2018.

[52] B. Ferdousi, S. M. Agsanullah, K. Abdullah-Al-Mamun and M. N. Huda, "Cough detection using speech analysis," in *2015 18th International Conference on Computer and Information Technology (ICCIT),* Dhaka, 2015.

[53] J. Amoh and K. Odame, "Deep Neural Networks for Identifying Cough Sounds," *IEEE Transactions On Biomedical Circuits And Systems,* vol. 10, no. 5, pp. 1003-1011, 2016.

[54] B. H. Tracey, G. Comina, S. Larson, M. Bravard, J. W. López and R. H. Gilman, "Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis," in *33rd Annual International Conference of the IEEE EMBS,* Boston, 2011.

[55] S. Larson, G. Comina, R. H. Gilman, B. H. Tracey, M. Bravard and J. W. Lopez, "Validation of an Automated Cough Detection Algorithm for Tracking Recovery of Pulmonary Tuberculosis Patients," *PLOS ONE,* vol. 7, no. 10, 2012.

[56] B. Rocha, P. De Carvalho and R. P. Paiva, "Detection of Explosive Cough Events in Audio Recordings by Internal Sound Analysis," in *IEEE Engineering in Medicine and Biology Conference*, 2017.

[57] P. Kadambi et al., “Towards a Wearable Cough Detector Based on Neural Networks,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Calgary, 2018.

[58] S. Matos, S. S. Birring, I. D. Pavord and D. H. Evans, “An Automated System for 24-h Monitoring of Cough Frequency: The Leicester Cough Monitor,” *IEEE Transactions On Biomedical Engineering,* vol. 54, no. 8, pp. 1472-1479, 2007.

[59] J. Martinek, P. Klco, M. Vrabec, T. Zatko, M. Tatar and M. Javorka, “Cough Sound Analysis,” *Acta Medica Martiniana,* vol. 13, no. Suppl 1, pp. 15-20, 2013.

[60] J. Ye, T. Kobayashi, M. Murakawa and T. Higuchi, “Robust Acoustic Feature Extraction For Sound Classification Based On Noise Reduction,” in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP),* 2014.

[61] T. Drugman, J. Urbain and T. Dutoit, “Assessment Of Audio Features For Automatic Cough Detection,” in *19th European Signal Processing Conference (EUSIPCO 2011),* Barcelona, 2011.

[62] S. H. Shin, T. Hahimoto and S. Hatano, “Automatic Detection System for Cough Sounds as a Symptom of Abnormal Health Condition,” *IEEE Transactions On Information Technology In Biomedicine,* vol. 13, no. 4, pp. 486-493, 2009.

[63] M. You, Z. Liu, C. Chen, J. Liu, X. H. Xu and Z. M. Qiu, “Cough detection by ensembling multiple frequency subband features,” *Biomedical Signal Processing and Control,* vol. 33, pp. 132-140, 2017.

[64] M. J. Doherty et al., “The acoustic properties of capsaicin-induced cough in healthy subjects,” *European Respiratory journal,* vol. 10, pp. 202-207, 1997.

[65] K. Kosasih, U. R. Abeyratne and V. Swarnkar, "High Frequency Analysis of Cough Sounds in Pediatric Patients with Respiratory Diseases," in *34th Annual International Conference of the IEEE EMBS*, San Diego, 2012.

[66] Q. Z. Wu, I. C. Jou and S. Y. Lee, "On-line signature verification using LPC cepstrum and neural networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 27, no. 1, pp. 148-153, 1997.

[67] T. Drugman et al., "Audio and Contact Microphones for Cough Detection," *INTERSPEECH*, 2012.

[68] A. Imran et al., "AI4COVID-19: AI Enabled Preliminary Diagnosis for COVID-19 from Cough Samples via an App," *Informatics in Medicine Unlocked*, vol. 20, 2020.

[69] J. M. Liu, M. You, Z. Wang, G. Z. Li, X. Xu and Z. Qiu, "Cough Detection Using Deep Neural Networks," in *IEEE International Conference on Bioinformatics and Biomedicine*, 2014.

[70] J. M. Liu, M. You, Z. Wang, G. Z. Li, X. Xu and Z. Qiu, "Cough event classification by pretrained deep neural Network," *BMC Medical Informatics and Decision Making*, vol. 15(Suppl 4), 2015.

[71] I. D. Miranda, A. H. Diacon and T. R. Niesler, "A Comparative Study of Features for Acoustic Cough Detection Using Deep Architectures," in *41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2019.

[72] M. Sterling, H. Rhee and M. Bocko, "Automated Cough Assessment on a Mobile Platform," *Journal of Medical Engineering*, vol. 2014, 2014.

[73] Y. A. Amrulloh, U. R. Abeyratne, V. Swankar and R. Triasih, “Automatic cough segmentation from non-contact sound recordings in pediatric wards,” *Biomedical Signal Processing and Control,* vol. 21, pp. 126-136, 2015.

[74] V. Swankar, U. R. Abeyratne, Y. Amrulloh, C. Hukins, R. Triasih and A. Setyati, “Neural Network based algorithm for automatic identification of cough sounds,” in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC’13),* Osaka, 2013.

[75] Liu, J M et al., “Cough Signal Recognition With Gammatone Cepstral Coefficients,” in *ChinaSIP,* 2013.

[76] P. M. Olia, P. Sestini and M. Vagliasindi, “Acoustic parameters of voluntary cough in healthy non-smoking subjects,” *Respirology,* vol. 5, no. 3, pp. 271-275, 2000.

[77] G. Yang, H. Mo, W. Li, L. Lian and Z. Zheng, “[The endpoint detection of cough signal in continuous speech],” *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi,* vol. 27, no. 3, 2010.

[78] K. Kosasih, U. R. Abeyratne, V. Swarnkar and R. Triasih, “Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis,” *IEEE Transactions On Biomedical Engineering,* vol. 62, no. 4, pp. 1185-1194, 2015.

[79] T. H. Dat, N. W. Terence, J. W. Dennis and L. Y. Ren, “Generalized Gaussian Distribution Kullback-Laiber Kernel For Robust Sound Event Recognition,” in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP),* 2014.

[80] J. Knocikova, J. Korpas, M. Vrabec and M. Javorka, “Wavelet Analysis Of Voluntary Cough Sound In Patients With Respiratory Diseases,” *Journal of Physiology and Pharmacology,* vol. 59, no. Suppl 6, pp. 331-340, 2008.

[81] C. Bales et al., "Can Machine Learning Be Used to Recognize and Diagnose Coughs?," in *IEEE International Conference on E-Health and Bioengineering*, 2020.

[82] F. Barata, M. Weber, P. Tinschert and E. Flelsch, "Towards Device-Agnostic Mobile Cough Detection with Convolutional Neural Networks," in *IEEE International Conference on Healthcare Informatics (ICHI),* Xi'an, 2019.

[83] L. Kvapilova et al., "Continuous Sound Collection Using Smartphones and Machine Learning to Measure Cough," *Digit Biomark,* vol. 3, no. 3, pp. 166-175, 2019.

[84] A. Murata, N. Ohota, A. Shibuya, H. Ono and S. Kudoh, "New Non-invasive Automatic Cough Counting Program Based on 6 Types of Classified Cough Sounds," *Internal Medecine,* vol. 45, no. 6, pp. 391-397, 2006.

[85] J. Monge-Alvarez, C. Hoyos-Barcelo, P. Lesso and P. Casaseca-de-la-Higuera, "Robust Detection of Audio-Cough Events Using Local Hu Moments," *IEEE Journal of Biomedical and Health Informatics,* vol. 23, no. 1, pp. 184-196, 2019.

[86] C. Brown et al., "Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data," in *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2020.

[87] A. A. Abaza et al., "Classification of voluntary cough sound and airflow patterns for detecting abnormal pulmonary function," *Cough,* vol. 5, no. 8, 2009.

[88] J. Ye, T. Kobayashi, M. Murakawa and T. Higuchi, "Kernel Discriminant Analysis for Environmental Sound Recognition Based on Acoustic Subspace," in *IEEE International Conference on Acoustic, Speech and Signal Processing*, 2013.

[89] A. Teyhouee and N. D. Osgood, "Cough Detection Using Hidden Markov Models," *Social, Cultural, and Behavioral Modeling,* 2019.

[90] P. Klco, M. Smetana, M. Kollarik and M. Tatar, "Application of Octonions in the Cough Sounds Classification," *Advances in Applied Science Research,* vol. 8(2), pp. 30-37, 2017.

[91] H. Chatrzarrin, A. Arcelus, R. Goubran and F. Knoefel, "Feature extraction for the differentiation of dry and wet cough sounds," *2011 IEEE International Symposium on Medical Measurements and Applications,* p. 162–166, 2011.

[92] S. A. Walke and V. R. Thool, "Differentiating Nature of Cough Sounds in Time Domain Analysis," in *International Conference on Industrial Instrumentation and Control (ICIC)*, 2015.

[93] A. Marata, Y. Taniguchi, Y. Hashimoto, Y. Kaneko, Y. Takasaki and S. Kudoh, "Discrimination of Productive and Non-Productive Cough by Sound Analysis," *Internal Medicine,* vol. 37, no. 9, pp. 732-737, 1998.

[94] A. Van Hirtum and D. Berckmans, "The fundamental frequency of cough by autocorrelation analysis.," in *EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event*, Aalborg, 2001.

[95] J. Smith and A. Woodcock, "New Developments in the Objective Assessment of Cough," *Lung,* vol. 186, pp. 48-54, 2008.

[96] G. A. Fontana and J. Widdicombe, "What is cough and what should be measured?," *Pulmonary Pharmacology & Therapeutics,* vol. 20, no. 4, pp. 307-318, 2007.

[97] Kelsall, A et al., “How to quantify coughing: correlations with quality of life in chronic cough,” *European Respiratory Journal*, vol. 32, pp. 175-179, 2008.

[98] Ç. Bilen, G. Ferroni, F. Tuveri, J. Azcarreta and S. Krstulović, “A Framework for the Robust Evaluation of Sound Event Detection,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, 2020.

[99] D. Wyatt, T. Choudhury and . J. Bilmes, “Conversation detection and speaker segmentation in privacy-sensitive situated speech data,” in *INTERSPEECH 2007, 8th Annual Conference of the International Speech*, Antwerp, 2007.

[100] K. J. Piczak, “ESC: Dataset for Environmental Sound Classification,” Harvard Dataverse, 2015. [Online]. Available: https://doi.org/10.7910/DVN/YDEPUT. [Accessed 2021].

[101] J. Carletta, “Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus,” *Lang Resources & Evaluation*, vol. 41, pp. 181-190, 2007.

[102] M. Mijic, “STI-to-PB-word score relation in direct listening,” *Acta Acustica united with Acustica*, vol. 74, no. 2, pp. 143-148, 1991.

[103] IEC 60268-16, “IEC 60268-16. Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index,” *International Electrotechnical Commission - International Standard*, vol. Edition 4.0, 2011.

Chapter 7

LIST OF EMPLOYABILITY SKILLS AND DISCIPLINE SPECIFIC SKILLS TRAINING

7.1 ACCUMULATED ECTS

The following ECTS were obtained upon successful completion of the annual evaluation 2020, TU Dublin modules, and external modules.

- 7.5 ECTS for research and professional development planning.
- 30 ECTS for successful completion of employability and discipline specific training.

The next sections provide details on the accumulated ECTS.

7.2 EMPLOYABILITY SKILLS TRAINING

- Research Integrity – 5 ECTS

Module	RESM1953	**CRN**	32660
Module provider	Technological University Dublin	**Module Coordinator**	Prof. Mary McNamara (TU Dublin)
Module Description: This programme is designed to help graduate and early career researchers answer many questions that will arise as they consider how to plan, carry out and report their research with integrity, and to deal with the complex situations in which they may find themselves.			

- D-REAL Setting Sail – 5ECTS

Module	INTL1000	**CRN**	32547
Module provider	Trinity College Dublin	**Module Coordinator**	Prof. Carol O'Sullivan (TCD)
Module Description: The purpose of Setting Sail was to introduce you to what it means to embark on a PhD, to raise awareness of aspects which are important for you to consider at this point, and to ensure that all students in your cohort have the same knowledge of digital platform research and fundamental digital media research.			

- D-REAL Smaointe Summer School – 5 ECTS

Module	COMP47760	**CRN**	60519
Module provider	University College Dublin	**Module Coordinator**	Prof. Julie Berndsen (UCD)
Module Description: Smaointe ("Reflections") Summer Schools consist of two types of activities. Firstly, building on the Dagstuhl model, themed workshops on big-ideas and hot-topics in Digitally Enhanced Reality (e.g. Ethical dilemmas in Digitally-Enhanced Reality) with Smaointe topics designed in collaboration with industry partners. The summer schools will facilitate communication, interaction, knowledge and skills transfer across d-real.			

7.3 DISCIPLINE SPECIFIC SKILLS TRAINING

- Advanced Topics in Computational Intelligence – 5 ECTS

Module	COMP9000	**CRN**	30484
Module provider	Technological University Dublin	**Module Coordinator**	Dr. Robert Ross (TU Dublin)
Module Description: Weekly seminar series with discussions on new topics in Computational Intelligence. Oral and written reviews on papers in AI/ Computational Intelligence.			

- Machine Learning – 10 ECTS

Module	INTL1002	**CRN**	32548
Module provider	Coursera	**Module Coordinator**	Andrew Ng
Module Description: A broad introduction to machine learning, datamining, and statistical pattern recognition. Topics include: Supervised learning (parametric/non-parametric algorithms, support vector machines, kernels, neural networks). Unsupervised learning (clustering, dimensionality reduction, recommender systems, deep learning). Best practices in machine learning (bias/variance theory; innovation process in machine learning and AI).			

7.4 Annual Evaluation

- Annual Evaluation 2020 – 7.5 ECTS

Module	PGRE9023	**CRN**	33822
Module provider	Technological University Dublin	**Module Coordinator**	Prof. Mary McNamara (TU Dublin)
Module Description: Annual report including descriptions of: - Research carried out in 2019/2020. - The plan of the future research.			