Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

Efficient Cardiac Audio Analysis for Mobile Devices

Conor Shane Gildea

May 14, 2021

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Masters of Computer Science

Declaration

I hereby declare that this dissertation is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at `http://www.tcd.ie/calendar`.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at `http://tcd-ie.libguides.com/plagiarism/ready-steady-write`.

Signed: ______________________ Date: ______________________

Abstract

Cardiac irregularity detection commonly occurs from consultation with medical practitioners. Given the prevalence of mobile devices capable of heart sound recording, this project implements and investigates digital signal process techniques, based on intrinsic characteristics of heart sounds, for on-device heart analysis.

The project implements systems to complete S1/S2 ('Lub'/'Dub') segmentation and heart sound category classification utilizing the PASCAL Heart Sounds Challenge, as the testing dataset. Mobile device heart recordings are completed by positioning the mobile's microphone against a suitable heart auscultation location. Successful completion of automated heart analysis, based on these recordings, would provide earlier detection tooling for cardiac irregularities, without the initial need for a medical practitioner.

For S1/S2 segmentation, the project utilizes Harmonic Percussive Source Separation, in combination with Onset Detection and 12 Chroma features inputted into a Support Vector Machine (SVM), of 31.7 KB in size, to complete S1/S2 segmentation from mobile recorded heart sounds. The proposed solution for S1/S2 segmentation performed with $\approx 23\%$ less Total Error than the challenge winning SLAC Stanford solution, on the provided mobile dataset.

For heart sound classification, an individual classification solution is proposed for each testable category. To classify heart murmurs, 6 parameters of summarised tonnetz based on the harmonics of each test heart audio, are inputted into an SVM of 4.1 KB in size. To classify extra heart sounds (double S1/S2), prior S1/S2 segmentation is utilised. To classify extrasystoles (irregular extra heart sounds), a heart rate variability metric for root mean square of successive differences (RMSSD) between heartbeats is evaluated. To classify artifacts (non-heart) sounds, autocorrelation and Durbin Watson statistics are utilised along with S1/S2 detection within a minimum expected heatbeat range. The proposed heart sound classification within the mobile test dataset achieved the highest F-Score for Heart Problem irregularities (Murmurs and Extra Heart Sounds) compared to prior winners, for other metrics within the dataset the solution performed near par.

The dissertation also evaluates the limitations, challenges and testing criteria utilised, contributing to a discussion of the viability for medical usage of mobile recorded heart sounds and subsequent automated heart analysis of mobile recordings.

Acknowledgements

Special thanks to my supervisor, Fergal Shevlin, for agreeing to supervise such an unusual project and providing continuous expert guidance and support. Thanks to my second reader, Meriel Huggard, for providing vital advice and expertise to grade this project.

Thanks to Peter Bentley, Glenn Nordehn, Miguel Coimbra, Shie Mannor and Rita Getz for the creation of the PASCAL Classifying Heart Sounds Challange at the center of this dissertation.

A *heartfelt* thanks goes towards all my peers and teaching staff in the Trinity College Dublin, School of Computer Science and Statistics who made the last five years utterly transformative.

Finally a huge thanks to my family and friends for supporting me throughout this intense course.

Contents

List of Figures

List of Tables

Nomenclature

S1	The first heart sound 'lub' of 'lub-dub' occurring from atrioventricular valve closure (15).
S2	The second heart sound 'dub' of 'lub-dub' occurring from semilunar valve closure (15).
Heart Murmur	"whooshing, roaring, rumbling, or turbulent fluid" noises of the heart between either "lub" and "dub" or "dub" and lub" (14).
Extra Heart Sound	Regularly occurring extra heart sounds e.g. "lub-lub dub" or a "lub dub-dub" (14).
Extrasystoyle	Irregularly occurring extra heart sounds e.g. "lub-lub dub" or a "lub dub-dub" (14).
Artifact	Non-heart sounds within the challenge dataset (14).
ECG	Electrocardiogram - Tracing of the changes of electrical potential during a heartbeat by an electrocardiograph (16).
PCG	Phonocardiogram - High fidelity graphic representation of heart sounds made by means of a microphone, amplifier, and recording equipment (17).
HPSS	Harmonic Percussive Source Separation (18).
SVM	Support Vector Machines: A supervised machine learning algorithm used for classification of data with N-dimensional points (19).
Chroma	Tonal feature descriptor for the tonal content of a musical audio signal in a condensed form (20).
Tonnetz	Tonal centroid features of harmonic changes within audio. Tonnetz produces 6-dimensional basis features, two-dimensional coordinates (x-axis, y-axis) for each; perfect fifth, minor third, and major third (11).
.xlxs	Excel Open XML Spreadsheet File Format (21).
.wav	Waveform Audio file format for uncompressed lossless audio produced by IBM and Microsoft (22).
.aif/.aiff	Audio Interchange Format File Format for uncompressed lossless audio produced by Apple (22).

1 Introduction

1.1 Motivation

Hearts have a 100% uptime requirement or fatality could occur. Hearts are a single point of failure (based in one geographical location, with no load balancing). Hearts lack sufficient monitoring or alerting, to complete efficient emergency incident response. Based on these characteristics, if the heart was a technical service, it would not be utilized in a production environment with a strong likelihood of continued success. Thus it is no wonder why, based on heart characteristics, 31% of deaths worldwide are due to cardiovascular diseases [1].

In an ideal world, hearts would be depreciated and replaced with an engineered heart with more optimal characteristics. However in lieu of drastic changes, this project aims to develop low-resource cardiovascular analysis algorithms based on widely accessible mobile microphone audio, enabling easier monitoring and alerting of hearts.

A low resource requirement is enforced due to access to hardware and privacy considerations. >75% of Cardiovascular deaths occur in low- and middle-income countries, thus it is beneficial for the algorithms in this project to be developed with low-resource devices in mind suitable for both low- and middle-income countries.

The heart displays concerning privacy characteristics. Heart audio can display vulnerabilities in an individual's health condition. Whilst living, humans are unable to opt-out of their heart being audible. Heart audio is difficult to alter even with significant lifestyle changes, meaning heart audio could be used as an additional method of permission-less identity verification. Thus ideally, this medically sensitive data should be processed efficiently with low-resource mobile devices.

1.2 Research Objectives

The central objectives of this project are:

1.2.1 Develop low-resource Heart Sound Segmentation

In a normal healthy heart of adults, two fundamental sounds heard are S1 ("lub") and S2 ("dub"). S1 occurs from atrioventricular valves closure signifying the beginning of systole (ventricles pumping blood) phase of the cardiac cycle. S2 occurs from semilunar valves closure signifying the beginning of diastole (ventricles filling with blood) phase of the cardiac cycle (15).

The low-resource heart sound segmentation algorithms will be developed with explainable justification for the computational techniques used based on characteristics of heart sounds.

Successful heart sound segmentation enables detection of the heartbeat, arrhythmia (rate or rhythm issues with the heart) and easier classification of additional heart sounds in each cardiac cycle e.g. "lub...dub.dub" or murmurs.

1.2.2 Develop low-resource Heart Sound Classification

Alongside the normal cardiac cycle sounds, additional heart sounds events can occur, including murmur (turbulent sounds of blood), extra sounds of S1/S2 and arrhythmia.

Whilst detection of these events in a heart audio does not necessarily mean a serious health issue, earlier detection can prompt a medical practitioner to appropriately investigate.

Given the of mobile phone recorded audio, classification of artifacts (non-heart) audio would be beneficial to highlight if the heart audio recording is suitable.

The low-resource heart sound classification algorithms will also be developed with explainable justification of the computational techniques used, based on explainable properties of these events.

1.3 Technologies Used

1.3.1 Python

Python is an open source high-level general-purpose programming language (23). Access to thousands of third-party modules through the Python Package Index (PyPI) and Python's language accessibility makes Python popular language in academic research. Python is the main language utilized in this project.

1.3.2 Librosa

Librosa is an open source Python package for music and audio analysis. Librosa provides tooling to aid the creation of music information retrieval systems (24). In this project, Librosa is utilized for core music information retrieval operations including; Harmonic Percussive Source Separation and Onset Detection in S1/S2 ('Lub'/'Dub') segmentation, Tonnetz Feature Extraction for murmur classification and Autocorrelation for artifact classification.

1.3.3 NumPy

NumPy is an open source Python package for scientific computing. NumPy provides multidimensional array objects with a variety of fast array operations (25). In this project, NumPy is utilized for mathematical and reshaping array operations, during all feature extractions.

1.3.4 SciPy

SciPy is an open source Python package for mathematics, science and engineering. SciPy is built on top of NumPy, offering additional functions for optimization, statistical, signal processing (26). SciPy's Butterworth Bandpass filter is utilized in this project.

1.3.5 Scikit-Learn

Scikit-learn is an open source Python package for advanced machine-learning algorithms. Scikit-learn also provides model selection, classification reporting and model metrics, to aid in appropriate machine-learning application (27). Scikit-learn is utilized in this project for Support Vector Machines (SVM) for S1/S2 Classification and Murmur Classification.

1.3.6 statsmodels

statsmodel is an open-source Python package providing classes and functions for statistical models, tests and data exploration (28). In this project, statsmodel is utilized for Durbin Watson Statistic calculation for auto-correlation of artifact audio.

1.3.7 Pickle

Pickle is a standard Python package, allows for binary serializing and de-serializing of python object structure (29). Pickle's usage is necessary in the project, to store pre-trained efficient machine-learning models, for later predictions.

1.3.8 Jupyter Notebooks

Jupyter Notebooks is an open-source project to provide interactive data science and scientific computing (30). In this project, Jupyter Notebooks through a web browser, were utilized for all Python project development.

1.3.9 Audacity

Audacity is open-source cross-platform application for audio editing and recording. Audacity allows for various audio effects applications (31). During project investigations, Audacity was utilized for efficient effect applications.

1.3.10 TinkerCad and Utilmaker Cura

TinkerCad is an easy-to-use online 3D design software (32). Utilmaker Cura is 3D printing software with an open-source silencing engine for 3D prints (33). TinkerCad and Utilmaker Cura were used in the design and printing of a 3D printed Stethoscope to Phone attachment during investigations of this project.

1.3.11 Android Studio

Audio Studio is the official Integrated Development Environment (IDE) for Android application development (34). During project investigations, Android studio was used for the creation of an optimal heart audio recording application.

2 Background

2.1 Audible Heart Fundamentals

To analyse heart audio, knowledge of core audible heart sounds is necessary:

2.1.1 S1/S2 ('Lub'/'Dub') Sounds

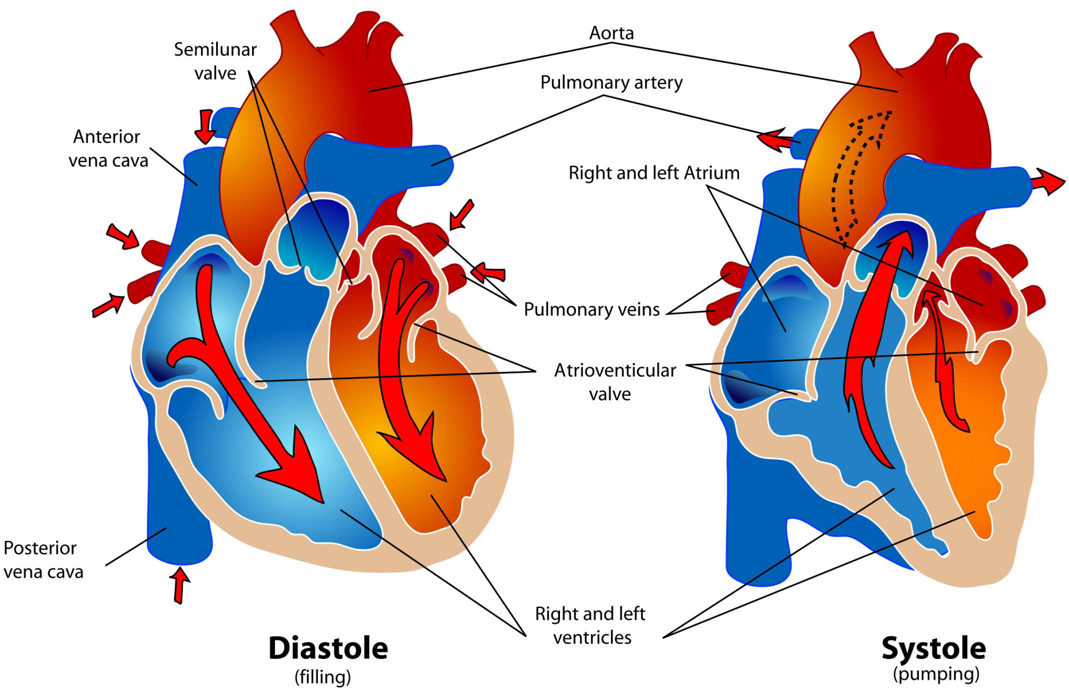


Figure 2.1: Healthy Heart in Diastole and Systole (1)

The cardiac cycle of a normal healthy heart consists of two phases, **diastole** and **systole**. During diastole, the atrioventricular valves open, this allows for blood to enter the heart from the pulmonary veins. As the systole phase begins, the atrioventricular valves close to prevent blood further filling the ventricles of the heart. This **atrioventricular valves closure causes the 'lub'** sound, formally known as S1 (15).

With the systole phase continuing, the semilunar valves open, this enables blood to leave the ventricles of the heart. When systole phase ends, these semilunar valves close. This **semilunar valves closure causes the 'dub'** sound, formally known as S2 (15). Then the diastole phase begins and the cardiac cycle starts again.

S1 and S2 the fundamental audible heart sounds. There is additional sounds S3, referred to as "ventricular gallop" due to ventricular filling and S4, referred to as "atrial gallop" due to atrial filling (15). These sounds usually indicate a stressed heart however are extremely low-pitched (15) thus are hard to detect even in clinical settings, due to this S1 and S2 segmentation remain the sole focus of this dissertation.

The pattern of S1 and S2 can be described with the illustration (14):

lub. . . . . .dub. . . . . lub. . . . . .dub. . . . .

2.1.2 Extra Heart Sounds

Extra heart sounds are identifiable from the presence of regularly occurring repeated heart sound of either S1 or S2 e.g. "lub-lub dub" or a "lub dub-dub" (14).

The pattern of extra heart sounds can be described with the illustration (14):

lub. . . . .**dub.dub**. . . . . .lub. . . . .**dub.dub**

or

lub.lub. . . . . .dub. . . . .**lub.lub**. . . . . .dub

2.1.3 Extrasystole

Extrasystole are extra beats or contractions of the heart. The difference between extra heart sounds and extrasystoles, is that extrasystoles are not regularly occurring extra heart sounds. In certain cases extrasystoles can be caused by heart diseases, thus earlier dection can lead to a more effective treatment (14).

The pattern of extrasystole hearts can be described with the illustration (14):

lub. . . . . .dub. . . . .**lub.lub**. . . . . .dub.. . .

or

lub. . . . . .dub. . . . . .lub. . . . .**dub.dub**

2.1.4 Murmurs

Heart murmurs is the terminology to describe "whooshing, roaring, rumbling, or turbulent fluid" noise occurring between either S1 to S2 and S2 to S1 within the heart however not during S1 or S2 (14). Based on the understanding of the heart this is justifiable given S1 and S2 indicate the movement of valve closures, not immediate blood movements.

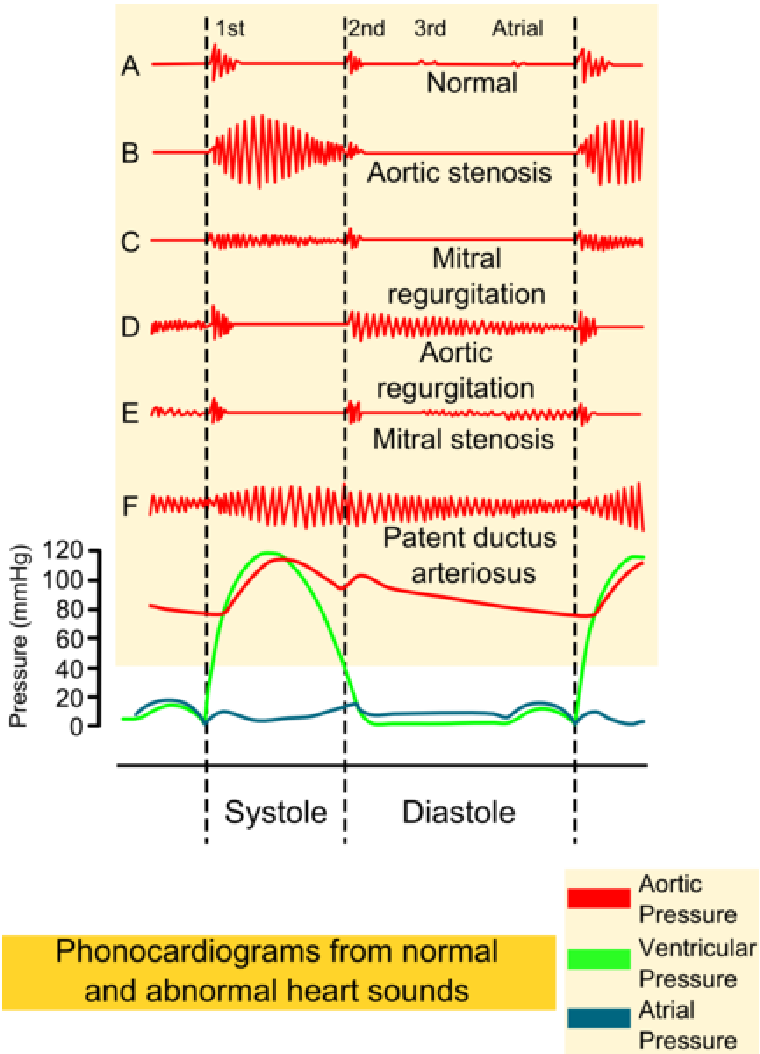


Figure 2.2: Phonocardiograms of normal and murmur hearts with blood pressure diagram (2)

Murmurs arise from two sources; valvular abnormality or abnormal blood flow pathway (15). Valvular abnormality means difficulty experience with complete valve closure within the cardiac cycle. Abnormal blood blood can be congenital or acquired (15). Murmurs can be a symptom of many heart irregularities which can be serious, thus earlier detection can lead to a better outcome.

2.2 Audibility of Mobile Recorded Hearts

2.2.1 Heart Frequency Range

The most challenging aspect from the onset of this project is the issue of amplitude and low frequency of heart sounds. Human hearing can detect sound vibrations from 16-18,000 Hz (35) however the audible range is dependant on the sound intensity. Thus the frequency range of heart sounds with are audible is approximately between 40-500 Hz (35), depending on pressure intensity.

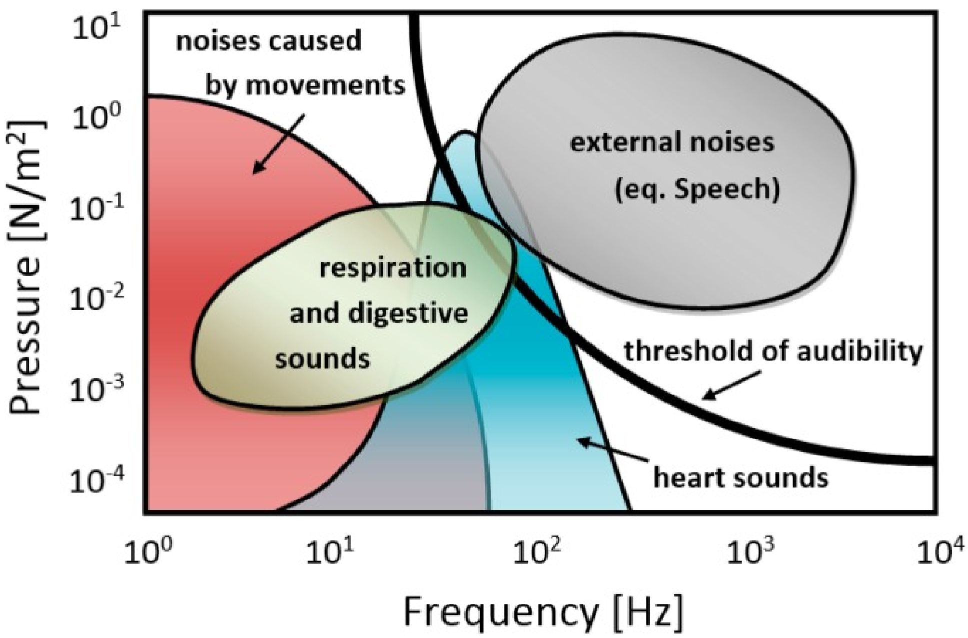


Figure 2.3: Threshold of Heart Sound Audibility (3)

The threshold of audibility of heart sounds explains why some medical practitioners use palpitation (touching) rather than auscultation (listening) when diagnosing low-pitched sounds such as S3 and S4 (36).

The threshold of heart sound audibility highlights the challenging nature of this project. Heart sounds, even in clinical conditions, can be challenging to detect and diagnose, let alone with environmental background noise, movement of the mobile device from the user and the frequency response from all manufactured mobile microphones.

2.2.2 Frequency Response of Mobile Microphones

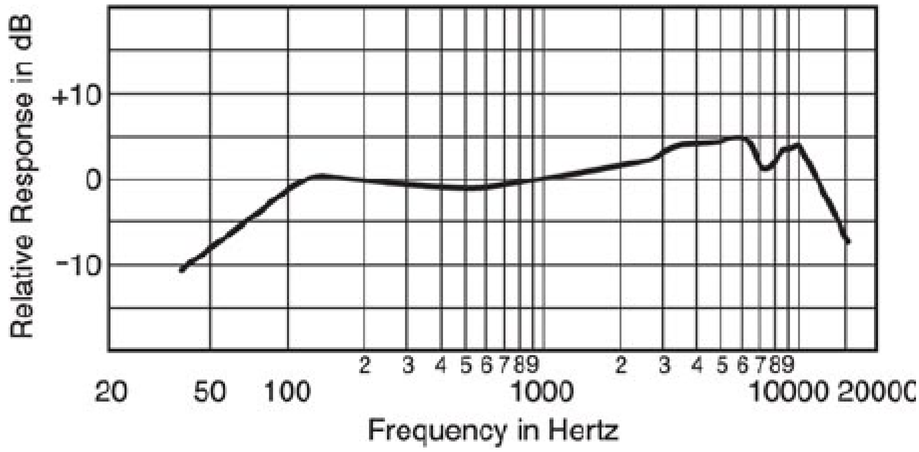


Figure 2.4: Mobile Microphone Frequency Response (4)

The frequency response of a microphone is the largest factor to determine the sound signature of a microphone (37). A microphone can have a flat or shaped response. A flat response means the microphone is equally sensitive across frequency ranges, whilst a shaped response microphone is more sensitive to certain ranges of frequency (37), this can limit the audibility of other frequency ranges.

The typical frequency response of mobile microphone, as expected, is shaped to favour high frequency ranges (4) individuals talk in, whilst reducing the frequency response of lower frequency ranges to prevent movements of the device being picked up by the microphone during communication.

For the sake of this project, typically the frequency response of mobile microphones work against the objectives of this project. A typical mobile microphone frequency response by design tries to limit the audibility of the frequencies this project relies on to detect and diagnose heart sounds.

The frequency response in combination with the threshold of audibility of heart sounds, highlights the largest challenge with this project from the offset. Even with sophisticated algorithmic approaches, at both hardware levels, human hearts and mobile devices, the audibility frequency range is challenging. Thus limitations from the audibility from mobile recorded hearts need to remain considered throughout any research in this topic area.

2.3 Classifying Heart Sounds Challenge

2.3.1 Introduction

Given the medical nature of this project, a source of truth dataset for training and testing was certainly a requirement for this project to proceed, given the author of this dissertation is a Computer Science Masters student.

The Classifying Heart Sounds Challenge (14), sponsored by PASCAL (high-tech and AI focused magazine), was discovered, which provided access to labelled S1/S2 dataset for normal heart sounds and labelled heart sound classifications for multiple categories, for both training and testing.

2.3.2 Background on Challenge

The challenge was publicized on 1st November 2011 with an extended deadline for submission of results till 13th April 2012, with winners announced on 24th April with a workshop at AISTATS (14). The timeline of the challenge is an important consideration as it gives scope to the prominent technical techniques at that time. For example, certain techniques utilised within this project's implementation only were publish relatively recently, such as Harmonic Percussive Source Separation in 2010 (18).

The challenge dataset provided .wav and .aif/.aiff file formatted audio for 832 audio samples, across two different datasets (14).

2.3.3 Dataset A - Mobile: iStethoscope Pro iPhone App

Peter J Bentley created an iPhone application, iStethoscope Pro, which received widespread media attention, which enabled the use of the iPhone microphone as a stethoscope, providing real-time amplifiers and filters to improve the audibility of heart sounds (38).

Dataset A consisted of heart sounds acquired from the iStethoscope Application. This was ideal for the scope of this project. The audio was 44100 Hz sampling rate and feature real world data collection from mobile microphones.

Dataset A was split into 5 folders:

Normal, Murmur, Extra Heart Sound, Artifact (non-heart) Sounds and Unlabelled.

2.3.4 Dataset B - DigiScope

Dataset B was acquired from usage of a digital stethoscope, DigiScope. This data was acquired in clinical trials within hospitals and featured a lower sampling rate of 4000 Hz. Dataset B's feature vary vastly from Dataset A's of high sampling rate but high noise mobile devices.

Dataset B was split into 4 folders:

Normal, Murmur, Extrasystole, Unlabelled.

2.3.5 Challenge 1: Heart Sound Segmentation

For both datasets normal heart sounds, Challenge 1 required the expected location of S1/S2 ('lub/'dub') to be recorded within an .xlsx spreadsheet for evaluation. This location was recorded in samples, which required sampling rate manipulation in implementation given the differing sampling rates (44100 Hz and 4000 Hz) of both datasets.

2.3.6 Challenge 2: Heart Sound Classification

Challenge 2 required heart sound classification of specified unlabelled dataset sounds, into the correct corresponding category from the provided labelled training data. For Dataset A, classifying into; 'Normal', 'Murmur', 'Extra Heart Sound' and 'Artifact'. For Dataset B, classifying into; 'Normal', 'Murmur', 'Extrasystole'.

2.3.7 Evaluation Method

The challenge also provided two locked .xlsx file formatted spreadsheets, for which the results of both challenges should be outputted. The .xlsx evaluation sheet provided automatic evaluation of proposed answers locally. This was a highly beneficial resource as allowed for locally testing without any internet connection, however the security of a locked .xlsx for testing evaluation will be discussed in subsequent chapters.

2.3.8 Results and Previous Contestants

The challenge also had prior results (39) for project comparison from three well-regarded finalist universities groups (ISEP/IPP Portugal, CS UCL and SLAC Stanford). It is worth noting SLAC Stanford completed Challenge 1, however not Challenge 2. SLAC Stanford however did produce the best result by a substantial margin, within the challenge's deadline, for S1/S2 Heart Sound Segmentation in Dataset A.

3 State of the Art

The State of the Art discusses previous attempted techniques to the Classifying Heart Sounds Challenge (14).

3.1 S1/S2 ('Lub'/'Dub') Segmentation

SLAC Stanford's winning result paper for S1/S2 Segmentation was unfortunately not available online anymore, due to this, the State of the Art discusses ISEP/IPP Portugal's system to complete S1/S2 Segmentation.

3.1.1 S1/S2 Detection

ISEP/IPP Portugal's S1/S2 detection examines a normalized average Shannon energy curve to identify peaks which the system would consider as S1/S2 detections.

The justification for usage of normalized average Shannon energy to identify peaks is well-founded in academic research given that normalized average Shannon energy "attenuates the effect of low value noise and makes the low intensity sounds easier to be found" (40). Making the properties of normalized average Shannon energy for peak detection of S1/S2 heart sounds strongly justifiable.

In preprocessing, a low pass filter was applied whilst maintaining the frequency components of S1/S2, then the signals were subsequently normalized to the absolute maximum length of this signal (39). Shannon Envelope of the normalized signal was calculated, to allow for Average Shannon Energy calculations in windows across the signal for use to calculate the normalized average Shannon energy (39). As noted in their explanation this technique was based on prior research utilising normalized average Shannon energy for heart segmentation (40). Lastly, across the normalized average Shannon energy, local maxima peak detection was utilised to find potential S1/S2 locations.

3.1.2 S1/S2 Classification

ISEP/IPP Portugal's S1/S2 Classification of the found S1/S2 locations was self-deemed "unsatisfactory" (39) however the solution presented is logical. It is challenging to implement their solution with successful results, as the domain is full of variability.

Assuming working S1/S2 detection, ISEP/IPP Portugal attempted to classify S1/S2 based on the distance from S2 to S1 and the distance from S1 to S2. The reasoning behind this decision was that for normal heart rates the distance from S2 to S1 was longer than the distance from S1 to S2 (41).

Heart cycle detection was then attempted with the goal to classify the longer of the two lengths corresponding to the diastolic period, with the peak on the right classified as S1 and the peak on the left as S2. Unfortunately given the variability with the datasets, this solution for S1/S2 classification wasn't as successful as hoped due to the variability in the dataset. (39)

3.2 Classifying Heart Sounds

For this section, the State of the Art will discuss the ISEP/IPP Portugal's system, which achieved the highest total precision and F-score using Multi-Layer Perceptrons. ISEP/IPP Portugal's system utilised 6 features, 4 based on distances between S1 and S2 peaks and 2 subsequent features: Rmedian (a ratio of the largest segments in the sample over the total mean) and R2 (a measure of linearity of the signal), these features were subsequently trained on an Multi-layer Perceptron for classification (39).

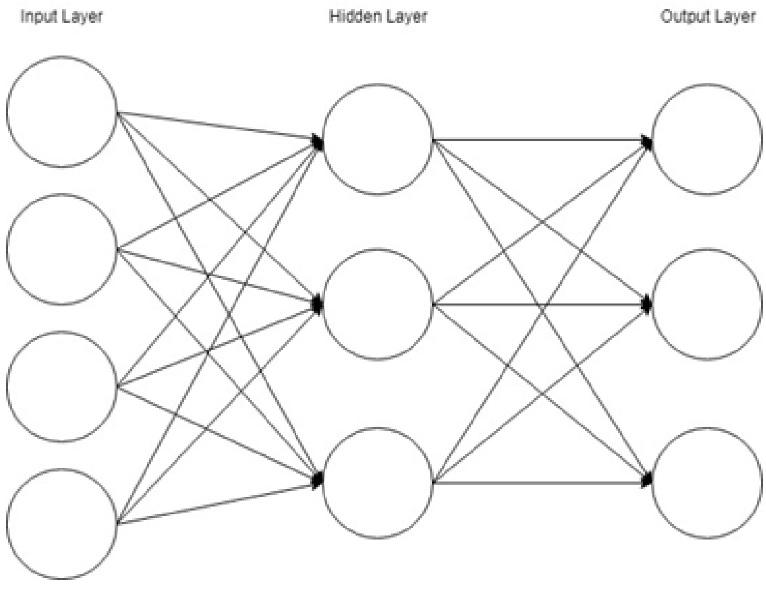


Figure 3.1: Multi-Layer Perceptron Illustration (5)

A Muli-layer Perceptron is suitable for use cases in which the inputted features are not linearly separable from a singular perceptron (5), thus making it justifiable, as an machine learning algorithm choice, for aiding in classification of heart sound types without further feature extraction, to make the data more linearly separable.

Multi-layer perceptrons form the core of neural networks. Inputs have a dot product operation applied with the weight between the input and hidden layer, this dot product is then used within an activation function at each layer, then passed to the next layer for the same repeated process at each layer until the output layer is reached (5). There can be multiple hidden layers within the Multi-layer perception. At the output layer either; a decision is made (e.g. heart sound classification) or in training back-propagation would occur to improve the Multi-layer perceptron. (5)

Whilst this solution clearly was performant for ISEP/IPP Portugal it is less justifiably explainable for why heart classifications occurred, especially in a small dataset. This project believes the techniques used should be justifiable based on the intrinsic properties of the heart, especially given the limited dataset size of 832, there is strong potential for a 'black box' approach to overfit.

3.2.1 Note on a Proposed High Precision Solution

During investigation, a paper (42) was discovered which utilised multiple machine learning approaches (include Naive Bayes, Support Vector Machines, Decision Trees, Ada-Boost, Random Forest and Gradient Boosting) for comparison to the Classifying of Heart Sounds within the PASCAL Classifying Heart Sounds Challenge. This paper was not part of the evaluated solutions given the time of production.

This paper showcased remarkably high precision values in comparison to the State of the Art however upon further investigation of the paper, it seems these precision values stated are based on averaged precision values from K-fold cross validation, not the separate evaluation on specified unlabelled test data.

"Each method above uses a K-fold cross validation of 4, which means that 25% of the dataset is used for the validation testing and 75% of the dataset is used for training. All final precision and score results listed are an average of the K-fold cross validation" (42).

Based on the understanding the final precision values were tested on a K-fold of the training dataset, not the separate unlabelled test dataset with accompanying evaluation sheet, thus this dissertation shall be not utilising the paper's results as a comparison for State of the Art.

4 Investigation

This section will outline the investigative steps taken near the offset of this project. This investigation influenced the key factors leading to the implementation of the project.

4.1 Data Acquisition - Primary

The intention of primary data acquisition was to assess the feasibility of mobile microphone recorded heart audio under the project's own controlled circumstances and assess where data acquisition alterations could be made to improve the input data, aiding the subsequent heart analysis.

4.1.1 Mobile Recording

Originally the feasibility of the heart audio recording was assessed through pre-built recording applications on a singular device. The recording output whilst under the same conditions varied dramatically between recording applications. Upon further investigation, two key factors were identified:

4.1.1.1 Application Setup

The first key factor was the configuration of recording application. Recording applications were trialed, when the environment remained constant (phone position with bottom of phone in aortic region of heart, background audio quiet and user laying down and resting) showed the recording applications produced varying quality of audio recordings. From research I established the programming implementation of the tested applications must be varied.

Mobile phones can feature multiple microphones and the application can select a microphone for the use-case (e.g. on android with Microphone Direction allows for selection of the microphone on the side of the phone as the user). The recording quality is also impacted by the selection of recording API and the selected recording settings in use with the API (e.g. on android AudioRecord (43) or MediaRecorder (44)).

MediaRecorder for instance, provides compressed audio and the recording methodology can be altered to account for different activities e.g. voice communication. MediaRecorder also allows an unprocessed microphone selection. AudioRecorder provides raw uncompressed audio from the microphone however isn't as convenient for developers, requiring managing buffers.

The differences between the two API is strong reasoning for why audio recording differs across applications. Both the APIs have justified use-cases but the audio for this project is low frequency and low magnitude. Low frequency and low amplitude audio is often filtered through selected MediaRecorder API operations e.g. normal handheld voice communication. The result of this discovery was to build my own android application, using AudioRecorder raw microphone input with the microphone direction set to the side facing the user. This provided the optimal application recording for this use-case, maintaining the low frequency components the project relies on, to solely apply the desired digital signal processing techniques.

4.1.1.2 Device Positioning

The second key factor impacting mobile recording quality was positioning. Positioning plays a crucial role in quality of the recording for heart analysis. Positioning needs to be evaluated from two areas, positioning of microphone on device and positioning of the device on the user.

4.1.1.2.1 Hardware Microphone Positioning

Modern mobile devices often feature at least two microphones, however the quantity and location of microphones can vary across the spectrum of devices. Primary data collection was completed, on the basis a microphone, was present on the bottom of the mobile device, near where a user would conventionally talk during phone calls. This microphone would also feature a hole in the phone body leading to the microphone.

Figure 4.1: Example of array of microphone, in modern phones, in an iPhone 8 (6)

This microphone hole enables a sealed medium to be formed when the microphone is placed against skin, similar to a stethoscope. In case a phone doesn't follow this convention, an external microphone would be able to be connected to the device instead. Not following the convention of a bottom microphone would be unlikely for a mobile phone given the need during calls.

During primary data collection, the microphone positioning was near ideal circumstances with the microphone on the bottom side of the device, capable of being placed directly on the user's skin, forming a sealed medium for audio to pass through.

As the primary data collection completed was successful and there were solutions for rare cases of microphone positioning changing (using an external microphone), the project moved focused on attempting to improve and analyze the signal acquired. Given the variety of devices and limited access to devices to test, it was infeasible to trial the success of this heart recording across a representative sample however this compatibility concern felt more suitable for a business with breadth to pursue, instead of an academic dissertation which is in pursuit of technical depth. In this case, to produce justifiable heart analysis algorithms.

4.1.1.2.2 User Device Positioning

The area at which the device is placed can drastically change the quality of the recording, from classifiable to inaudible. To decide on appropriate locations to place the device, the project relied on optimal stethoscope positioning acquired from established heart auscultation practices. During primary data collection, Erb's point was most commonly used for comparison of the collection, as from a singular point, Erb's point, is optimal to hear both S1 ("Lub") and S2 ("Dub") and heart sounds have the highest intensity (45).

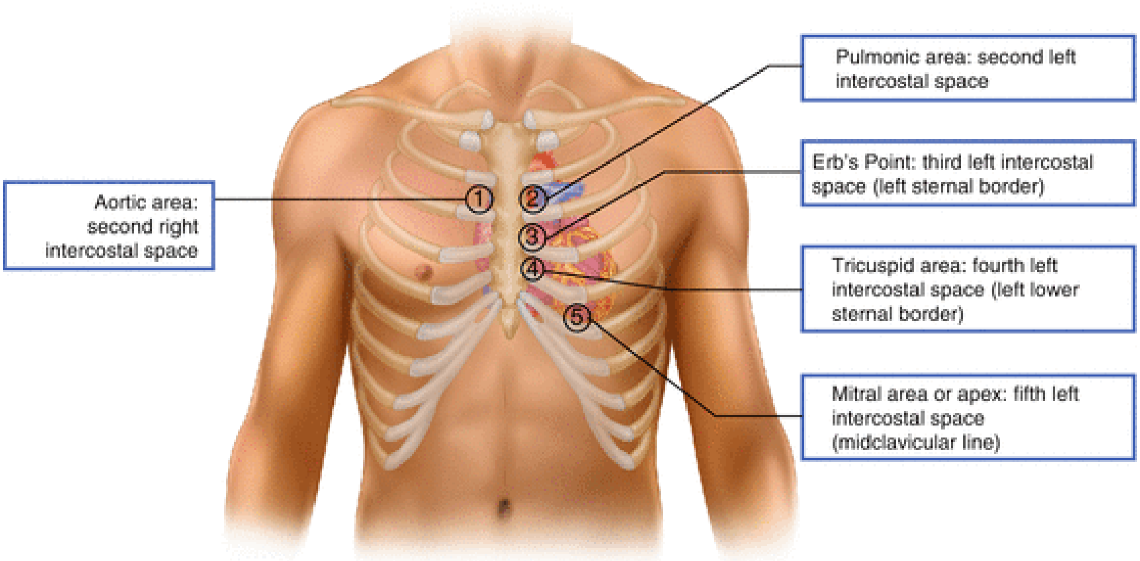


Figure 4.2: Heart Auscultation Optimal Positions (7)

In comparison to stethoscopes, mobile microphones have a smaller area of sealed medium between the skin and the device. This means the user needs to be more precise with the device positioning to acquire an adequate signal. Unlike a stethoscope, the orientation of the device also plays an important role, as microphones are often off-center on a side of a device, with a much smaller area than either side (bell or diaphragm) of a stethoscope. The device orientation must be altered to have the microphone hole in the central location the stethoscope would ordinarily be, to produce the optimal signal.

4.1.2 3D-Printed Stethoscope-to-Mobile Attachment

Traditional stethoscopes have a bell and diaphragm side, the bell (smaller diameter side) is utilised to hear low pitched sounds via low pressure of placement on the heart (15).

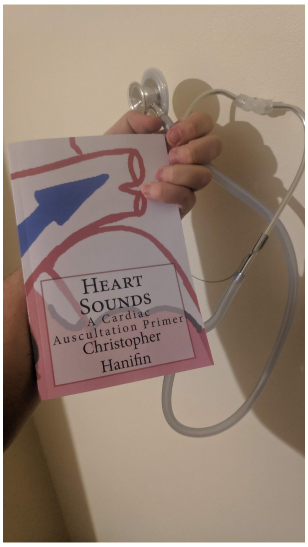


Figure 4.3: Stethoscope used during project investigations

As part of primary data collection enhancement, a 3D Printed Stethoscope to Mobile attachment was produced. The usage of the attachment was to benefit from the existing stethoscope design whilst making it convenient to record the audio file on mobile device without the need for a digital stethoscope.

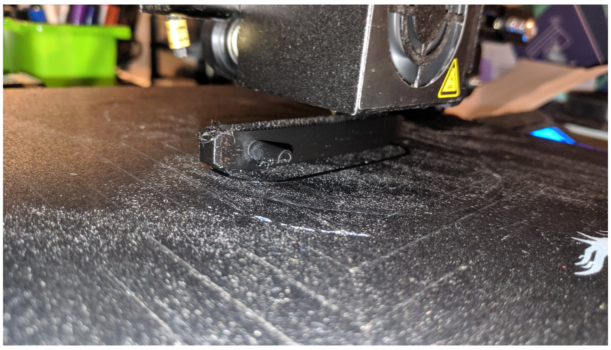


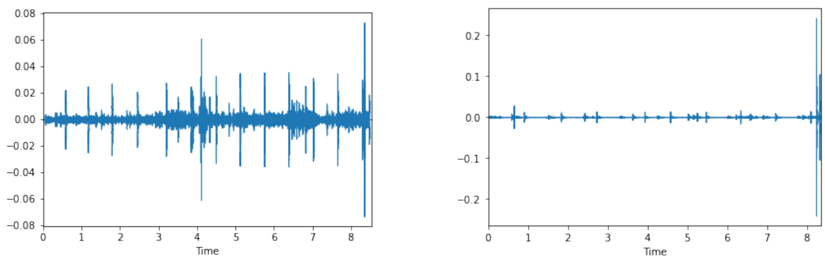
Figure 4.4: 3D Printing Self-Made Attachment to connect Stethoscope to Mobile

Despite numerous iterations on the design of this attachment along with additional sealing to aid formation of vacuum, the audio produced was found to be less audible than directly placing the device in Erb's point on the heart. Overall this discovery was of benefit for the use-case. Suitable audible quality could be achieved only with a device with a microphone placed against an adequate location near the heart. Further improvements could be made from different materials from Polylactic acid (PLA) and alternations to design however the convenience not requiring a 3D printed part and stethoscope whilst still attaining an adequate signal means a phone for recording alone is suitable for the project.

4.1.3 Key Findings from Primary Data Collection

4.1.3.1 Blood Pressure Impact

The first key finding impacting heart mobile recording was blood pressure. As would be expected the higher the blood pressure correlates to an increase in the audibility of the heart sounds whilst particularly low blood pressure samples could be less audible without amplification.



(a) Afternoon (b) Night

Figure 4.5: Audio Recordings Variablity with Day

These characteristics can be seen from primary testing. Heart sounds were recorded in the afternoon with a higher blood pressure and also in night with a lower blood pressure. The timing correlates to our bodies natural change in blood pressure throughout the course of the day (46). This finding of blood pressure impacting audibility also discussed in the literature for cases of hypertensive (medically classified high blood pressure) patients (47).

(a) Afternoon (b) Night

Figure 4.6: Blood Pressure Readings

For the design of sound based heart analysis algorithms, the impact of blood pressure on sound, highlights the need for algorithms that account for variability in heart sounds even in the same heart. Characteristics such as blood pressure, lead the project to focus on developing algorithms based on aggregated patterns across a signal to handle the variability. The audible impact of variation in blood pressure highlights the challenges faced building solutions to analyze biologically varied systems.

4.1.3.2 Stethoscope Pressure Impact

From researching heart auscultation, the project discovered the importance of applying low pressure to the surface being listened to detect low pitched sounds (15). Low pressure allows for low pitched noises to maintain enough amplitude for audibility. This has significant considerations for users utilizing their phones for heart recording, as given the size of the device compared to a stethoscope, users could be more include to press the phone firmly against the area being recorded. This poses a further challenge to the suitability of mobile recorded heart sounds for analysis.

4.2 Data Acquisition - Secondary

Before deciding on proceeding with Classifying Heart Sounds Challenge, as the core dataset for training and testing evaluation of heart sounds. The project explored the use of two other datasets, which both lead to key findings prompting further action.

4.2.1 Michigan University Heart Sounds and Murmurs dataset

Michigan University Heart Sounds and Murmurs dataset is only 23 heart recordings provided for educational purposes (48). The fidelity of the recordings provides a much greater understanding of each classification of murmur and see the importance of positioning when completing heart auscultation.

4.2.2 PhysioNet/Computing in Cardiology Challenge 2016 Database

The PhysioNet database provided the largest quantity of heart sound recordings featuring 2,435 separate recordings, collected from 1,297 mixed of both subjects with a variety of cardiac conditions and no conditions (49). Crucially parts of the dataset included electrocardiograms (ECGs) recorded simultaneously with phonocardiograms (PCG). Despite the dataset having large variety, the challenge itself was to classfy abnormal/normal hearts and classify the audio as noisy/clean, This dataset thus had the quantity but arguably not the granularity designed in a system to develop algorithms to complete heart analysis.

4.2.3 Key Findings from Secondary Data Collection

4.2.3.1 Harmonic Percussive Relationship

From sampling audio between from the Classifying Heart Sounds Challenge and the Michigan University Heart Sounds and Murmurs dataset, it became abundantly clear the importance of the relationship between harmonic and percussive components in heart sounds. Percussive components gave a clear indicator to each beat of the heart S1 and S2, whilst the harmonic components seemed vital in describing the majority of heart murmurs you would want to detect. This justification makes sense given S1 and S2 are closures of valves which is by nature a percussive action and turbulent blood flow is a harmonic action.

4.2.3.2 Background Noise

The level of background noise present in some of the sample data, made distinguishing the characteristics of the heart seem increasingly challenging. Enhancing the provided audio through additional preprocessing during implementation would be necessary to make accurate analysis of the heart audio. Particularly for iStethoscope data, the amplifying techniques within the application aided audibility of the heart sounds however also the background noise.

4.2.3.3 Relationship between ECG and Heart Sounds

Despite not selecting as a main dataset, the PhysioNet dataset (49) providing simulatenous recordings of ECG and sounds of the heart, allows for the relationship between ECG and PCG to be explored. Whilst researching the application of this dataset, was informed of the abundance of ECG algorithms to diagnosis hearts. I furthered investigated this relationship, to see could producing "Synethic ECGs" from audio, to utilize the existing ECG algorithms be a viable solution this project.

4.3 Synthetic ECGs: Audio to ECG

Electrocardiogram (ECGs) are the tracing of the changes of electrical potential during a heartbeat by an electrocardiograph (16). The benefits of an ECG include the ability to analyze the heart from various directions depending on location of the electrodes, this allows for precise diagnosis of electrical impulse faults within the heart, which in turn could produce audible faults. There is many decades of research completed in the ECG domain to analyse hearts so producing an accurate 'Synthetic ECG' would be worthwhile, if possible.

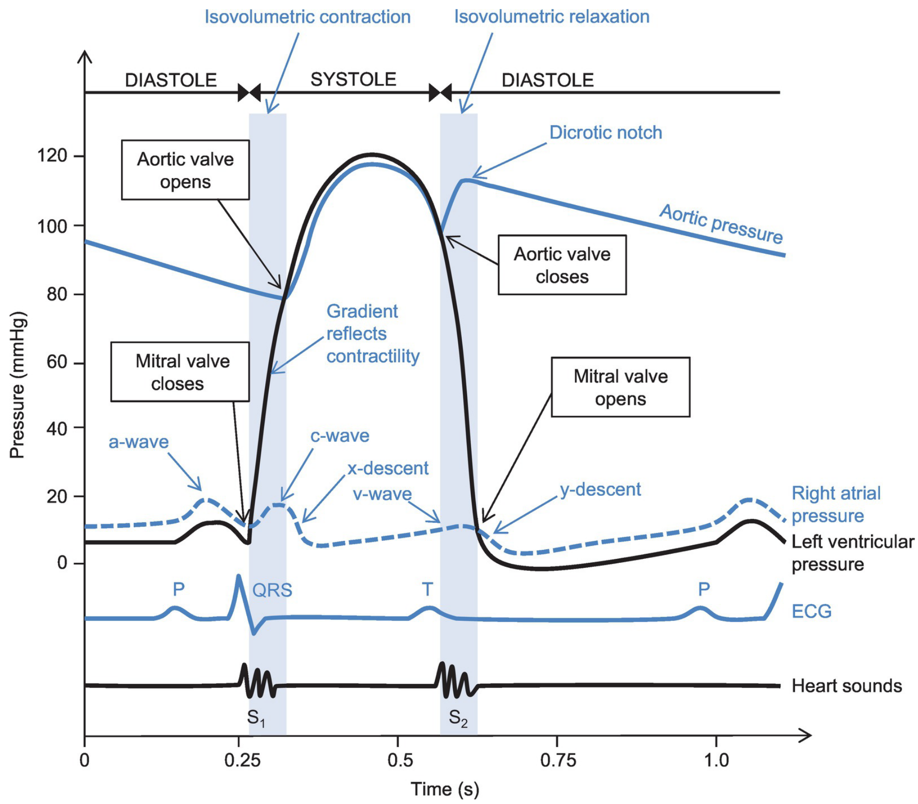


Figure 4.7: Cardiac Cycle - Pressure, ECG and Heart Sound Relationship (8)

Within a normal healthy heart, a time-series relationship is visible between certain electrical potential changes of the heart and the subsequent heart sounds. However upon further research, 'Synthetic ECGs' were not continued, for a variety of reasons:

4.3.1 Positioning

ECG output is based on the positioning of electrodes, different positioning of electrodes impacts the ECG produced. This is a huge benefit of an ECG, as with a 12-lead ECG you are capable of analysing the electrical changes of a heart from a variety of sides (50). To infer a Synthetic ECG' from audio would require inferring the polarization for all different positions of an ECG or limit the positioning to one. This would be far too much inferring of data without strong justification to be able to translate between audio and ECG.

4.3.2 Value Inferring

ECGs output the electrical changes within a heart depending on the positioning of electrodes (50), to attempt to produce a 'Synthetic ECG' would require inferring values for the electrical changes, which isn't proven viable from heart sounds alone.

4.3.3 P-wave and Relationship

The P-wave within an ECG cardiac cycle (50), before S1 heart sounds triggering the heartbeat, produces no audible sound. Thus unless inferring based on the rhythm was proceeded with, it would be impossible to accurately plot a Synthetic ECG's P-wave.

This highlights how there is a relationship between electrical changes and heart sounds however it isn't necessary a direct relationship. Not all heart sounds produce electrical changes within the heart faults and not all electrical changes within the heart faults produce clearly audible heart faults.

4.3.4 No Substantial Benefit to Leave Audio Domain

Besides for visualization purposes, there is no clear benefit to producing a 'Synthetic ECG' for this project. All heart sounds feature extractions would always occur in the audio domain and the classifications for this project could also remain with the audio domain.

However a useful finding from the study of ECG's was the usefulness of heart rate variability metrics (12), this later would be utilised for classification but within the audio domain.

Figure 4.8: Raspberry Pi connected ECG for project investigations

5 Implementation

Code examples of implementations can be found in the appendix and include .zip folder on submission.

5.1 S1/S2 ('Lub'/'Dub') Segmentation

5.1.1 S1/S2 Detection - HPSS and Onset Detection

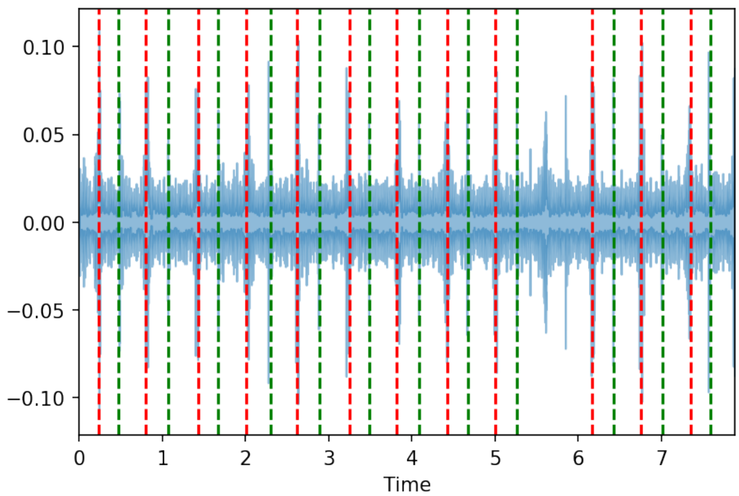


Figure 5.1: Mobile Recorded S1/S2 Segmentation Example
S1 (red) and S2 (green)

Note: The above training example within the S1/S2 Segmentation dataset shows the subjectivity of S1/S2 classification. In the above example what looks like beats between 5th and 6th second are not labelled, in the training labels. This could be human error or the subjective labelling not able to assert are the sounds S1 or S2.

5.1.2 Butterworth Bandpass

Given the targeted dataset is mobile devices, the majority of this audio contains unwanted noise across the frequencies, due to to this a Butterworth Bandpass Filter is applied. All implementation solutions utilize the Butterworth Bandpass Filter to reduce noise. For S1/S2 detection Butterworth Bandpass was applied for higher frequencies above 350 produced during S1/S2 heart sounds.

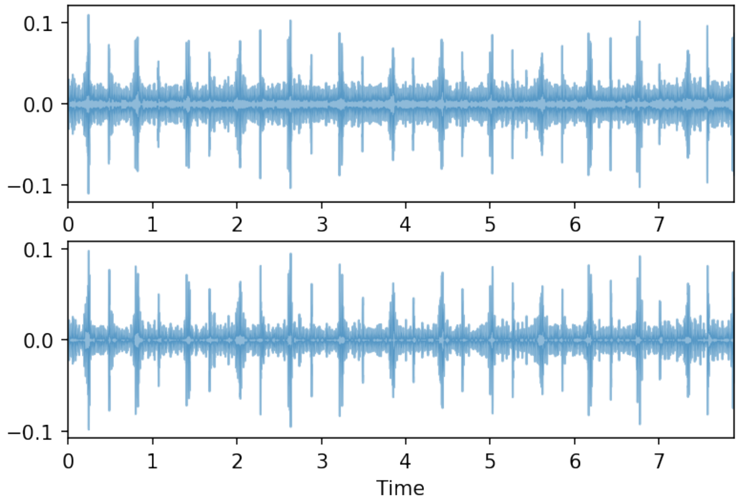


Figure 5.2: Mobile Recorded Training Example before and after Butterworth Bandpass Filtering

The justification for a Butterworth Bandpass filter is the flat as possible frequency response within the passband (51). A flat frequency response is desirable as preferably don't want shaped frequency responses further impacting the recorded heart audio quality, like within previously discussed phone microphones.

With a filtered signal to reduce noise, this project utilised the key findings during investigations, that the closure of atrioventricular/semilunar valves within the heart for both S1/S2 sounds is intrinsically a percussive action, thus the solution to detect S1 and S2 replies on the percussive action detection from applying Harmonic Percussive Source Separation and then onset detection.

5.1.3 Harmonic Percussive Source Separation

Harmonic Percussive Source Separation is a remarkably powerful, yet easy to implement technique. Harmonic Percussive Source Separation is the process from horizontal and vertical median filtered on the spectogram of an audio signal, the harmonic and percussive components can be extracted.

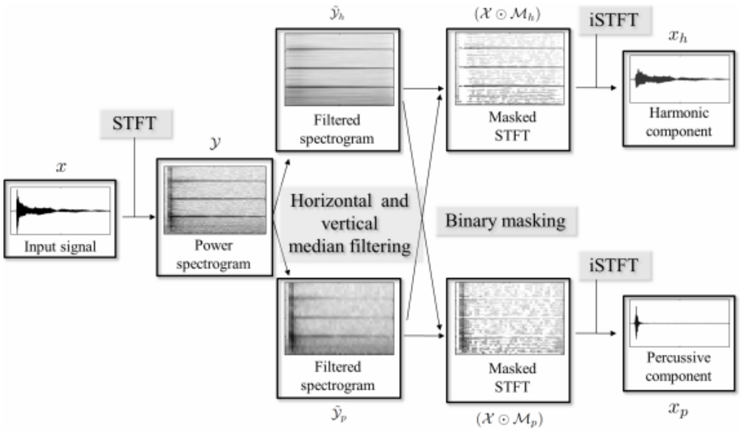


Figure 5.3: Harmonic Percussive Source Separation Demonstration (9)

For harmonics the median filtering is completed horizontally on the successive frames to suppress percussive events and enhance harmonic components. For percussive events, median filtering is completed vertically on the frequency bins to suppress the harmonic components (18).

The two filtered spectograms for both are then utilised as a mask which is applied to the original spectogram. This masking is what produces the percussive and harmonic components of the track.

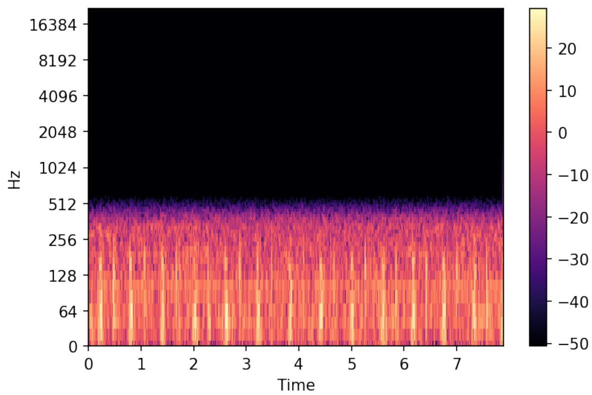
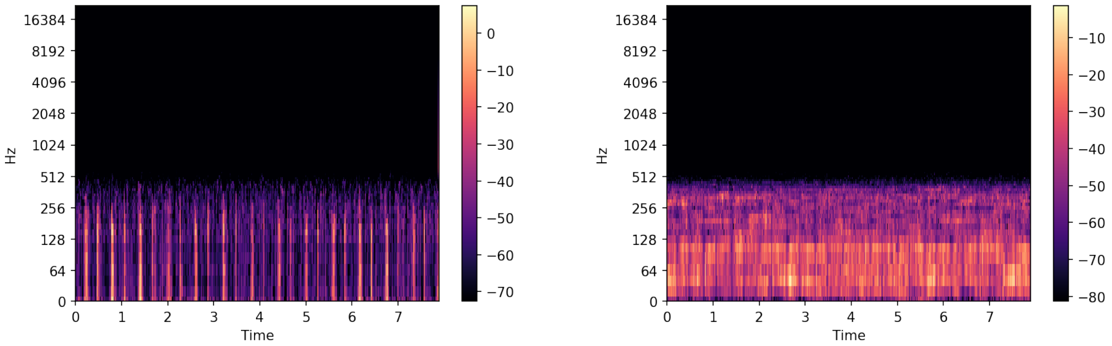


Figure 5.4: Before Harmonic Percussive Source Separation



(a) Percussive Component

(b) Harmonic Component

Figure 5.5: After Harmonic Percussive Source Separation

5.1.4 Onset Detection

Now with the filtered percussive component of the heart sounds, onset detection can be applied to give an accurate S1/S2 locations.

Onset detection is a peak picking within a spectral novelty function, this means it denotes local changes such as energy (52). However just because a novelty function is triggered does not necessarily mean it should be perceived as the onset, it could only be a higher energy triggering the novelty function corresponding to a prior onset (52).

Thus this can be resolving through the use of backtracking. Librosa has a feature called backtrack, for onset detection, which allows for the initial onset of a succession of onsets to be selected as the only onset. This reduces the possibility of duplicate S1/S2 detections of the same S1 or S2, within the given range.

A criticism of this implementation of onset detection for S1/S2 detection is that there is specified hyperparmeters that needs to be defined for how the peak picking operates. In the finalized solution, this utilizes a large pre-max and post-max values of 20 frames, this operates well for high sampling rate data however for DigiScope data these hyperparmeters of peak picking are less performant, requiring an if statement to change the parameters depending on input. As mentioned as a future work, potentially Grid Search to iterate through all viable hyperparmeters could be utilised to select the best hyper parameters for each sampling rate in usage.

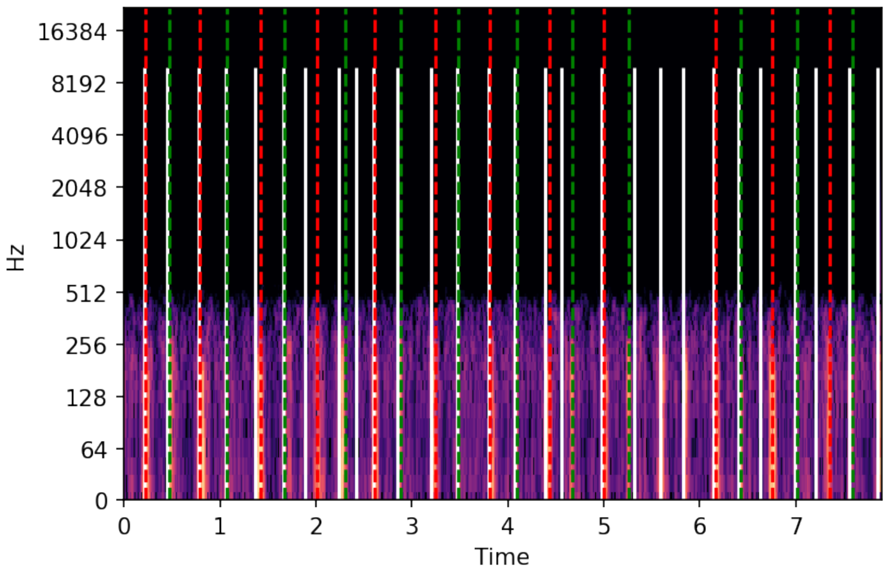


Figure 5.6: Onset detection results. In this example the training labelled points for S1 (red) and S2 (green) are displayed. The white lines indicate the onsets detected from this implementation.

5.1.5 S1/S2 ('Lub'/'Dub') Classification

Now accurate S1/S2 locations haven been detected, the solution must classify is it S1 or S2 at each location. This is a remarkably challenging problem, especially for the lower sampling rate DigiScope dataset. The system uses Chroma features to complete S1/S2 classifications, in a Support Vector Machine.

5.1.6 Chroma

Chroma is tonal feature descriptor for the tonal content of a musical audio signal in a condensed form (20). In this solution a 12-element summarising chroma vector describing the energy (52) in each pitch class is used over a window.

To evaluate is it a S1 or S2 at each S1/S2 detected location, the system takes a window of the surrounding Chroma vectors which was extracted using constant-Q transform. The system summarises the full chroma window into 12 summarised chroma values for the entire window (of 0.16 seconds). This summarised 12-element feature is then utilised for classification between S1 and S2 in a Support Vector Machine (SVM). Giving the average energy of each pitch for the 0.16 seconds window.

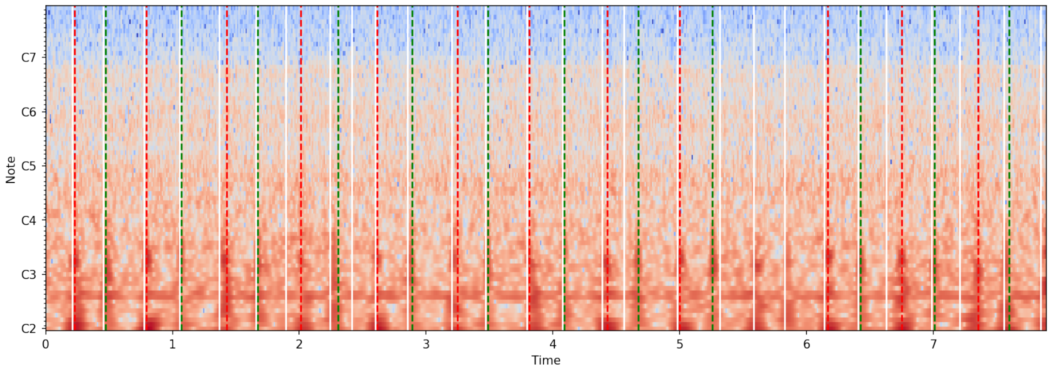


Figure 5.7: Chroma Distribution. In this example the training labelled points for S1 (red) and S2 (green) are displayed. The white lines indicate the onsets detected in this implementation.

5.1.7 Support Vector Machines

SVMs are supervised machine learning methodology which attempts to achieve a hyperplane which separates data into classes e.g. S1/S2. The optimal hyperplane has a maximised margin between support vectors (points which are closest to the hyperplane) (53).

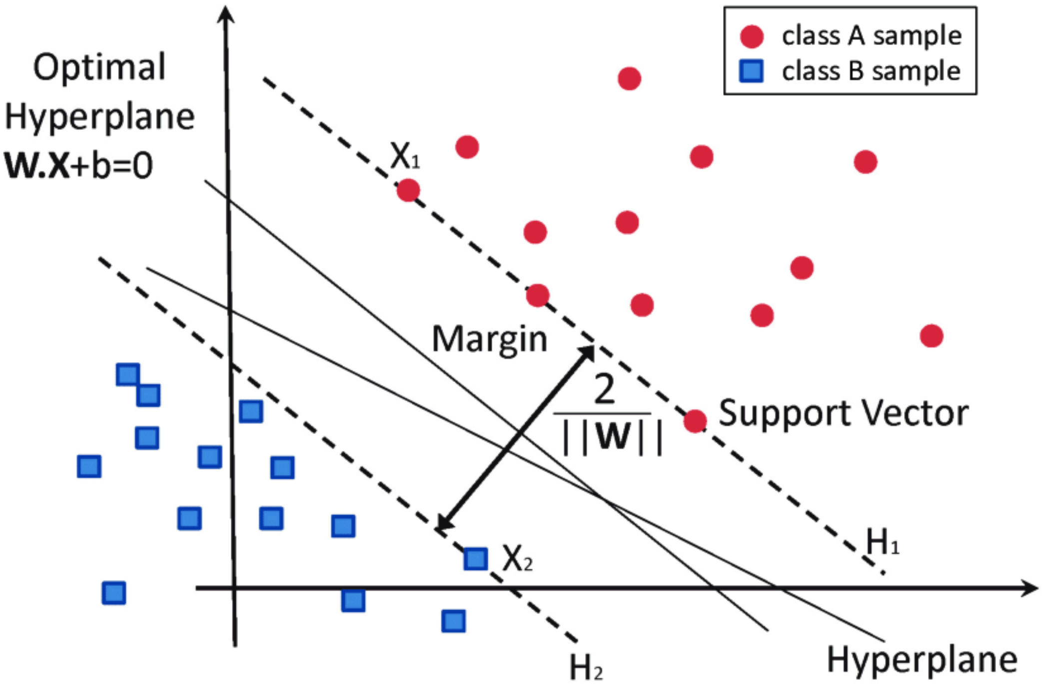


Figure 5.8: Support Vector Illustration (10)

To improve performance the kernel trick can be used on a SVM, which can augment the features to make data more linearly separable (53). For S1/S2 Classification the kernel trick was not necessary within the summarised Chroma based SVM. This was proven, using gridsearch (trialing a series of hyperparamaters) across a test and training data split in order to find the most performant hyperparameters for the S1/S2 SVM.

5.2 Heart Murmur Classification

Given the key finding during investigations that murmurs are considered generally harmonic, given murmurs are described as, "whooshing, roaring, rumbling, or turbulent fluid" (14). For heart murmur classification, this system utilised Tonnetz features to describe the tonal changes. Below you can see a comparison between normal and murmur heart sounds, which highlights the harmonic nature of most murmurs.

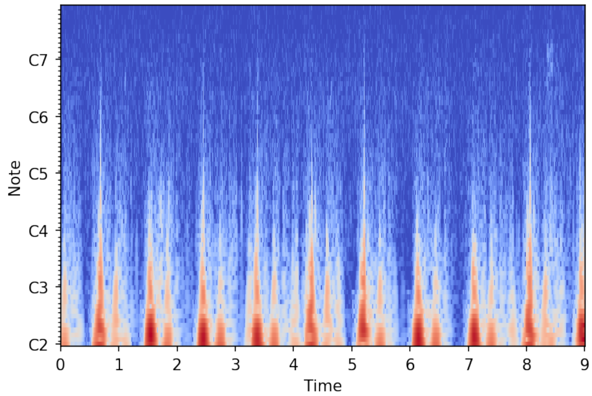
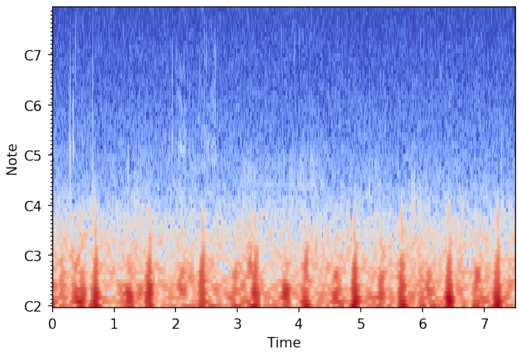
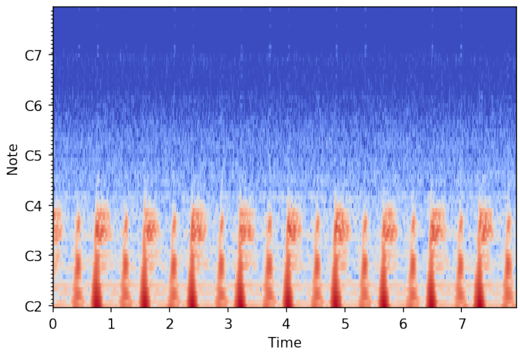


Figure 5.9: Normal Heart Sound



(a) Murmur Heart Sound



(b) Murmur Heart Sound

5.2.1 Pre-processing

Thus for murmur classification, a butterworth bandpass filter of up to 600 Hz is applied to reduce background noise whilst maintaining higher frequency murmur sounds (35), then harmonic percussive source separation is applied and the harmonic component of the audio remains for feature extraction.

5.2.2 Tonnetz

With the harmonic component, Tonnetz are utilised. Tonnetz are tonal centroid features of harmonic changes within audio. Tonnetz produces 6-dimensional basis features, two-dimensional coordinates (x-axis, y-axis) for each; perfect fifth, minor third, and major third (11).

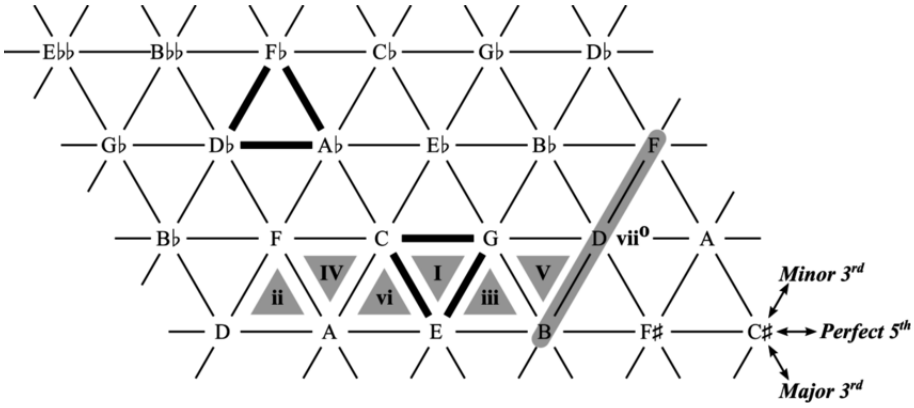
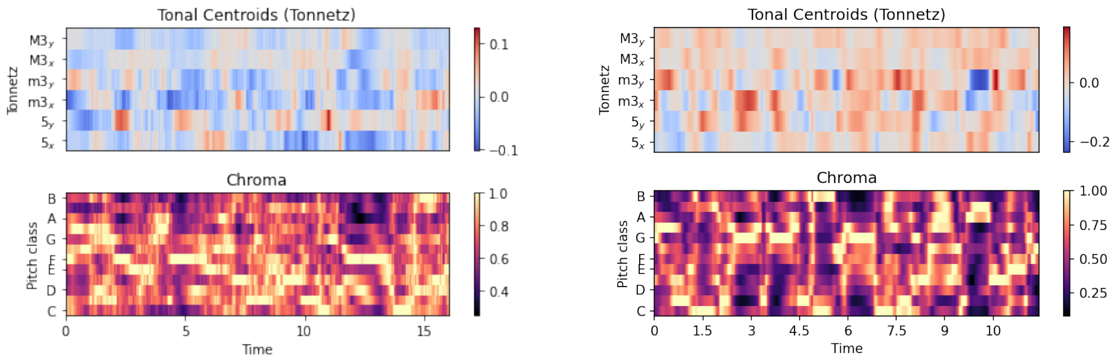


Figure 5.10: Tonnetz Diagram (11)

By using Tonnetz alone you can describe any harmonic change within audio. Summarised tonnetz of the entire harmonic component of the audio is extracted from the audio track. This means an entire audio clip of heart sounds can be reduced to 6 parameters, describing the average tonal centroid changes across a track. The effectiveness can be see below:



(a) Normal (b) Murmur

Figure 5.11: Tonal Centroids and Chroma

These images highlights the effectiveness of Tonnetz as a feature to classify against heart murmurs. The feature is highly justifiable given murmurs are the turbulent flow of blood.

5.2.3 Support Vector Machines

The summarised tonnetz features across the entire track are then inputted to a support vector machine for classification. Just as in S1/S2 Classification, Gridsearch is utilised to find the optimal hyperparameters. The excellent usage of well reasoned summarised tonnetz features produced a much more precise model especially for higher sampling rate data sources. The summarising of tonnetz also meant this could work irrespective of the sizing of the audio.

The separability of the summarised tonnetz features during training can be seen below:

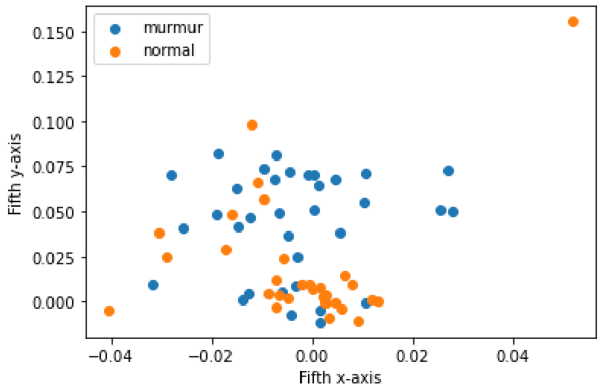
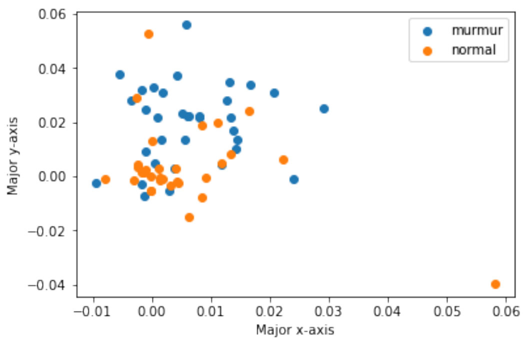
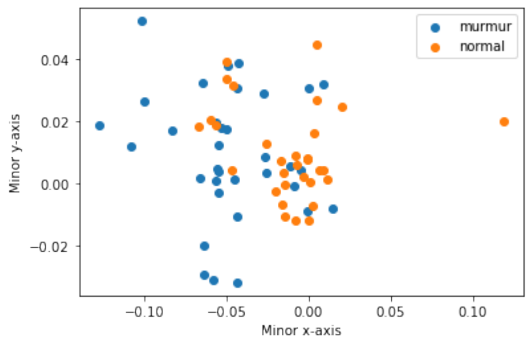


Figure 5.12: Mobile Dataset - Perfect Fifth Tonnetz



(a) Mobile Dataset - Major Tonnetz



(b) Mobile Dataset - Minor Tonnetz

Figure 5.13: Tonnetz features utilised in murmur classification SVM

5.3 Extrasystole Classification

Extrasystole is irregularly occurring extra heart sounds. To classify extrasystole, the project utilised the findings of Heart Rate Variability.

5.3.1 Heart Rate Variability: RMSSD

Root Mean Square of Successive Differences (RMSSD) between normal heartbeats, is normally an ECG metric using the distances between R intervals, however given this is the audio domain, this system utilizes S1 instead of R for a corresponding metric.

The RMSSD was calculated on every second S1/S2 detection, this reduce reliance on the accuracy of S1/S2 classification. The test sample would be classified as extrasystole if the RMSSD value was above an established threshold.

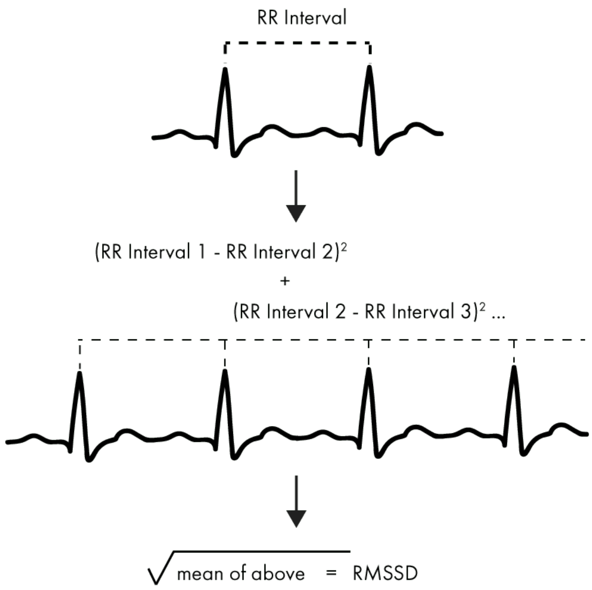


Figure 5.14: RMSSD of normal heartbeats explained (12)

5.4 Artifact Classification

Artifacts are any non-heart sounds, due to this artifacts were tested against two criteria:

5.4.1 Reasonable Minimum Lub Dubs

Firstly, S1/S2 Segmentation was completed on the audio track and there was a threshold number of minimum detected heartbeats based on 70% of the minimum expected heartbeats per minute of an adult.

5.4.2 Autocorrelation and Durbin Watson Statistic

Autocorrelation is a correlation coefficient of a signal with a copy of itself (54). Autocorrelation is a useful tool to detect non-randomness in data, given that even irregular sounding hearts have a periodic nature it would be expected that there is a strong autocorrelation for all heart sounds.

$$DW = \frac{\sum_{t=2}^{T}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{T}\hat{e}_t^2}$$

Figure 5.15: Durbin Watson Test Statistic Formula (13)

For this second criteria of artifact classification, the Durbin-Watson Statistic is evaluated. Durbin-Watson tests the autocorrelation of residuals (13). A Durbin-Watson value of 2 is considered to have no autocorrelation, 0 to <2 is a positive autocorrelation and >2 to 4 is negative autocorrelation. Sounds less strongly autocorrelating by Durbin-Watson was deemed to be an artifact.

5.5 Extra Heart Sounds Classification

To detect extra heart sounds (regularly occurring extra heart sounds), S1/S2 segmentation is utilised and the number of detected high confidence double S1/S2s are measured. Once a threshold is reached, the sound is classified as a Extra Heart Sound. Given S1/S2 Classification is extremely challenging, a threshold below what would be expected, in an ideal world of approx. 2:1 of S1s:S2s/S2s:S1s is set.

Extra Heart Sound Classification highlights the benefit of explainable techniques over 'black box' approaches, with improvements in S1/S2 segmentation, the Extra Heart Sound Classification would base classifications against the precise criteria using in a clinical setting.

5.6 Normal Heart Sounds Classification

The system to classify normal heartbeats is based on the inverse of other classifications. For example, if a sound isn't a 'Murmur', 'Extra Heart Sound' or 'Artifact', the heart sound is assumed to be normal within the test dataset. This solution would need large expansion of categories to be suitable for real-world classification in which the number of classifications is much greater.

6 Testing

In order the evaluate the success of the project implementation, the project utilized the PASCAL: Heart Sounds Challenge (14) testing evaluation methods.

6.1 Testing Criteria

6.1.1 Challenge 1: S1/S2 Segmentation

In Challenge 1, the system must output locations (in samples) of where an S1/S2 has occurred, in normal heart audio. To evaluate the effectiveness of the S1/S2 ('lub'/'dub') segmentation, total error is calculated. Total error is considered to be the average distance from the labelled S1/S2 locations and the calculated S1/S2 locations for each test audio, summated for all test audio. Distance and locations are measured in audio samples; 44100 Hz for Dataset A (Mobile) and 4000 Hz for Dataset B (DigiScope).

This total error is calculated according to the following formulas:

$$\delta = \sum_{k=1}^{j} \delta_k \tag{1}$$

Where:
δ is the total error across all test audio
j is all test audio

$$\delta_k = \frac{\sum_{i=1}^{\left\{\frac{N_k}{2}\right\}} \left(|RS1_i - TS1_i| + |RS2_i - TS2_i|\right)}{N_k} \tag{2}$$

Where:
δ_k is the average distance from labelled S1/S2 locations of the k-th sound clip
$RS1_i/RS2_i$ is the labelled location (in samples) of S1/S2
$TS1_i/TS2_i$ is the calculated location (in samples) of S1/S2
N_k is the total number of S1 and S2 in the k-th test audio

6.1.2 Challenge 2: Heart Sound Classification

In Challenge 2, Heart Sound Classification, different classifications must occur depending on the dataset. In the Dataset A (Mobile Audio), must be classified into either: 'Normal', 'Murmur', 'Extra Sound' or 'Artifact'. In Dataset B (DigiScope Audio) must be classified into either: 'Normal', 'Murmur' or 'Extrastoyle'.

The following metrics are utilized to evaluate the effectiveness of the classification:

6.1.2.1 Precision

Precision of all classification types for each dataset is evaluated.

$$Precision = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \tag{3}$$

6.1.2.2 Youden's Index

Youden's Index evaluates the system's failure avoidance. In Dataset A, the Youden Index of Artifact Classification is evaluated. In Dataset B, the Youden Index of problematic heartbeats (Extrasystole and Murmur) is evaluated.

$$Precision = \text{Sensitivity} - (1 - \text{Specificity}) \tag{4}$$

$$Sensitivity = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{5}$$

$$Specificity = \frac{\text{True Negatives}}{\text{True Negatives + False Negatives}} \tag{6}$$

6.1.2.3 F-score

F-score can be utilized to evaluate balance between precision and specificity depending on the value of β, in this challenge it is detailed (14) that $\beta = 0.9$, favouring precision. F-score is only evaluated in Dataset A for problematic heart sounds (Murmur and Extra Heart Sounds).

$$F = \frac{(\beta^2 + 1) * Precision * Sensitivity}{\beta^2 * Precision + Sensitivity} \tag{7}$$

6.1.2.4 Discriminant Power

Discriminant power can evaluate the system's ability to distinguish between positive and negative examples. Discriminant Power as stated in the task overview of the challenge (14) is only evaluated in Dataset B for problematic heart sounds (Murmur and Extrasystole). A discriminant power <1 is considered a poor algorithm, a value 1< but >2 is considered limited algorithm and a value 2< but <3 is considered a good performing algorithm.

$$DP = \frac{\sqrt{3}}{\pi}(\log X + \log Y) \tag{8}$$

where $X = \frac{\text{Sensitivity}}{1-\textit{Sensitivity}}$ and $Y = \frac{\text{Specificity}}{1-\textit{Specificity}}$

6.2 Results

Evaluation between past top solutions and this project's solutions will occur in the subsequent chapter, Evaluation.

6.2.1 Challenge 1: S1/S2 Segmentation

Dataset	Total Error (Samples)
Dataset A (iStethoscope App)	959,236
Dataset B (DigiScope)	89,544

Table 6.1: Challenge 1: S1/S2 Segmentation Results

6.2.2 Challenge 2: Heart Sound Classification

Dataset A Metrics	Result
Precision of Normal	0.47
Precision of Murmur	0.63
Precision of Extra Sound	0.17
Precision of Artifact	0.85
Artifact Sensitivity	0.69
Artifact Specificity	0.53
Youden Index of Artifact	0.22
F-Score of Heart Problem	0.22
Total Precision	2.11

Dataset B Metrics	Result
Precision of Normal	0.78
Precision of Murmur	0.10
Precision of Extrasystole	0.06
Heart Problem Sensitivity	0.37
Heart Problem Specificity	0.19
Youden Index Heart Problem	-0.32
Discriminant Power	-0.42
Total Precision	0.94

Table 6.2: Challenge 2: Heart Sound Classification Results for Dataset A (Mobile) and Dataset B (DigiScope)

6.3 Past Winning Results

		ISEP/IPP Portugal J48 / MLP	CS UCL	SLAC Stanford
Challenge 1 A	Total error	4 219 736.5	3 394 378.8	**1 243 640.7**
Challenge 1 B	Total error	**72 242.8**	75 569.8	76 444.4
Challenge 2 A	Precision of Normal	0.25 / 0.35	**0.46**	
	Precision of Murmur	0.47 / **0.67**	0.31	
	Precision of ExtraS	**0.27** / 0.18	0.11	
	Precision of Artifact	0.71 / **0.92**	0.58	
	Artifact Sensitivity	0.63 / **0.69**	0.44	
	Artifact Specificity	0.39 / **0.44**	**0.44**	
	Youden Idx Artifact	0.01 / **0.13**	-0.09	
	F-score	**0.20** / **0.20**	0.14	
	Total Precision	1.71 / **2.12**	1.47	
Challenge 2 B	Precision of Normal	0.72 / 0.70	**0.77**	
	Precision of Murmur	0.32 / 0.30	**0.37**	
	Precision of ExtraS	0.33 / **0.67**	0.17	
	Heart prb Sensitivity	0.22 / 0.19	**0.51**	
	Heart prb Specificity	0.82 / **0.84**	0.59	
	Youden Idx Hrt prb	**0.04** / 0.02	0.01	
	Discriminant Power	0.05 / 0.04	**0.09**	
	Total Precision	1.37 / **1.67**	1.31	

Figure 6.1: PASCAL: Classifying Heart Sounds Results (14)

6.4 Critical Analysis of Testing

6.4.1 Locked Excel Sheet Evaluation

The evaluation method of the challenge should be updated to prevent individuals optimizing their solution for the test data answers, either by; accessing the answers directly or repeatedly evaluating their answers.

A locked Excel sheet is used for automatic evaluation. A locked Excel sheet is not secure, if an individual is familiar with machine learning processes, they could likely overcome a lock on an Excel sheet. For the sake of academic integrity, I purposefully avoided accessing these answers during development however did verify the Excel sheet is capable of unlocking.

An online evaluation solution could be optimal, which would allow restricted testing against the test dataset, without providing access to the test data answers.

6.4.2 Subjective S1/S2 Locations

In Challenge 1, the S1/S2 calculated locations are evaluated against the labelled locations within the Excel evaluation sheet, however these labelled locations are slightly subjective to the individual who labelled them.

A system that performs perfectly in this task, may not necessarily find S1/S2 in the same location consistently but be a system that agrees with the subjective labelling consistently.

6.4.3 Total Error S1/S2 Punishments

Challenge 1: S1/S2 ('Lub'/'Dub') Segmentation is evaluated via a 'Total Error' calculation, this metric does display downfalls.

If an S1/S2 is not detected at the end of longer length audio, there is a larger penalty averaged across the total error for that test case. If an S1/S2 is not detected at the start of audio there is a smaller penalty averaged across the error for that test case, however the subsequent calculated locations could be sequentially misaligned for evaluation.

6.4.4 Sequential S1/S2 Location Evaluation

In the Evaluation Sheet for Challenge 1: S1/S2 ('Lub'/'Dub') Segmentation, the output must show the list of S1 and S2 locations (in samples) in aligned spreadsheet columns. If an S1/S2 is not detected at the start of S1/S2s in the test case, all S1/S2 detections could be evaluated against wrong S1/S2 locations.

A solution could be to require systems to output individual S1/S2 detection objects with a location value, with the evaluation method comparing is an individual detection object's location within a reasonable window range of a labelled S1/S2 object's location. This would allow for easier detection of false positives and true negatives, which are currently obscured behind the 'Total Error' evaluation metric.

6.4.5 S1/S2 Dataset Limitations

In Challenge 1: S1/S2 ('Lub'/'Dub') Segmentation, all test audio is stated to be normal audio. Thus this means the classification of S1/S2 isn't correctly evaluated, a solution could assume, given it is a normal heart audio, that an S1 is followed by an S2. A system which can detect the presence of either S1/S2, is thus capable of performing well at this challenge without requiring accurate classification.

Non-normal heart audio still has an S1/S2. To fully assess S1/S2 segmentation for all hearts, it could be optimal to evaluate against more variety of S1/S2 labelled heart audio, beyond normal heart audio e.g. Hearts with a Murmur or Arrhythmia

6.4.6 Sequential Classifications

In Challenge 2: Heart Sound Classification, each test audio is required to be classified into one of multiple categories, depending on dataset being tested. This means systems, which classify the heart audio into an individual category e.g. 'Murmur' present/not present, are impacted by the sequential order in which they evaluate each of the classifications per audio test.

Classifications during Challenge 2 are penalized if there is more than one classification made. Yet, heart audio could feature multiple classifications e.g. Murmur and Extra Sound. This means systems performing well on this challenge's evaluation, would still miss any other applicable classifications.

A solution to this testing issue, could be to expand the testing to an individual test for each category, this would reduce the classification accuracy of one part of the system impacting the subsequent classification accuracies.

6.4.7 Source Selection

Both Challenge 1 and Challenge 2 feature two separate datasets for evaluation, however these datasets display vastly difference properties, given their sources (Mobile/Digiscope) and environments (General Public/Clinical).

The intended goal of the two datasets sources is to develop general-purpose solutions suitable for both environments however focusing on a single source within a challenge e.g. Mobile Audio from the General Public, could allow for greater focus on algorithm improvement for that large specified source than handling algorithm compatibility for both sources.

The argument for general-purpose solutions also seemed somewhat dismissed given the classification categories vary between the two datasets. Dataset A requires 'Normal', 'Murmur', 'Artifact' and 'Extra Sound' classification, whilst Dataset B requires 'Normal', 'Murmur' and 'Extrastoyle' classification. If the goal is to develop general-purpose algorithms, ideally all categories of classification should be present in each dataset being evaluated.

7 Evaluation

7.1 Results Comparison

7.1.1 Challenge 1: S1/S2 Segmentation

Dataset A (Mobile): The implemented S1/S2 segmentation system achieves a Total Error of 959,235.52, which is $\approx$ 23% less Total Error than the previous best solution from SLAC Stanford (1,243,640.7).

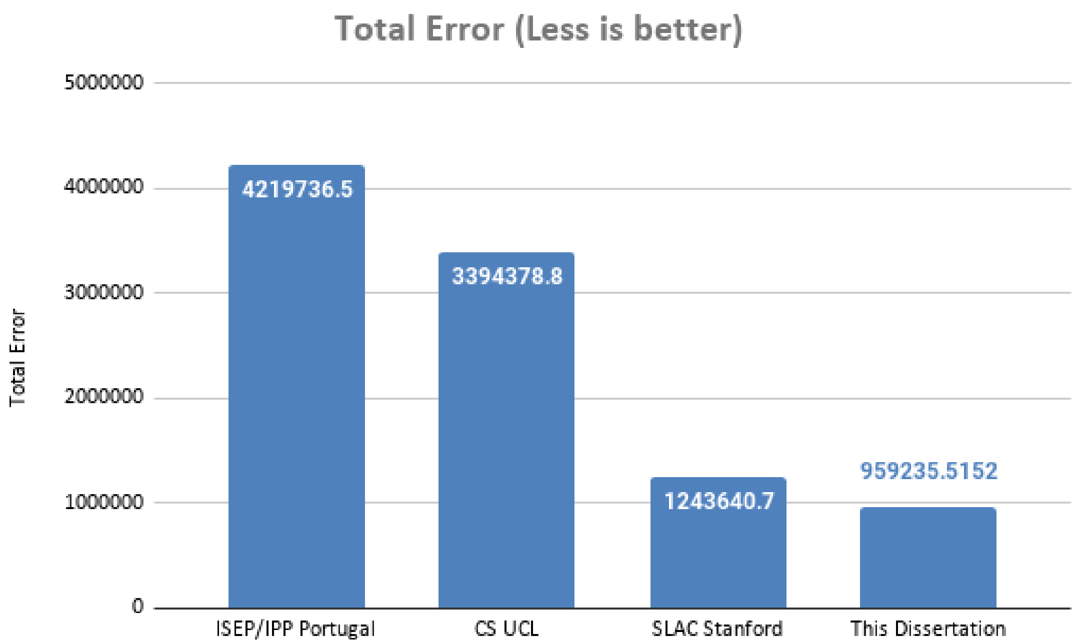


Figure 7.1: Dataset A: Comparison of Total Error Results

Dataset B (DigiScope): The implemented S1/S2 segmentation system achieves a total error of 89,543.83. This result is less satisfactory compared to prior solutions however the implementation of the system was intended for for high sampling rate and high noise mobile device audio, not low sampling rate and high fidelity DigiScope (4000 Hz) audio.

7.1.2 Challenge 2: Heart Sound Classification

Dataset A (Mobile): The system achieved a Total Precision of 2.11 which almost matched the previous winner (2.12), whilst utilizing lower resource models based on justified feature extractions. Ignoring artifacts (non-heart) audio, the system produced in this dissertation had the highest combined Total Precision in the three remaining categories ('Normal', 'Murmur' and 'Extra Sound').

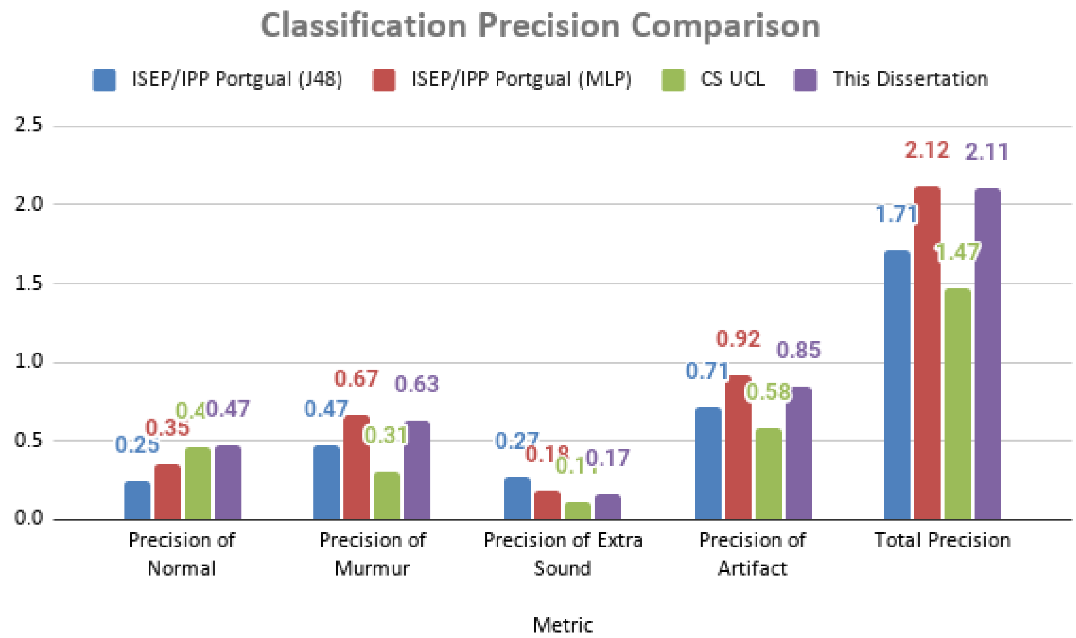


Figure 7.2: Dataset A: Classification Precision Comparison

The system crucially achieved a 10 % improvement in F-Score for heart problem sounds (Murmurs and Extra Heart Sounds) in comparison to prior best solutions.

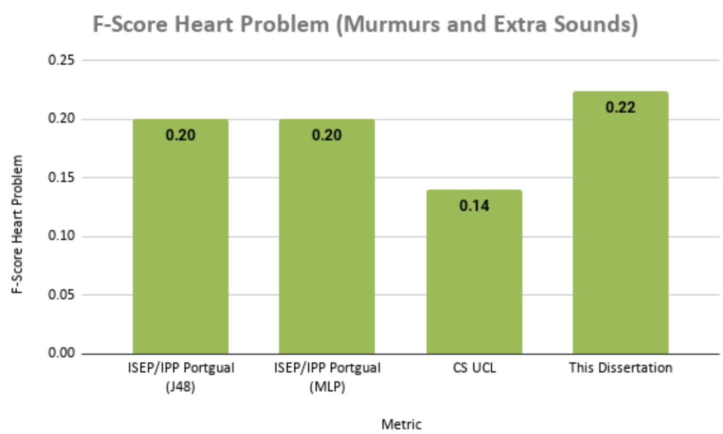


Figure 7.3: Dataset A: F-Score Heart Problem Comparison

Dataset B (DigiScope): In Dataset B, the system displayed the highest precision value for normal heart sounds (0.78) however elsewhere the system's performance on DigiScope data wasn't as performant. This coincides with Challenge 1, where the solution implemented still favours high sampling rate with high noise data, which was the focus of this project.

7.2 Critical Analysis of Implementation

7.2.1 Challenge 1: S1/S2 Segmentation

7.2.1.1 S1/S2 Detection

The system utilised in S1/S2 detection was successful having achieved $\approx$ 23% less Total Error, than the past winner of the competition SLAC Stanford in Dataset A (Mobile).

This further justifies the use Harmonic Percussive Source Separation in combination with Onset Detection, based on the the Harmonic and Percussive relationship of heartbeats previously noted in investigation key findings.

However critically could argue the techniques are not general-purpose and should be reserved for high sampling rate despite high noise environments, given the performance of of the solution was less performant from Dataset B (DigiScope).

7.2.1.2 S1/S2 Classification

S1/S2 Classification was not able to be appropriately tested by this challenge. Only normal heart beat audio was stated to be provided in the test dataset. This meant, assuming correct S1/S2 detection was completed, the subsequent detections only needed to be classified correctly for the first detection and the subsequent detections could be inferred as the correct classification.

Thus due to only normal heart sounds being tested, to improve S1/S2 segmentation results, this system completed S1/S2 detection and then checked was the confidence of the first onset high enough to justify reordering the start of the detection to be S2. This solution was designed given the time between S1 and S2 is shorter than the time between S2 and S1, thus unless there was strong confidence of S2, any heart sounds audio is more likely to have S1 detected first.

The complexity in S1/S2 Classification can be noted by past winners in the challenge, who deemed their classification "unsatisfactory" (39). This further supports the complexity of the classification, perhaps with a larger quantity of training data and further labeling including the start and end segmenting of S1, performance could be improved.

7.2.2 Challenge 2: Heart Sound Classification

7.2.2.1 Note on all Classifications

Classification performance could not be independently tested, per classification, in this challenge with a sequential classification approach. Given incorrect performance of prior classification could impact the performance of subsequent classifications, depending on the sequential order the classifications are completed.

7.2.2.2 Murmur Classification

Within Dataset A, Murmur classification performed with a total precision of 0.63, mildly below the prior winner at 0.67. It is worth noting the F-Score (Murmurs and Extra Sounds) was the most performant of all solutions. The murmur classification models also had an extremely low resource requirement, by design, for storage and computational complexity with a Support Vector Machine size of 4.1 KB, only requiring 6 parameters.

Within Dataset B, the solution was less performant, this yet again suggests that the solution favours high sampling rate despite high noise environments. The ability to extract valid tonnetz summarised features could also be impacted by the lower sampling rate.

Overall given the performance in Dataset A, the implementation of Summarised Tonnetz, to describe the tonal changes across the harmonics of heart sounds is further justified for high sampling rate audio. This outcome would be expected given murmurs are "whooshing, roaring, rumbling, or turbulent fluid" noise (14), which by nature would be considered harmonic, thus suitable for tonal change feature extraction for classsification.

7.2.2.3 Artifact Classification

Artifact classification implementation was performant, achieving the second highest precision (0.85) against the winning solution (0.92).

The implementation is justified given, it should be expected to recieve a minimum number of heart beats within heart audio and given the periodic nature of heart beats (even for problematic heartbeats), we would expect there to be an extremely high autocorrelation relationship in the Durbin Watson statistic.

7.2.2.4 Extra Sound Classification

Extra sound classification implementation performed third best in precision. The proposed solution for extra sound classification required S1/S2 Segmentation to be valid for accurate extra sound classification. As noted by this dissertation and past contestants, S1/S2 classification is remarkably challenging within audio data alone, thus detecting extra sounds consistently was also challenging.

Ideally a solution for extra heart sounds would expect an extra sound consistently to occur however given the challenge of consistent accurate S1/S2 classification, this solution relied on detecting a threshold of strong confidence extra heart sounds to be classified as having an extra sound.

7.2.2.5 Extrasystole Classification

Extrasystole classification could only able to be tested against Dataset B, which was disappointing given this system is shown to be more performant in Dataset B. Extrasystole classification requires consistent and accurate S1/S2 classification, which has been proven to be complex.

Due to this, the solution opted for heart variability testing, which inferred within a normal heart sound, each second S1/S2 detection should be the same classification, thus there should be less variability between timings, measured by the root mean square of successive differences (RMSSD) between normal heartbeats.

If there was consistently accurate S1/S2 Classification, the solution to extrasystole classification would instead be to look for inconsistent extra heart sounds of S1/S2. Perhaps if extrasystole classification was possible in a dataset similar to Dataset A (Mobile), such a solution would be more possible given the system favouring high sampling rate despite high noise environments for classifications. However given the lack of consistently accurate S1/S2 classification within Dataset B, heart variability testing was more suitable.

7.2.2.6 Normal Classification

Normal heart sound classification was optimized for the testing criteria, if no other classifications occur (e.g. 'Murmur', 'Extra Sound', ' Artifact'), the test sample would be inferred as a normal heart beat.

For real-world use case, this would need to be further advanced to classify against more classifications before considering the heart sound to be normal. However given the performance of this solution had the highest precision, it is proven justified approach to the problem.

7.3 Challenges and Limitations

7.3.1 Suitability of Mobile Heart Audio for Analysis

Despite this solution performing strongly with; $\approx 23\%$ less Total Error in S1/S2 Segmentation, second highest Total Precision performance and achieving the highest F-Score for Problematic Heart Sound Classification (Murmur and Extra Heart Sounds) of 0.22, the results of this dissertation and past winners of the competition are arguably too low for mobile recorded heart audio to be suitable for a medical environment.

This concern highlights the limits from the complexity of the challenge, given the frequency response of mobile microphones and the limits of audibility of heart sounds. The recorded audibility of heart sounds, from a mobile microphone alone, could be too difficult for valid data collection and thus too difficult for valid heart analysis, beyond non-critical usage for heart beat detection.

Perhaps with access to larger, more granular datasets and further sophisticated algorithms, automated Mobile Heart Audio Analysis could be deemed more suitable for a medical environment.

7.3.2 Medical Domain Knowledge

This project was completed by a Masters Computer Science student within any prior medical domain knowledge. This meant this dissertation required accurate dataset as a source of truth and research outside of the authors domain. Lack of prior Medical Domain Knowledge, provided additional challenge for designing the systems produced and was a limitation causing the project to be more reliant on the accuracy of the dataset utilised (14).

7.3.3 Variability

As discovered during investigations, the variability of data collection from; variety in recording APIs utilised, positioning of microphone on device and device on the user, each individual heart displaying design variations, heart rate changes and blood pressure impact further adds to the complexity of consistent audio recording of heart sounds.

Such variability increased the complexity of the challenge however was approached by applying techniques which benefit from the high sampling rate of mobile recorded data to counteract the high noise in the data. The solution also responded to variability by applying explainable justified techniques, based on the core features of heart sounds, which responded to aggregated patterns within the audio, capable of adjusting to variability e.g. Feature Extraction of Summarised Tonnetz on Harmonics for Murmur Classification.

7.3.4 Classifications Granularity

Classification granularity within training and test data remained a limitation of this project. Ideally a dataset which provided even more granular classifications would enable even more robust systems to be designed.

Such further classification granularity could include; classify specific types of murmurs instead of categorizing all murmurs together, classify S1/S2 across a variety of non-normal and normal heart sounds, classify the precise location the inconsistent extra S1/S2 occurs in extrasystoles heart sounds.

7.3.5 Dataset Quantity

Although this challenge was of remarkably beneficial to this project, the size of the dataset was limiting during implementation. The current challenge (14) provides 832 audio files between Dataset A and Dataset B. Ideally even larger quantities of training and test data within a dataset would be preferred, for robust system design.

For example, at times during the Support Vector Machine implementations, there was so few classifications within the training dataset that upon training and testing data split no classifications of one classes would be present in the training/test data split for performance measuring.

Given this challenge is intended for machine learning applications, it could be especially beneficial to expand the dataset in both training and testing to prevent solutions overfitting the data.

8 Conclusion

8.1 Review and Reflection

This project highlights the appropriate use of background research justified techniques, for a problem set can produce capable and even better performing systems than more resource intensive processes such as Multi-Layer Perceptrons.

Successful low-resource S1/S2 segmentation was achieved for a mobile recorded dataset. The proposed solution performed with $\approx 23\%$ less Total Error than the previous SLAC Stanford solution winner. The success in this challenge was achieved by using the justified techniques of Harmonic Percussive Source Separation and Onset Detection to detect valid S1/S2 locations. S1/S2 classification was completed whilst only requiring extraction of 12 parameters per classification for the Support Vector Machine, of 31.7 KB in size, based on Chroma CQT features. This greatly reduced the computational and storage resource requirements for efficient S1/S2 segmentation.

Successful heart sound classification for a mobile recorded dataset was also completed, achieving the; highest F-Score for Problematic Heart Sound Classification (Murmur and Extra Heart Sounds) of 0.22, second highest total precision for artifacts (non-heart) and heart sounds and the highest total precision for solely heart classifications. Success within classification was achieved through the accuracy of the S1/S2 Classification and the appropriate usage of Summarised Tonnetz Features on the Harmonic Components of Murmurs, to produce a classifying Support Vector Machine. From 6 parameters extracted over an entire audio sample, a Support Vector Machine of 4.1 KB was capable of murmur detection. The capability was proven having achieved the highest F-Score for Problematic Heart Sounds Classification (Murmurs and Extra Heart Sounds).

The solutions proposed were performant on high sampling rate mobile data, however adjustments could be required for lower sampling rate environments. Despite this successes within this project, the complexity of this problem cannot be understated. Variability of hearts, the lack of prior medical domain knowledge and the suitability of mobile recorded heart audio ensured this project remained consistently challenging.

The viability of mobile phone recorded heart audio for cardiac analysis would need further research given the results from any proposed solutions to the PASCAL: Heart Sounds Classification arguably does not achieve sufficient results for medical usage.

Mobile recorded heart audio reaches the limitations; from both heart sound audibility, mobile microphone frequency response capability and user positioning. Despite the challenges, the viability is worth further assessment given 31% of deaths worldwide are due to cardiovascular disease [1].

It is hoped the work produced within this project can be a research step forward towards more accessible and efficient detection of cardiac irregularities, using strongly justifiable processing techniques.

8.2 Future Work

8.2.1 Port to Mobile Device Suitable Language

Converting the Python implementations of these algorithms to a suitable mobile language would be beneficial to trial the algorithms in the public domain. There is a lack of digital signal processing third party libraries available for usage in general, not to mention for mobile device languages, so this would be a porting these algorithms would require building utilised library techniques from scratch. A priority could be to develop the algorithms for usage with JavaScript for device compatibility and mainly to be accessible to lower-resource devices such as KaiOS devices (55).

8.2.2 Adapt for Lower Sample Rate

The implementations featured in this project focused on 44100 Hz sample rate audio from mobile devices. The implementations performed worse on lower sample rates of 4000 Hz from the DigiScope dataset. Adapting hyperparamaters in the implementations for particular in onset detection, could lead to much greater performance for lower sample rates however for the scope of this dissertation, usage on high sampling rate but high noise mobile devices was the target.

8.2.3 Gridsearch for Onset Detection hyperparameters

Applying Gridsearch techniques to search for onset detection hyperparameters could enable much better performance during S1/S2 ('Lub'/'Dub') Segmentation however doing so also risks the potential for overfitting for the hearts present in the training dataset.

8.2.4 Combine Prepossessing with Advanced Machine Learning

The pre-processing techniques and approaches featured in this dissertation are both strongly justifiable and provide explain ability for outputs. The combination of this strongly justified processing and feature extractions in combination with more advanced machine learning techniques from existing approaches could provide much better results at the cost of additional resource requirements.

8.2.5 Augment Labelled Training Data

Augmenting the current labelled training data through addition of randomized noise or subsampling the audio, could enable more accurate systems, especially for the trained efficient Support Vector Machines in S1/S2 and Murmur classification.

8.2.6 Further Primary Dataset Acquisition

Collecting larger quantities and more granular data would enable improvement of the work featured in this dissertation. Primary data collection in collaboration with medical practitioners, could ensure all environmental favours are controlled and potentially provide greater insight into the required techniques to produce a more robust system.

8.2.7 Explore Multiple Simultaneous Audio Recordings of a Heart

Given ECG investigations, there is compelling evidence to evaluate multiple simultaneous audio recordings of the heart at different heart auscultation areas, this could provide a more robust classification of heart issues. Multiple microphones could be less invasive than requiring attendance at a medical practice for more invasive investigation e.g. ECG.

8.2.8 Murmur Classification for Non-Harmonic Murmurs

The current implementation of murmur classification prefers harmonic changing murmurs, which is justifiable given murmurs are "whooshing, roaring, rumbling, or turbulent fluid" within the heart. However certain murmurs demonstrate less harmonic changes, such as patent ductus arteriosus murmurs meaning an additional separate solution for this murmur edge case could be necessary.

Bibliography

[1] Mariana Ruiz Villarreal. File:human healthy pumping heart en.svg - wikimedia commons, 2021. URL `https://commons.wikimedia.org/wiki/File:Human_healthy_pumping_heart_en.svg`.

[2] Artwork by user: MadHero88. File: Phonocardiograms from normal and abnormal heart sounds.png - wikimedia commons, 2021. URL `https://commons.wikimedia.org/wiki/File:Phonocardiograms_from_normal_and_abnormal_heart_sounds.png`.

[3] Dawid Gradolewski, Giovanni Magenes, Sven Johansson, and Wlodek J Kulesza. A wavelet transform-based neural network denoising algorithm for mobile phonocardiography. *Sensors (Basel, Switzerland)*, 19(4), February 2019. ISSN 1424-8220. doi: 10.3390/s19040957. URL `https://europepmc.org/articles/PMC6412858`.

[4] Hristo Bojinov, Yan Michalevsky, Gabi Nakibly, and Dan Boneh. Mobile device identification via sensor fingerprinting, 2014.

[5] 2021. URL `https://deepai.org/machine-learning-glossary-and-terms/multilayer-perceptron`.

[6] Paul Briggs. Where is the mic on iphone 8 and how to fix it, 2021. URL `https://iphone8userguide.com/where-is-the-mic-on-iphone-8-and-how-to-fix-it/`.

[7] Meghan C. Kusko and Kathryn Maselli. *Introduction to Cardiac Auscultation*, pages 3–14. Springer London, London, 2015. ISBN 978-1-4471-6738-9. doi: 10.1007/978-1-4471-6738-9_1. URL `https://doi.org/10.1007/978-1-4471-6738-9_1`.

[8] David Chambers, Christopher Huang, and Gareth Matthews. *Cardiac Cycle*, page 117–120. Cambridge University Press, 2 edition, 2019. doi: 10.1017/9781108565011.031.

[9] Lab course: Harmonic percussive source separation. URL https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/02-teaching/2016w_mpa/LabCourse_HPSS.pdf.

[10] Esperanza García-Gonzalo, Zulima Fernández-Muñiz, Paulino Jose Garcia Nieto, Antonio Sánchez, and Marta Menéndez. Hard-rock stability analysis for span design in entry-type excavations with learning classifiers. *Materials*, 9:531, 06 2016. doi: 10.3390/ma9070531.

[11] Gilberto Bernardes, Diogo Cocharro, Marcelo Caetano, Carlos Guedes, and Matthew Davies. A multi-level tonal interval space for modelling pitch relatedness and musical consonance. *Journal of New Music Research*, 45:1–14, 05 2016. doi: 10.1080/09298215.2016.1182192.

[12] Heart rate variability - how to analyze ecg data, 2021. URL https://imotions.com/blog/heart-rate-variability/.

[13] *Durbin–Watson Test*, pages 173–175. Springer New York, New York, NY, 2008. ISBN 978-0-387-32833-1. doi: 10.1007/978-0-387-32833-1_122. URL https://doi.org/10.1007/978-0-387-32833-1_122.

[14] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor. The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results. http://www.peterjbentley.com/heartchallenge/index.html.

[15] Christopher Hanifin. *Heart sounds: A cardiac auscultation primer*. Createspace Independent Publishing Platform, 2010. ISBN 9781452840055.

[16] URL https://www.merriam-webster.com/dictionary/electrocardiogram.

[17] Definition of phonocardiogram. URL https://www.merriam-webster.com/dictionary/phonocardiogram.

[18] Derry Fitzgerald. Harmonic/percussive separation using median filtering. *13th International Conference on Digital Audio Effects (DAFx-10)*, 01 2010.

[19] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000. doi: 10.1017/CBO9780511801389.

[20] Ayush Shah, Manasi Kattel, Araju Nepal, and D. Shrestha. Chroma feature extraction. 01 2019.

[21] .xlsx file extension - what is an xlsx file?, 2021. URL https://fileinfo.com/extension/xlsx.

[22] What's the real difference between .wav, .aiff, .mp3, and .m4a?, 2021. URL `https://www.izotope.com/en/learn/whats-the-difference-between-file-formats.html`.

[23] Python Software Foundation. Python.org: About, 2021. URL `https://www.python.org/about/`.

[24] Brian McFee, Vincent Lostanlen, Alexandros Metsai, Matt McVicar, Stefan Balke, Carl Thomé, Colin Raffel, Frank Zalkow, Ayoub Malek, Dana, and et al. librosa/librosa: 0.8.0. Jul 2020. doi: 10.5281/zenodo.3955228.

[25] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL `https://doi.org/10.1038/s41586-020-2649-2`.

[26] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[28] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[29] Guido Van Rossum. *The Python Library Reference, release 3.8.2*. Python Software Foundation, 2020.

[30] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain

Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter notebooks – a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press, 2016.

[31] Audacity Team. Audacity® software © 1999-2021. URL https://www.audacityteam.org/.

[32] AutoDesk®. Tinkercad® © 2021 autodesk®, 2021. URL https://www.tinkercad.com/.

[33] Ultimaker®. Ultimaker cura, 2021. URL https://ultimaker.com/software/ultimaker-cura.

[34] Jetbrains® Google®. Android studio, 2021. URL https://developer.android.com/studio.

[35] Xuan Zhang, L Durand, Lotfi Senhadji, Howard Lee, and J.-L Coatrieux. Time-frequency scaling transformation of the phonocardiogram based of the matching pursuit method. *Biomedical Engineering, IEEE Transactions on*, 45:972 – 979, 09 1998. doi: 10.1109/10.704866.

[36] Aubrey Leatham. *Auscultation of the heart and phonocardiography*. Churchill Livingstone, 1975.

[37] 2021. URL https://www.shure.com/en-US/performance-production/louder/mic-basics-frequency-response.

[38] Peter J Bentley. istethoscope pro - instructions, 2021. URL http://www.peterjbentley.com/ispinstructions.html.

[39] Elsa Gomes, Peter Bentley, Miguel Coimbra, E. Pereira, and Y. Deng. Classifying heart sounds: Approaches to the pascal challenge. *HEALTHINF 2013 - Proceedings of the International Conference on Health Informatics*, pages 337–340, 01 2013.

[40] H. Liang, S. Lukkarinen, and I. Hartimo. Heart sound segmentation algorithm based on heart sound envelogram. In *Computers in Cardiology 1997*, pages 105–108, 1997. doi: 10.1109/CIC.1997.647841.

[41] Soraia Cruz Oliveira, Elsa Ferreira Gomes, and Alípio M. Jorge. Heart sounds classification using motif based segmentation. In *Proceedings of the 18th International Database Engineering Applications Symposium*, IDEAS '14, page 370–371, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450326278. doi: 10.1145/2628194.2628197. URL https://doi.org/10.1145/2628194.2628197.

[42] Angela Chao, Shirley Ng, and Linda Wang. Listen to your heart: Feature extraction and classification methods for heart sounds, 2021. URL https://lindawangg.github.io/projects/classifying-heartbeats.pdf.

[43] Android developers: Audiorecord. URL https://developer.android.com/reference/android/media/AudioRecord.

[44] Android developers: Mediarecorder. URL https://developer.android.com/reference/android/media/MediaRecorder.

[45] Piero Colli Franzone, Luca Pavarino, and Simone Scacchi. *Mathematical Cardiac Electrophysiology*, volume 13. 01 2014. ISBN 978-3-319-04800-0. doi: 10.1007/978-3-319-04801-7.

[46] Sheldon Sheps. Blood pressure: Does it have a daily pattern?, 2021. URL https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/expert-answers/blood-pressure/faq-20058115.

[47] J. BARLOW and P. KINCAID-SMITH. The ausculatory findings in hypertension. *British heart journal*, 22(4):505–514, Sep 1960. ISSN 0007-0769. doi: 10.1136/hrt.22.4.505. URL https://pubmed.ncbi.nlm.nih.gov/13687049. 13687049[pmid].

[48] 2021. URL http://www.med.umich.edu/lrc/psb_open/html/repo/primer_heartsound/primer_heartsound.html.

[49] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo Abad Juan, Francisco J. Chorro, Francisco Castells, José Millet Roig, Ikaro Silva, Alistair E. W. Johnson, Zeeshan Syed, Samuel E. Schmidt, Chrysa D. Papadaniil, Leontios Hadjileontiadis, Hosein Naseri, Ali Moukadem, Alain Dieterlen, Christian Brandt, Hong Tang, Maryam Samieinasab, Mohammad Reza Samieinasab, Reza Sameni, Roger G. Mark, and Gari D. Clifford. An open access database for the evaluation of heart sound algorithms. *Physiological measurement*, 37(12):2181–2213, Dec 2016. ISSN 1361-6579. doi: 10.1088/0967-3334/37/12/2181. URL https://pubmed.ncbi.nlm.nih.gov/27869105. 27869105[pmid].

[50] Dale Dubin. *Rapid interpretation of EKG's*. Cover Publishing Company, 2007.

[51] Butterworth bandpass filtering in python, 2021. URL https://www.geeksforgeeks.org/digital-band-pass-butterworth-filter-in-python/.

[52] Meinard Muller. *Fundamentals of Music Processing*. Springer, 2015.

[53] Support vector machines — scikit-learn 0.24.2, 2021. URL https://scikit-learn.org/stable/modules/svm.html.

[54] Gwilym M Jenkins. *Time Series Analysis: Forecasting and Control*. Wiley, 2015.

[55] KaiOS Technologies. KaiOS homepage © kaios technologies, 2021. URL https://www.kaiostech.com/.

[56] Scipy cookbook: Butterworth bandpass code, 2021. URL https://scipy-cookbook.readthedocs.io/items/ButterworthBandpass.html.

A1 Appendix

A1.1 Code

All code for evaluation and an install script for the PASCAL: Heart Sounds Challenge datasets has been supplied in the additional .zip file upon submission with this report.

A1.2 Butterworth Bandpass Filtering

Butterworth Bandpass Filter utilised from SciPy Cookbook example (56).

```
def butter_bandpass(lowcut, highcut, fs, order=5):
        nyq = 0.5 * fs
        low = lowcut / nyq
        high = highcut / nyq
        sos = butter(order, [low, high], analog=False, btype='band',
            output='sos')
        return sos

def butter_bandpass_filter(data, lowcut, highcut, fs, order=5):
        sos = butter_bandpass(lowcut, highcut, fs, order=order)
        y = sosfilt(sos, data)
        return y
```

A1.3 Code: S1/S2 ('Lub'/'Dub') Segmentation

```
def lub_dub_detection(x, sr, source_name):
    if source_name != 'mobile' and source_name != 'digiscope':
        raise Exception("Invalid source_name inputted")
    x_filtered = butter_bandpass_filter(x, 100, 350, sr, order=10)
    X = librosa.stft(x_filtered)
    Xmag = librosa.amplitude_to_db(np.abs(X))
```

```
    H, P = librosa.decompose.hpss(X, margin=1)
    Hmag = librosa.amplitude_to_db(np.abs(H))
    Pmag = librosa.amplitude_to_db(np.abs(P))
    p = librosa.istft(P)
    onset_samples = []
    loaded_model = None
    if source_name == 'mobile':
        onset_samples = librosa.onset.onset_detect(p, sr=sr, pre_max=20,
            post_max=20, backtrack=True, units='samples')
        loaded_model = pickle.load(open('lub_dub_mobile_model', 'rb'))
    elif source_name == 'digiscope':
        onset_samples = librosa.onset.onset_detect(p, sr=sr, pre_max=5,
            post_max=10, backtrack=True, units='samples')
        loaded_model = pickle.load(open('lub_dub_digiscope_model', 'rb'))
    features = onsets_feature_aggregator(x_filtered, sr, onset_samples)
    probs = loaded_model.predict_proba(features)
    return onset_samples, probs
```

A1.4 Code: Murmur Classification

```
def classify_murmur(x, sr, source_name):
    if source_name != 'mobile' and source_name != 'digiscope':
        raise Exception("Invalid source_name inputted")
    tonnetz = get_single_tonnetz(x, sr)
    test = np.array(tonnetz).reshape(1, -1)
    loaded_model = pickle.load(open('murmur_'+source_name+'_model', 'rb'))
    return loaded_model.predict(test)[0]

def get_single_tonnetz(x, sr):
    tonnetz = harmonic_tonnetz(x, sr)
    summarized_tonnetz = []
    for i in range(len(tonnetz)):
        summarized_tonnetz.append(np.mean(tonnetz[i]))
    return summarized_tonnetz

def harmonic_tonnetz(x, sr):
    x_filtered = butter_bandpass_filter(x, 1, 600, sr, order=10)
    X = librosa.stft(x_filtered)
    H, P = librosa.decompose.hpss(X, margin=1)
    h = librosa.istft(H)
```

```
    return librosa.feature.tonnetz(y=h, sr=sr)
```

A1.5 Code: Extrasystole Classification

```
def rmssd(onsets):
    summation = 0
    count = 0
    for i in range(len(onsets)):
        if i+2 < len(onsets):
            summation = summation + ((onsets[i] - onsets[i+2]) ** 2)
            count = count + 1
    if count == 0:
        return 0
    mean = summation / count
    return math.sqrt(mean)
```

A1.6 Code: Extra Heart Sounds

```
def classify_extra_sound(x, sr, source_name, confidence):
    detections, probs = lub_dub_detector(x, sr, source_name)

    if detections == []:
        return 0

    double_s1_count = 0
    double_s2_count = 0
    for index, prob in enumerate(probs):
        if index+1 >= len(probs):
            break
        if probs[index][0] > confidence and probs[index+1][0] > confidence:
            double_s1_count = double_s1_count + 1
        if probs[index][1] > confidence and probs[index+1][1] > confidence:
            double_s2_count = double_s2_count + 1

    # Given difficulty in high confidence Lub/Dub Classifications
    # More than 2 occurances of 2 repeated Lub/Dubs
    # Is considered an extra heart sound
    # Ideally this would be approx. 1:3 of lub:dubs or vice versa ratio.
    if double_s1_count >= 2:
```

```
        return 1
    if double_s2_count >= 2:
        return 1
    return 0

def onsets_feature_aggregator(x, sr, onsets, time=0.08, hop_length=64):
    chroma = librosa.feature.chroma_cqt(x, sr=sr, hop_length=hop_length)
    sample_size = librosa.time_to_samples(time, sr=sr)
    onsets_features = []
    for onset in onsets:
        start_index = onset-sample_size
        if start_index < 0:
            start_index = 0
        end_index = onset+sample_size
        if end_index > len(x):
             end_index = len(x)
        onsets_features.append(summarize_chroma(chroma, start_index, end_index,
            hop_length))
    return onsets_features

def summarize_chroma(chromagram, start_index, end_index, hop_length):
    summarized_chroma = []
    for i in range(len(chromagram)):
        selected_chroma =
            chromagram[i][int(start_index/hop_length):int(end_index/hop_length
        summarized_chroma.append(np.mean(selected_chroma)]))
    return summarized_chroma
```

A1.7 Code: Artifact Detection

```
def classify_artifact(x, sr, detection_times):
    r = librosa.autocorrelate(y=x)
    durb = durbin_watson(x)
    if (durbin_watson(x) > 0.025) or
        ((len(detection_times)/librosa.get_duration(y=x, sr=sr)) < 1.8):
        return 1
    return 0
```