Stress Detection and Alleviation via Electrodermal Activity and Generative Music

Christopher Corradine

MSc by Research

University of Salford

School of Science, Environment & Engineering 2023

Abstract

Accurate psychological stress detection systems have been created using a variety of methodologies and can provide users with real-time stress monitoring. Such systems can aid with providing early intervention and therapies for alleviation on order to chronic stress which is known to be detrimental to health. Previous research has shown music listening to be an effective form of stress alleviation and there is a wealth of knowledge regarding the associations between music parameters and induced emotional states. This work focuses on bridging the gap between these distinct areas to create a single system capable of detecting stress and using the resulting stress level to inform the generation of music for alleviation. A stress detection model has been created by training a random forest classifier on features extracted from samples of electrodermal activity measured during multiple affective states. MIDI data was then generated using a Markov model trained on a bespoke MIDI dataset, and musical parameters such as mode, velocity and tempo were modulated using the stress classification to apply the iso-principle. The resulting generative model is therefore a hybrid between stochastic and rule-based models. A proof-of-concept system has successfully been built along with footage of it functioning at the following link https://www.youtube.com/watch?v=SMfPrT2OJ-o&ab_channel=ChrisCorradine. This work emphasises the need for higher resolution stress detection methods and makes suggestions for a real-time system. The best performing stress detection model was subject dependent and achieved an accuracy of 86% and an F-Measure of 0.94.

Acknowledgements

The author would like to thank Duncan Williams and Bruno Fazenda for their guidance throughout the project, family and friends for their support, and AECOM for their sponsorship.

Declaration

This is my own work. The work of others used in its completion has been duly acknowledged. Experimental or other investigative results have not been falsified. I have read and understood the University Policy on the Conduct of Assessed Work. By submitting this assessment, I am declaring that I am fit to do so. I understand that personal mitigating circumstances request which relate to the standard I have achieved in this assessment may become null and void.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Background and Motivation of Research

Chronic stress is associated with some of the leading causes of death globally such as coronary heart disease and strokes. This has led to the development of stress detection technologies to improve diagnosis, increase awareness and apply early interventions or therapies for mitigation. Music is a relatively cheap and easily distributable form of stress alleviation making it ideal for early intervention. Generative algorithms are now capable of creating emotion targeted music and offer the ability to tailor music to each user and their specific emotional state in real-time. This project uses stress data to inform algorithmically generated music for alleviation which to the author's knowledge has not been done before.

1.2. Aims

The aims are as follows:

- Determine the most suitable metric based on previous research and context
- Create a stress detection system
- Determine the best performing features extracted for the chosen biometric
- Use conclusions from relevant research studies to create a music generation system that is informed by the stress data for alleviation
- Evaluate the performance of both the classification and generative models
- Combine both systems into a single system and make recommendations for future designs

1.3. Contribution to Knowledge

This work helps bridge the gap between the two research areas of stress detection and computer-generated music by creating a proof-of-concept system. A new electrodermal activity[1] (EDA) feature set has been created by applying forward selection to a set of most commonly and successfully used features gathered from the literature. The features in the final set are also ordered based on performance. Rules that alter musical parameters for

[1] Electrodermal activity is the umbrella term used for defining autonomic changes in the electrical properties of the skin. The most widely studied property is the skin conductance, which can be quantified by applying an electrical potential between two points of skin contact and measuring the resulting current flow between them (Braithwaite et al., 2013).

application of the iso-principle[2] have been suggested and a bespoke MIDI dataset has been created for the hybrid music generation system developed specially for stress alleviation.

1.4. Outline

Firstly, a review of the relevant literature is presented. Since the content of this thesis is broad the review is split into five distinct sections. Relevant psychological terminology is defined with a variety of theories presented where necessary. The second section discusses the universality of emotions followed by the emotion models commonly used in this context. The penultimate section discusses current understanding of the relationship between music, emotion and stress. Lastly, a review of algorithmic composition is presented with a focus on methods that can be applied in this context.

The methodology section is split into two main sections; stress detection and music generation. The former will explain the choice of biometric, dataset and classification algorithm as well as the machine learning pipeline applied, and the testing methodology used. Similarly, the latter will describe the MIDI generation algorithm choice and dataset, as well how the music generation parameters will be mapped to the stress signal. The results and discussion sections will also be split into stress detection and music generation and will present the findings from the work. Finally, the conclusions and further work will be presented.

[2] The iso principle is a music therapy concept that describes how music can be used to guide a patient's emotional state to a target through careful song choice. Pieces that reflect their current emotional state is played first to engage them, then songs that are incrementally closer to the target emotion are played (Heidersheit & Madson, 2015; Richardson et al., 2008).

2. Literature Review

Psychological stress is one of the most common work related health problems across the European Union and causes a variety of diseases and disorders (Setz et al., 2010). The estimated health and economic effects of this are severe but could be minimised through raising awareness and developing techniques for stress detection and alleviation (Kalia, 2002).

2.1. Definition of Emotion, Affect, Feeling and Stress

The terms emotion, affect and feeling are frequently used in psychology to describe similar phenomena but each have slightly different definitions (Kleinginna & Kleinginna, 1981). A variety of definitions have been reviewed in each case and the resulting descriptions used in this thesis are the most used at the time of writing. Sloboda and Juslin concluded that emotions are often defined as an automatic appraisal system used for quick decision making resulting in changes of physiology, behaviour and mental state (Sloboda & Juslin, 2001). Through the lens of evolution, emotions have developed for an individual to judge an environment quickly when not all information can be gathered and processed, but action may be required. Hence emotions can be intense, illogical and short lasting. Feelings on the other hand tend to be longer lasting and involve cognition.

Affect is commonly defined as a combination of both emotion and feeling, meaning affects last for extended periods of time involving multiple emotional states (Cambria et al., 2017). Stress can therefore be described as an affect that often manifests with intense emotions such as sadness or fear. Although this definition is suitable for the work in this project, the definition is somewhat vague leading some to argue it should be further restricted. For example, Koolhaas et al. suggest restricting the definition to only when “an environmental demand exceeds the natural regulatory capacity of an organism, in particular situations that include unpredictability and uncontrollability” (Koolhaas et al., 2011). To account for the positive and negative aspects of stress the state is sometimes separated into two forms; distress (negative) and eustress (positive). However, some studies suggest that this difference in appraisal does not create physiological changes different enough to be measured and

classified as different states (Schmidt, Reiss, Duerichen, & van Laerhoven, 2018).

2.2. The Universality of Emotions

Stress is usually associated with low valence and high arousal emotional states, but an important note is that one could feel a range of emotions as part of the stress response. A vital consideration for a stress detection system is whether these emotional states are universal. The debate between whether emotions are innate or culturally learnt is ongoing and the answer likely lies somewhere in between these two possibilities. The most frequent method used to test this has been showing images of facial expressions to groups from different cultures, particularly those who are illiterate or unfamiliar with western culture. Facial expressions are used since they're easy to record and reproduce, but also because it is thought that some emotion-associated facial expressions have evolved as a reflex to increase the chances of survival. The seminal work by Ekman et al. used images displaying one of six emotions (happy, fear, disgust-contempt, anger, surprise, and sadness) to populations from the United States, Brazil, Japan, New Guinea, and Borneo and found that there are universal emotions with associated expressions and physiology (Ekman, 1972; Ekman et al., 1968).
It is now widely accepted that there are basic emotions however there are notable critiques of the commonly cited methodologies used, as well as studies that have found contrasting results. Wierzbicka (Wierzbicka, 1986) and others have noted that language may influence the emotions experienced and question how the universal emotions can be so neatly listed in the language of the researchers. There are many cases where languages have specific words for emotions where other languages do not. This doesn't mean that the emotion is not experienced just because the language the individual speaks, doesn't have a single word to describe it. By limiting a study to the words of a single language rather than considering all emotions in all languages, a culturally specific framework is imposed on the study. The review carried out by Russel et al. (Nelson & Russell, 2013) showed that, particularly among illiterate non-western groups, basic emotions such as sadness, disgust fear and anger did not achieve the recognition threshold considered appropriate to conclude that these emotions are universal. They go on to suggest issues with testing methodology, such as posed faces and forced choice responses. Although such critiques exist, the consensus is that emotions are universal, which suggests that methods could be developed to detect emotion. This is

described further in the following sections.

2.3. Stress and Emotion Models

Several models have been developed to describe emotions quantitatively and qualitatively and they can be split into four main classes: discrete, dimensional, miscellaneous, and music-specific (Tuomas Eerola & Vuoskoski, 2013). Ekman's work discussed earlier is an example of a discrete emotional model where words are used to label specific emotions, but this does not quantify similarities between emotions, nor does it lend itself to computation.
Dimensional models describe emotions as being a point in N-dimensional space where each dimension describes an aspect common to all emotions. Russel's circumplex model of affect is one of the most frequently used dimensional emotion models in music psychology and is made up of two dimensions: valence and arousal. It was developed using four separate techniques: Ross' technique, Multi-Dimensional Scaling, Uni-dimensional scaling and Principal Component Analysis (Russel, 1980) and is a proven reliable model.

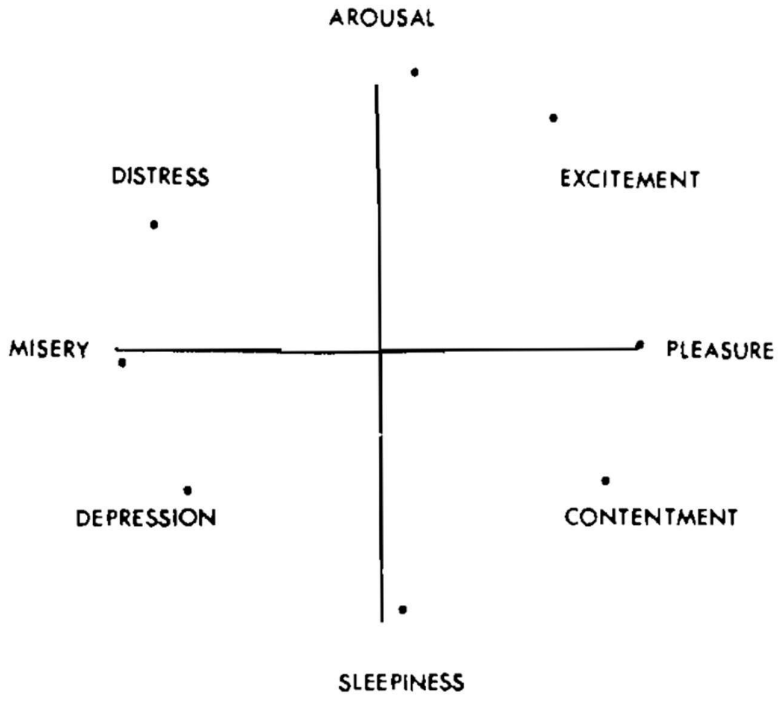

Figure 1 Russel's circumplex model of affect

Although "stress is not a basic emotion but there is a link between the dimensional models and stress" (Schmidt, Reiss, Duerichen, & van Laerhoven, 2018) this model offers a useful

framework to describe emotional states and compare what information physiological data provides. So reference will be made to valence and arousal throughout this work.
A further key point is that emotions can be confused (such as anger and fear). This has led to the use of an additional dimension known as dominance. Furthermore, none of the tests involved in the original work used music as a stimulus, though recent research has shown this model to be effective with musically induced emotions (Tuomas Eerola & Vuoskoski, 2011).
Several methods can be used to measure emotional state using this model and each can use a different component of emotion (subjective experience, biophysiological and behavioural). Self-Assessment such as Self-Assessment Manikins and the Affective Slider (Bradley & Lang, 1994) are effective but assume the user knows what emotion they are experiencing (these rely on memory requiring user action, which can cause distraction from the experience being measured).
Behavioural models such as speech detection or facial recognition have also been created (Kurniawan et al., 2013; Mishra, 2019; Neerincx & Kraaij, 2014). These have showed promising results but neither were considered suitable for the type of system being proposed. It was thought unlikely that someone would be speaking or be in front of a camera whilst using the system. In contrast, biophysiological signals can be readily recorded and analysed unobtrusively in real-time, for this reason these were considered to provide an ideal measurement method in this context.

2.4. Stress Measurement

A general review of stress measurement will be presented followed by a focus on the use of EDA. Emotion and stress measurement systems are closely related. A key difference is that stress measurement aims to classify a single condition (stress vs non-stress) whereas emotion recognition systems aim to recognise many conditions. Stress can be measured through a variety means such as the detection of behavioural changes, measuring hormone levels or analysing biophysiological signals. The main advantage of the latter is that biosignals (electroencephalography, electrocardiogram etc.) can often be measured non-intrusively and continuously.
Smith et al. provide an excellent chapter discussion of the physiological changes that accompany emotions as well as how the field has developed (Smith et al., 2004). Notable areas will be highlighted, and the reader can look to this work for more detailed information.
Ekman et al. were among the first to provide evidence that measurable physiological changes

occur in response to the basic emotions (Ekman et al., 1983). Heart rate, temperature, skin conductance and forearm muscle tension were measured whilst the participant made facial expressions associated with the basic emotions as well as reliving an emotion. They found statistically significant differences between the negative and positive emotions.

Since this work, the relationship between physiology and emotions has been explored in detail using several physiological signals and methodologies. The research area combines numerous disciplines such as psychology and computer science and is now usually called affective computing, a term coined by Rosalind Picard who is a pioneer in the area (Picard, 1999; Picard et al., 2001).

Discrete emotions have identifiable bodily changes yet the same physiological signal may be produced by multiple emotions (Smith et al., 2004). As such it is often recommended that multiple physiological signals are used to provide classification algorithms with the most amount of useful information possible (Bota et al., 2019). Such systems are known as multimodal. Commonly used physiological signals include electrocardiography (ECG), EDA, photoplethysmography (PPG), respiration (RESP) and electroencephalography (EEG).

Egmilmes et al. separate stress detection systems into two categories; event based and minute based (Egilmez et al., 2017). The former describes systems that detect features of specific events that occur in the physiological signal in response to a stressor, such as sudden spikes in skin conductance. The latter describes a method where physiological signals are windowed over a period such as one minute, and features are extracted to be used for pattern recognition. Since other pieces of research have successfully used window lengths shorted than one minute, a more appropriate term would be “window based”. Window based stress detection follows a standard machine learning pipeline which will be discussed in detail in the methodology (Shukla et al., 2019).

EDA is one of the most informative biosignals for stress detection. It has been used to understand psychological states from as early as 1906, where famed psychoanalyst Carl Jung used galvanometer needle deflection (galvanic skin response, GSR) to determine unconscious complexes in his patients (Jung, 1919). EDA encompasses various related metrics such as galvanic skin response, skin resistivity and skin conductance, the commonality being they each describe a measurement of electrical current passing through the skin.

EDA is now a vital part of many seminal stress measurement systems. Healey and Picard recorded electrocardiogram, electromyogram, skin conductance, and respiration whilst drivers navigated open roads in Boston and achieved an accuracy of 97% (Healey & Picard, 2005). Minguillon et al. created a portable system that utilised EEG, ECG, EMG and GSR

and classified between three conditions; stressed, relaxed and neutral achieving an accuracy of 86% (Minguillon et al., 2018). Ayata et al. used the GSR data from the DEAP dataset to compare four machine learning algorithms, decision tree, random forest, support vector machine and k nearest neighbours (Ayata et al., 2017). They found the Random Forest and Support Vector Matrix to be the best performing classifiers and showed that EDA contains information about valence as well as arousal. They also noted that GSR is not a stationary signal and so used the discrete wavelet transformation and empirical mode decomposition, rather than the short time Fourier transform for frequency analysis. Most systems are categorical classifiers, but some models offer better resolution such as that by Salazar-Ramirez et al. who used a fuzzy algorithm (Salazar-Ramirez et al., 2018). The validation of stress measurement models generally involves using held out data sets however more sophisticated methods have been used such as using stress hormones in the body as biomarkers (Nath et al., 2020).

2.5. Music, Emotion and Stress Alleviation

Appropriate therapies should be used to mitigate the negative effects of stress and prevent the state being prolonged. Numerous therapies have proven effective such as micro meditation, warm stones, good news, pharmacological interventions and music (Akmandor & Jha, 2017). Music is present in all known societies and accompanies many of the most important aspects of our lives such as celebrating marriages and mourning the loss of loved ones (Mehr et al., 2019). It is used for a variety of purposes and is often called the language of emotion, but studying the complex relationship between music and emotion is an ongoing challenge (Chamorro-Premuzic & Furnham, 2007). Music has been shown to reduce stress level and has been successfully applied alongside other forms of therapy (Clark et al., 2006; Hatta & Nakamura, 1991). In more extreme cases it has even been shown to help with pain regulation (Kwekkeboom, 2003; Tsai et al., 2014).

Yehuda et al. provide an excellent summary of music and stress and cite several studies showing the effects music can have on the stress response, as well as describing in detail the physical mechanisms by which music is processed (Yehuda, 2011). The study of brain biochemistry helps explain how music induces an emotional response and can be used to alleviate stress. Pleasurable music has been shown to release dopamine (a neurotransmitter released by the nucleus accumbens that is prominent in the reward system) as well as elevating serotonin and endorphin levels. This produces a relaxing effect and a feeling of

wellbeing and has been shown to prevent levels of the stress hormone cortisol from elevating further (Khalfa et al., 2003).

There are several techniques that use music for stress alleviation: music listening, guided imagery, progressive music relaxation and instrumental group improvisation (Yehuda, 2011). Though effective, instrumental group improvisation requires physical resources and multiple people making it difficult to be incorporated into a readily available system for early intervention. The system being proposed in this study just uses music listening. However, progressive music relaxation and guided imagery could also be easily implemented.

Before continuing to discuss music-emotion associations it is worth noting the downfalls of the current literature. Most research has used populations from western society creating the potential for bias and generalisations to be made, without considering cultural learning. Cultural learning can manifest in multiple ways, such as what music the listener was brought up listening to, as well as the degree of musical learning. This is important because such differences can change the associations a certain individual has (Midya et al., 2019).

Eerola and Vuoskoski conducted a review on music and emotion studies and found several patterns as well as giving multiple useful considerations for future projects (Tuomas Eerola & Vuoskoski, 2013). They found that results are generally incoherent due to the variety of methodologies used and although differing methodologies are used due to their individual benefits, a more specific framework should be created for future work so that results can be more easily compared, and repeatability ensured. There is a bias towards classical music in the literature even though other genres such as R&B and soul are more popular. Cross-cultural studies on music and emotion are sorely needed to estimate the degree to which findings in the field can be generalised.

Associations of minor keys with negative valence and major keys with positive valence are often taught in western societies from an early age. Some of the first work investigating these associations was done by Hevner who confirmed that there was an underlying truth to them, but that it was complex and not applicable in all cases (Hevner, 1935). Furthermore, other musical parameters such as register and intensity could mask these associations. This is partly why some genres (such as funk and soul) can be in a minor mode and yet still induce emotions with positive valence.

Several studies studied this further by including all modes. One example is the study by Van Der Zwaag et al. which investigated how tempo, mode and percussiveness affected emotional state and psychophysiology (van der Zwaag et al., 2011). It was found that an increase in tempo correlated with an increase in arousal, and an increase in arousal was also correlated

with minor songs more than major songs. This aligns with the findings of Ramos et al. who also investigated the relationship between affect and arousal with mode. A key distinction was that increasing the tempo was found to make major music seem happier and minor music seem more angry (Ramos et al., 2011). Temperley and Tan studied the effects of modality on emotional qualities and found the emotions evoked aligned with the number of flats and sharps in the mode relative to the Ionian mode (Temperley & Tan, 1973).

To become accustomed to the sounds of different modes musicians often use an exercise where a drone is used to set the tonic and patterns of intervals change the mode. This is similar to the suggestion by Temperley and Tan which was applied in the study carried out by Bostwick et al. A drone was used to define the tonal centre and melodies were generated using the intervals of the mode (Bostwick et al., 2018). Their results supported previous work where Ionian was rated as the happiest with Lydian and Mixolydian, the only change between the excerpts was the drone which changed the tonal centre suggesting the method can be used to modulate emotional state.

A clear progression of this work is using musical parameters such as modality to guide listeners through the affective space. In this work, the framework used for this purpose is the iso-principle which was established in music therapy to alter patients' emotional states using music. Songs are selected to reflect the patients' current emotional state and gradually songs closer to the target state are played. This is done to secure the attention of the listener by matching their emotional state (Sloboda & Juslin, 2001).

Jiang et al. found that music preference is also a considerable contributing factor whereas musical training had no effect (Jiang et al., 2016). They outlined the difference between state and trait anxiety, which may require different methods for alleviation. Furthermore, the emotions induced by music are also individual to the listener which adds complexity to systems aiming to guide emotions using music since subjectivity ought to be taken into account (Sagha et al., 2015).

2.6. Algorithmic Composition

Algorithmic composition describes the use of algorithms or rules to create music. One of the first forms of algorithmic composition was singing in canon where rules would be given for singers to follow based upon simple variations of the lead melody (Maurer, 1999). Other composers made use of natural randomness such as Mozart who used a technique called "dice music". Computers have been used for algorithmic composition since 1955 when Hiller and

Isaacson created the Illiac Suite. Iannis Xenakis then created compositions using note probability densities and a random generator, which along with the dice music technique are known as stochastic models. Once music could not only be reproduced but written using computer code the aims expanded, becoming both utility and philosophically based. Current aims of the field include passing the music adapted Turing test, creating real-time backing tracks, composing music with specific affect associations and creating tools for composers. Different end goals require a variety of bespoke models with many different methodologies applied.

Generally there are five groups of generative model; random, rule, math, music grammar and machine learning (Jin et al., 2020). Multiple approaches can be used to create hybrid models enabling the best characteristics of different algorithms to be combined. Neural networks have been used extensively in music generation with promising results (Briot & Pachet, 2020; Kalingeri & Grandhe, 2016; Mao, 2018; Wu et al., 2017; Yu & Canales, 2019). Jin et al. created a long short term memory neural network for MIDI generation and trained it using a discriminator network (Jin et al., 2020). To generate music of a specific genre, rule-based constraints were defined to encourage patterns linked to that genre, such as note interval differences being less than an octave for classical music.

Most models use MIDI as the music representation due to the versatility and ease of processing but these models rely on virtual instruments to create the timbre, requiring mixing or post processing after generation. Recently Deep Mind's Wavenet bypassed this requirement by creating raw audio at a sample by sample level, but the algorithms for creating this kind of model are extremely sophisticated and would not offer the kind of musical flexibility required in this context (Oord et al., 2016). Music is categorised into genres based on shared characteristics of songs. Music preference has been shown to be an important mediating factor for stress alleviation and so models that can generate genre specific music are particularly well suited to this challenge (Jiang et al., 2016).

Jazz and classical music are often used since the structure of such songs are less defined (Briot & Pachet, 2020). That said, these genres usually have greater harmonic complexity making them difficult to replicate. This also reflects a bias towards western music in the research as mentioned in the previous section. Although attention is rightfully turning towards deep learning for these challenges, the performance of these models is dependent on multiple factors such as computational power, availability of training data and requirements for user input. Hence although sophisticated algorithms such as LSTMs have performed well

at specific tasks it does not mean they outperform other methods in all contexts (Wiriyachaiporn et al., 2018).

Specific to this project are models that create music to induce a desired affect. This requires generating music using computers as well as making use of the most up to date understanding of how people react emotionally to music. Wallis et al. split the affective space into six areas using the modes and randomly played music generated from each mode to eleven subjects to study the strength of the relationship between modes and affect (Wallis et al., 2011). They modulated the mode, upper extensions, pitch register, voice spacing and voice leading creating a system that successfully induced the desired affect. However, it was noted that the perceived valence was affected by the intended arousal. It was shown that note density correlated most with arousal and that the relationship between music and the affective space is complex, meaning that small changes in parameters can affect the resulting valence and arousal unpredictably.

3. Methodology

3.1. High Level System Design

Most of the processing was accomplished using MATLAB due to its detailed documentation, Machine Learning toolbox, and MIDI toolbox (T. Eerola & Toiviainen, 2004). The stress detection and alleviation systems were developed separately and connected in the final stage. Figure 2 shows the flow of data through the system beginning with input of the physiological signal, processing and classification, melody generation informed by the user's stress level, and ending with audio output. Pure Data was used as a bridge to send MIDI to the standalone synthesiser plugin Analog Lab 4 via Open Sound Control (OSC). In a real-time system, the EDA signal would be a live stream of data from a measurement device on the user. However, in this implementation, the system was developed 'offline' using previously captured EDA traces on participants. A final system could be developed to work in real-time with the only constraints being the speed of processing and the classifying of the EDA signal.

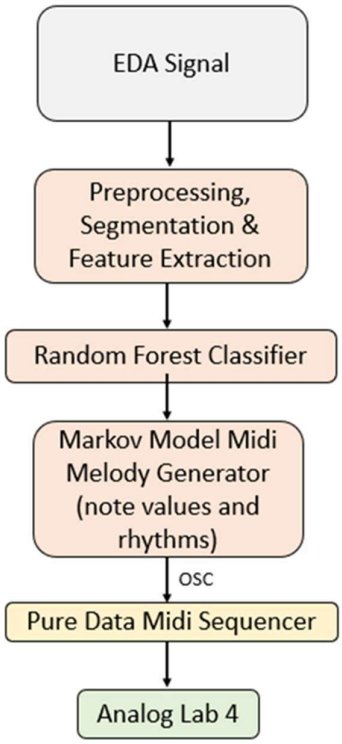

Figure 2 Flow chart of high-level system design

3.2. Stress Detection Using Electrodermal Activity

This problem lends itself to Machine Learning (ML) due to the complexity of mapping the bio-physiological signal onto an emotional state. Simple mappings can be coded using conditionals, but this becomes unmanageable as the number of potential inputs, thresholds, and associated outputs increases. Rather than explicitly programming all cases, ML relies on recognising patterns and creating inferences from example data so that accurate decisions can be made when given new data (Jordan & Mitchell, 2015). The stress detection system followed a common ML pipeline as shown in Figure *3*. EDA data can be collected either by creating a bespoke experiment or by using an already existing dataset, the latter approach was used in this case. The WeSad dataset was used since the experimental design and acquisition techniques met the requirements of the project aim, further justification can be found in Chapter 3.2.2 along with a detailed description of the dataset. Once data is collected, it is then checked for artefacts, cleaned if required, segmented so the model can detect variations in emotion over time, and standardised if required. An initial set of features thought to correlate with the target classes are then extracted, and an optimal feature set is found using feature selection. This refined feature set is then split into training and testing sets via a sampling method so the model can be tested on data that wasn't used during training (Bota et al., 2019).

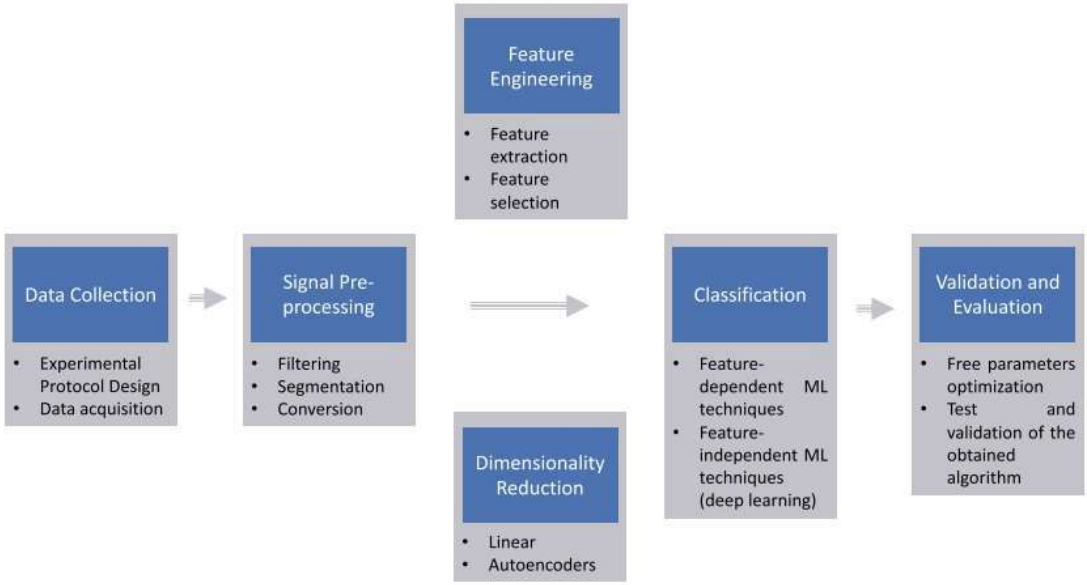

Figure 3 Schematic representation of machine learning process for emotion recognition (Bota et al., 2019)

3.2.1. Choice of Biosignal

There are two groups of biosignal: physical (such as pupil dilation, eye movements, and blinking) and physiological (such as ECG and EMG). Both groups can be used to determine emotional state without conscious control of the subject and can be recorded continuously, giving a steady stream of data for classification. Biosignals vary based on what dimensions of affect (valence or arousal) they correlate with, leading many researchers to create multimodal systems where biosignals are combined. Biosignals commonly used for stress detection are: EDA, EEG, ECG, and EMG (Giannakakis et al., 2019). Combining biometrics can increase performance however it can also increase model complexity and data preparation time, as well as requiring additional sensors. For these reasons only a single biometric was chosen, although the potential for additional metrics is discussed in Further Work.

When making system comparisons, it can be difficult to determine the source of differing system performances since different biometrics, datasets, stimuli, and classification methods are frequently used in studies. Accuracy is the most cited evaluation criterion but fails to give a full description of performance so other metrics such as F-Measure are often cited also (see section 3.2.6.2 for the more detail on F-Measure). Summaries such as that provided by Giannakakis et al. allow researchers to quickly analyse previous work in the area and determine which methods have been successful (Giannakakis et al., 2019).

It is important to consider context when deciding which biometric to use. For example, speech was considered unsuitable since it would have required the user to talk consistently throughout its use. Pupil dilation was also considered unsuitable as it would have needed the user to keep their eyes open. As the aim was to create algorithms that could be applied in real-time, there was also a processing time limit where all pre-processing, classification and music generation had to be completed within the few seconds between each window of data. To minimise obtrusiveness, it was also noted that minimal sensors should be used meaning information about valence and arousal must be provided as efficiently as possible.

EDA offers a relatively non-intrusive window into a user's emotional state requiring only a single sensor placed on the wrist, palm, or foot. It has been shown to correlate with valance and arousal and has well-studied measurement and processing techniques, with widely available sensors (Ayata et al., 2017). EDA has also consistently performed well in stress recognition tasks suggesting it would be well suited to this task (Hsieh et al., 2019; Liu & Du,

2018; Zangróniz et al., 2017).

3.2.2. WeSAD Dataset

Several datasets have been created for benchmarking emotion recognition algorithms (DEAP, AMIGOS, LUMED-2 and SWELL) and care must be taken when choosing which to use. Choice depends on the development objectives and experimental protocol. Stress data was needed but it was also important to expose the model to other conditions because in real use it would have to distinguish between every emotional state and individual experiences. That said, this is an early prototype and machine learning practitioners advise adding complexity slowly and incrementally once each phase is achieved, suggesting a dataset with several well recorded conditions would be suitable (Google, 2021).

The WESAD dataset is described as a "Multimodal Dataset for Wearable Stress and Affect Detection" (Schmidt, Reiss, Duerichen, & Van Laerhoven, 2018). It contains data from 15 participants who undertook a Trier Social Stress Test (TSST), a well-studied stress induction procedure relying on social evaluation and high mental load. Participants had 3 minutes to prepare for a 5-minute speech about their strengths and weaknesses, which was given to a panel of actors posing as human resources specialists. A mental arithmetic task was then carried out, with the participants given 5 minutes to count down from 2023 in increments of 17. Measurements of ECG, EDA, EMG, respiration rate (RESP), skin temperature (TEMP), and motion (ACC) were recorded with chest and wrist worn devices. These metrics were measured during baseline (20 minutes), stress (10 minutes), and amusement (392 seconds) conditions, producing a labelled dataset.

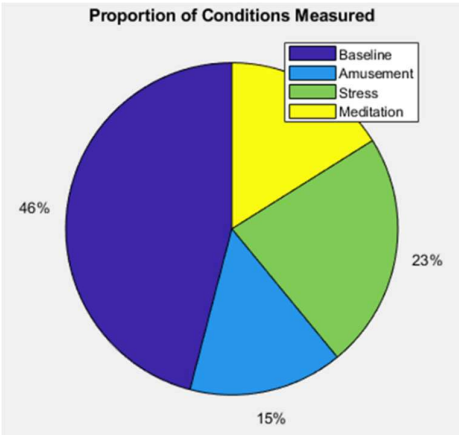

Figure 4 Proportion of condition measurements in the WeSad dataset

There were more observations of the non-stress condition than the stress condition resulting in a class mismatch (Figure 4). This can lead to high accuracies that suggest the model is performing well when the model is simply learning to classify most observations as the majority class. This gave two options: data from the majority class could be removed until conditions were equally represented; or evaluation metrics other than accuracy could be used. The latter option was chosen with F-Measure being adopted as explained in further detail in Chapter 3.2.6.

Good experimental designs control confounding variables whilst providing data that reflect real scenarios. Confounding variables can be more easily managed by conducting the experiment in a controllable environment such as a laboratory, but this puts the individual in an artificial environment which could have unintended consequences. For example, baseline recordings may not be true baselines since the individual may be stressed in the unfamiliar environment. Statistical tests were therefore carried out on the data to determine if the intended conditions had successfully been induced. Schmidt et al. conducted PANAS and DIM tests on the WeSad dataset as well as a Wilcoxon signed-rank test, and it was determined the desired states were evoked.

As mentioned, additional conditions were required to ensure the model could distinguish between stress from other states. Amusement was considered appropriate since it is a high arousal and high valence emotional state meaning successful classification would need the model to discriminate between valence as well as arousal.

3.2.3. Analysis of EDA

The wrist SC measurements from the WESAD dataset were recorded using an Empatica E4 with a sampling rate of 4Hz. Once innervated, the eccrine sweat glands release sweat into the skin, which increases the SC. These glands are most densely populated on the palms and soles, defining the points in the body where measurement devices are usually placed. SC is measured by recording the current between two electrodes placed on the skin and the standard units are micro-siemens μS. Figure *5* shows an example EDA trace in the time domain with each of the conditions highlighted. EDA signals can have a characteristic known as low frequency drift which requires a correction before using for classification models.

None of the data in this set exhibited low frequency drift.

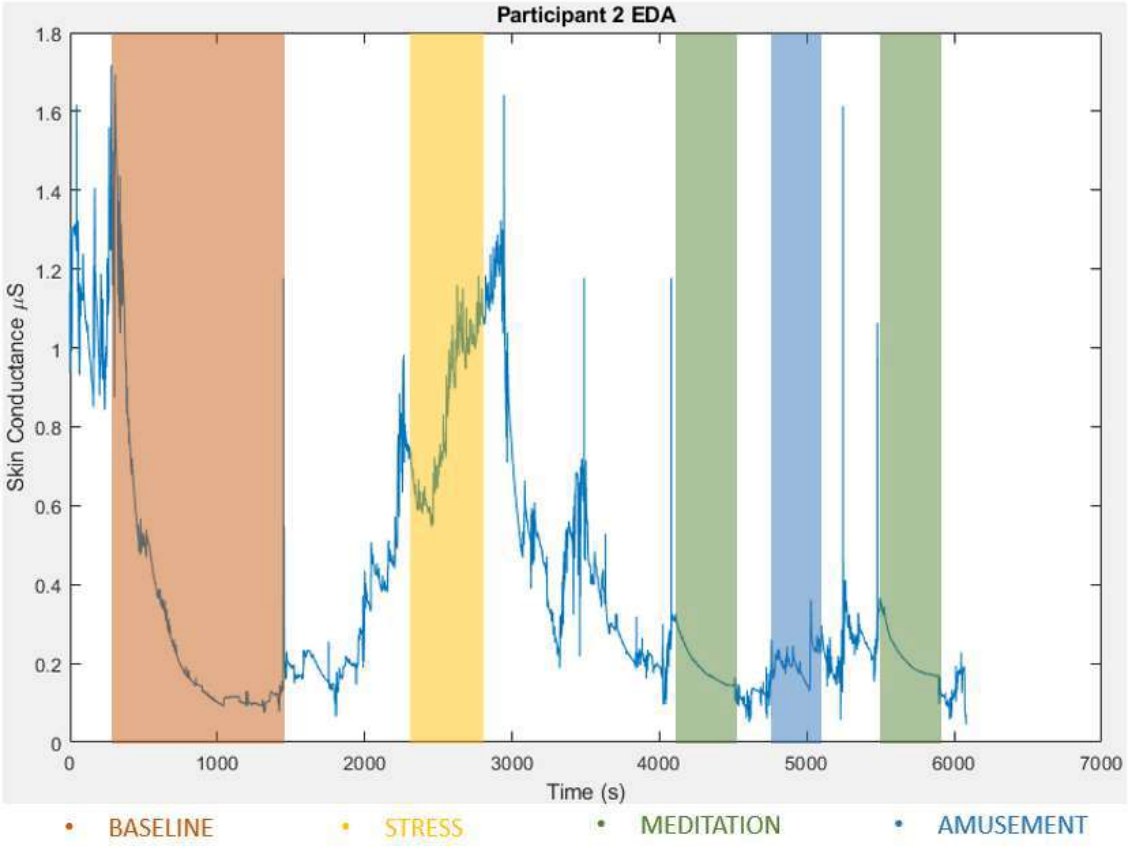

Figure 5 Example EDA trace

SC is a non-stationary signal made up of two components, tonic and phasic. The tonic component is the slow moving, underlying level (also known as the Skin Conductance Level). The phasic component is the faster, more reactive part (commonly referred to as the Skin Conductance Response). A SC trace can be converted into tonic and phasic parts using algorithms such as Continuous Deconvolution Analysis and General Linear Convolution. Two commonly used programs for this purpose are Ledalab (Benedek & Kaernbach, 2010b) and SCRalyze (Bach, 2014; Bach et al., 2009). Traces are commonly decomposed into the phasic and tonic part for analysis in psychology experiments, so these signals were considered for feature extraction. Unfortunately, these algorithms couldn't be implemented due to the processing time restrictions. Furthermore, SCRs tend to occur in response to sudden changes, rather than to gradual changes like those from music stimuli, so it was decided decomposing the signal into these components would not be best suited in this

context.

3.2.4. EDA Pre-processing

The WESAD dataset contains three conditions: stress, baseline, and amusement, so the problem was posed as stress vs non-stress classification as done by Schmidt et al (Schmidt, Reiss, Duerichen, & van Laerhoven, 2018). Firstly, the data was analysed visually by plotting against time and checks were made for common issues such as motion artefacts or loose-fitting electrodes by looking for sudden, extreme changes in skin conductance. None were detected. The data was then converted to the frequency domain via the discrete Fourier transform (Figure *6*) and checked for signs of aliasing, again none were found in this dataset. A 40Hz low pass filter is sometimes applied to EDA data as a smoothing function but this was considered unnecessary since the cut off sampling rate of the Empatica E4 was 4Hz. The dataset was labelled using the label vector provided, which required down sampling from 700Hz sampling rate to 4Hz.

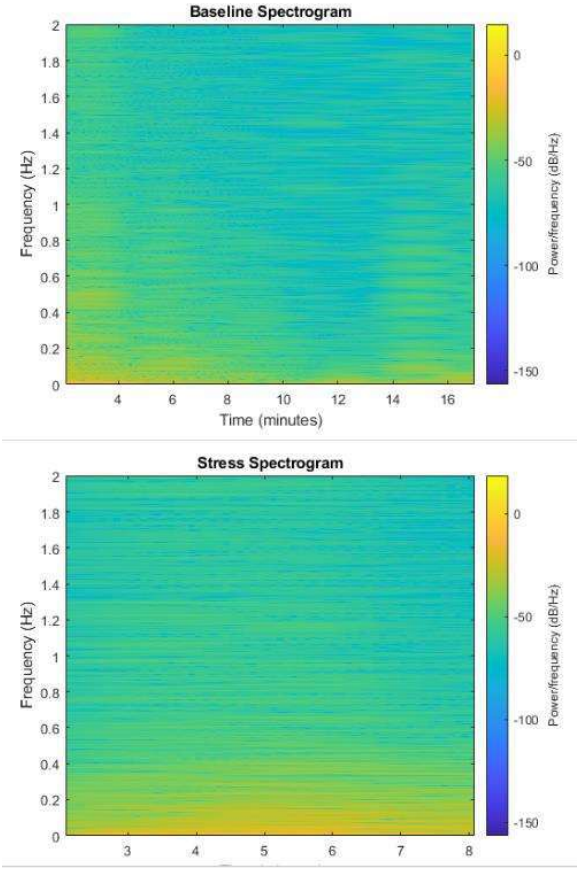

Figure 6 Spectrograms of baseline and stressed signals for Participant 1

Standardisation is the process of transforming data gathered from different sources so it can be more easily compared. EDA differs greatly between individuals leading some to train models based on standardised data. A benefit of using the Random Forest algorithm was that it is unaffected by the magnitude of features so no standardisation was applied.

3.2.5. Feature Extraction

The purpose of feature extraction in classification problems is to condense the data whilst maximising meaningful differences that allow the conditions to be separated. This can increase classification performance, model interpretability, and reduce processing time (Motoda & Liu, 2002). Many EDA features have been extracted and tested in the literature, but often using different datasets, algorithms, and performance metrics, making comparisons difficult. The initial feature set was chosen based on how often these features appeared in the literature, how well they performed and if they could be implemented with the processing time limitations imposed by the generative loop.

The windowing technique used limited the smallest window length to one minute. Ayata et al. (2017) showed accuracy decrease as segment size increases and that overlapping windows increases accuracy. They found the difference in accuracy between three second long segments and one minute long segments to be approximately 5%. A different technique would have taken too long to develop and other studies using window lengths of one minute have shown to be effective, so one minute window lengths with 5 second overlap was adopted (Egilmez et al., 2017; Kurniawan et al., 2013; Shi et al., 2010).

The distribution of the time domain EDA signal usually has a higher mean and variance for the stressed condition which makes sense intuitively but is not universal. For this reason, it is good to search through multiple features extracted from multiple domains. Table *1* shows the domains and features that were used to create the initial feature set. The justifications and analysis of each domain will be explained in the following sections.

Feature Extraction	
Time Domain	Maximum
	Minimum
	Mean
	Median
	Standard Deviation
	Skewness
	Kurtosis
	Area Under Curve
Intrinsic Mode Function	Maximum
	Minimum
	Mean
	Median
	Standard Deviation
	Skewness
	Kurtosis
First Derivative	Mean
	Standard Deviation
	Peak Frequency
	Peak Mean
	Peak RMS
MFCC	Mean
	Standard Deviation
	Median
	Skewness
	Kurtosis

Table 1 Summary of features extracted from each domain

3.2.5.1. Time Domain Features

Figure *7*, Figure *8* and Figure *9* show each participants' data in the time domain during the three conditions. The stress signal generally has a larger magnitude and variation, suggesting that measures of central tendency and spread can be helpful in detecting it. The variability between subjects is also clear, emphasising the difficulty of predicting emotional state from raw physiological signals.

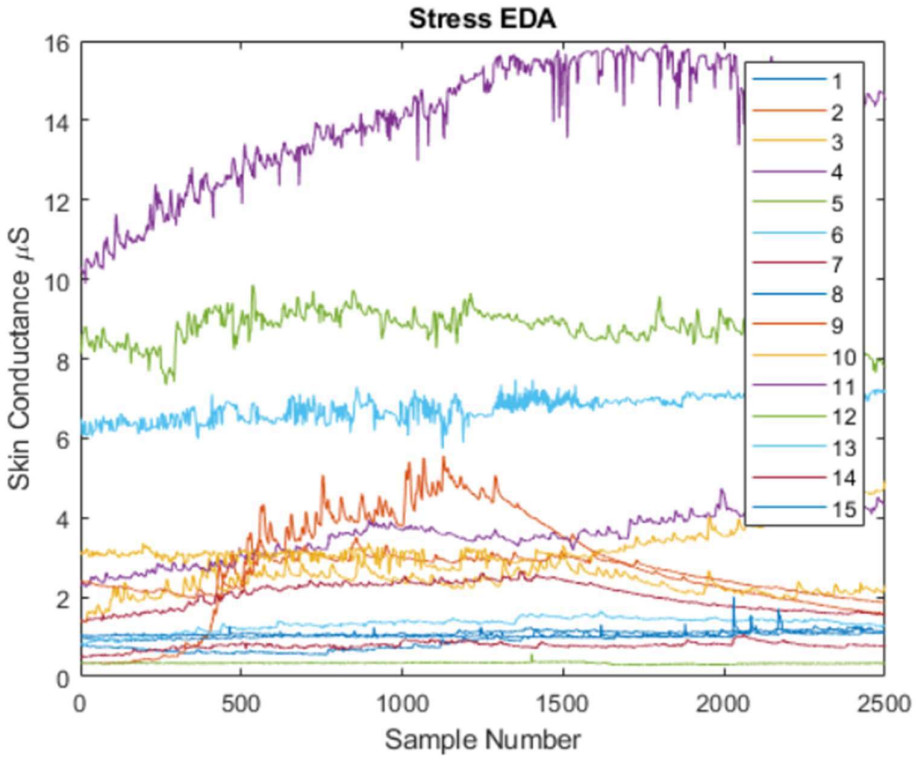

Figure 7 Example time domain EDA signal during the stress condition

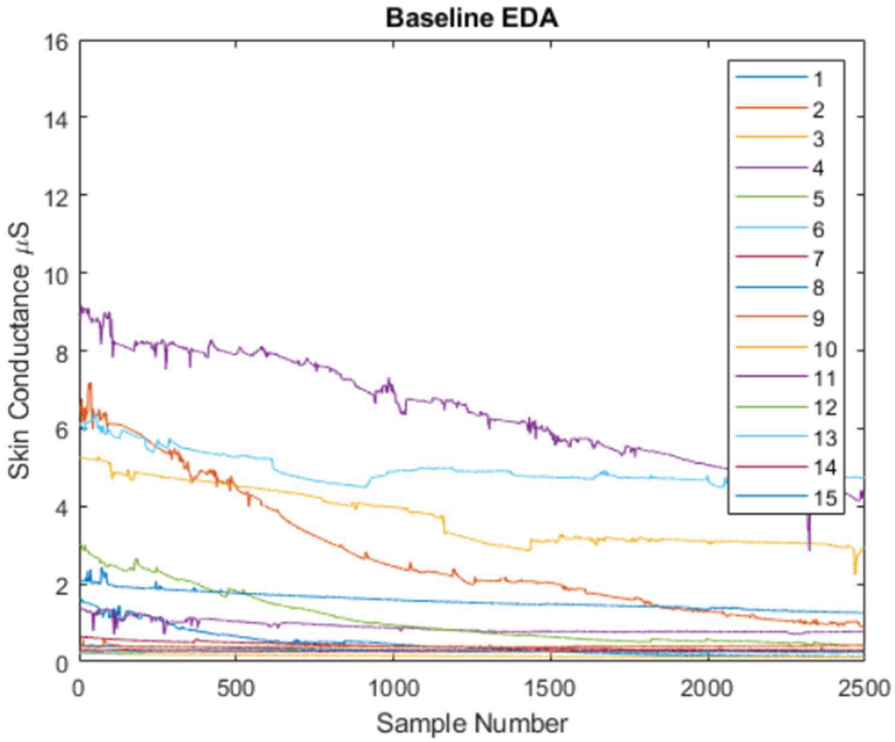

Figure 8 Example time domain EDA signal during the baseline condition

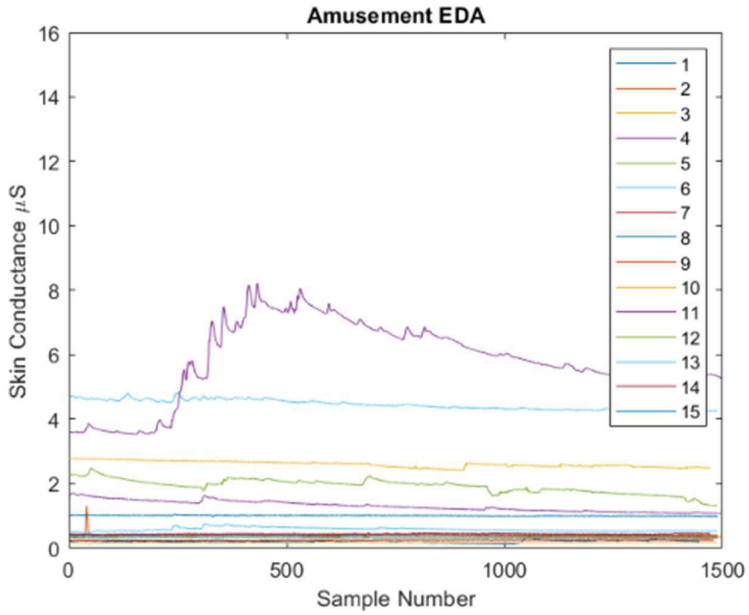

Figure 9 Example time domain EDA signal during the amusement condition

3.2.5.2. Intrinsic Mode Functions

EDA traces are non-stationary signals meaning the time-period and frequency are variable. This violates the assumption of periodicity made by the Fourier transform and suggests an alternative method should be used to analyse frequency content. One such method is finding the Intrinsic Mode Functions (IMFs) via Empirical Mode Decomposition (EMD), which accommodates for signal variability whilst also extracting its "periodic" components. This process is repeated until the resulting signal is monotonic. The EMD algorithm is shown in Figure *10* (Maheshwari & Kumar, 2014).

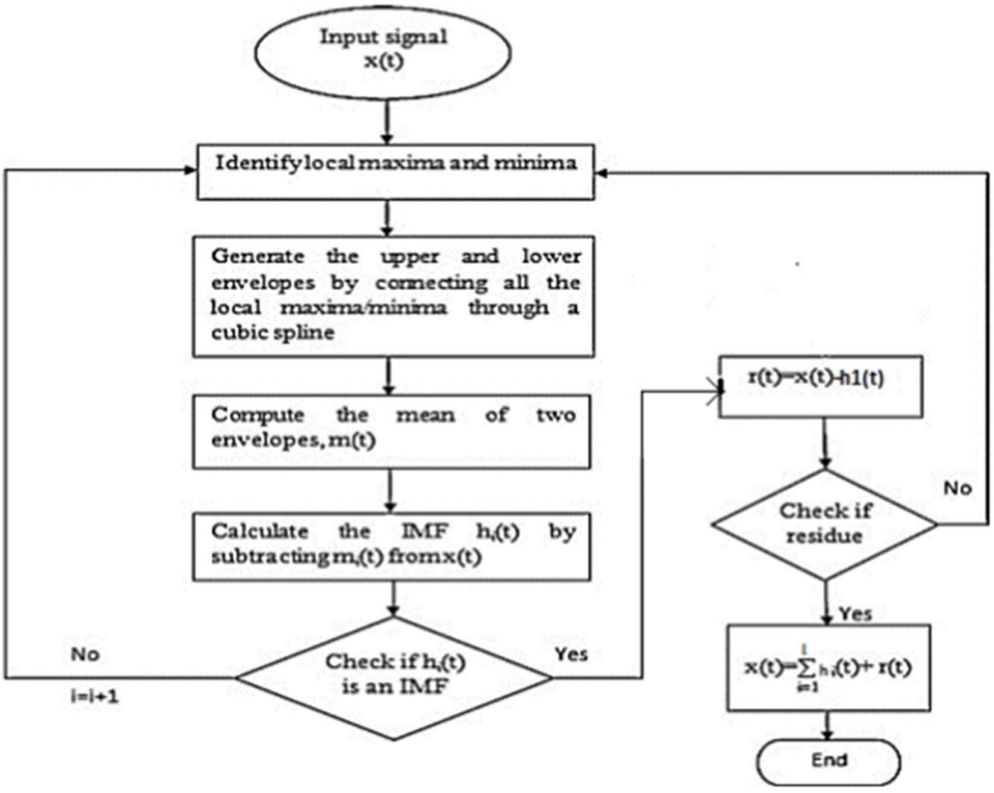

Figure 10 Flow chart of EMD process (Maheshwari & Kumar, 2014)

This process produces several IMFs each describing a separate frequency band over time, with the first representing the highest frequency band and the last representing the lowest.

Figure 11 shows the first IMF extracted from Participant 3's EDA trace. The stress signal has considerably more high frequency content than the baseline signal, as expected intuitively. *Figure 12* shows the last IMFs extracted, i.e. the lowest frequency bands, and both are monotonic functions suggesting the algorithm worked as expected. Magnitudes of both IMFs are similar but with opposite polarity. The distinct differences between conditions in both cases suggests this representation offers classification power. Again, descriptive statistical measures were used to summarise the IMF signals efficiently and used as features.

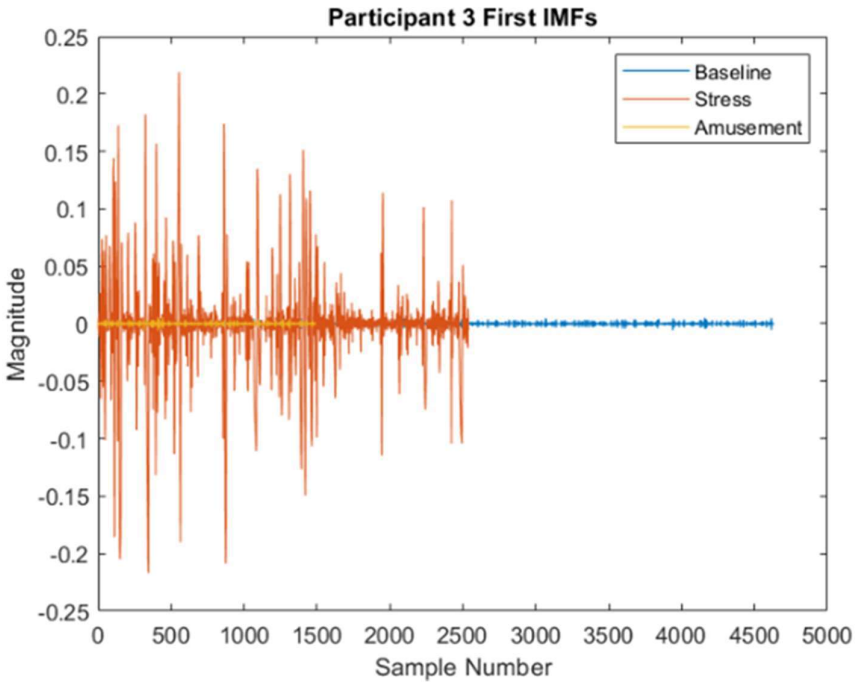

Figure 11 Comparison of first IMFs for stress and baseline conditions participant 3

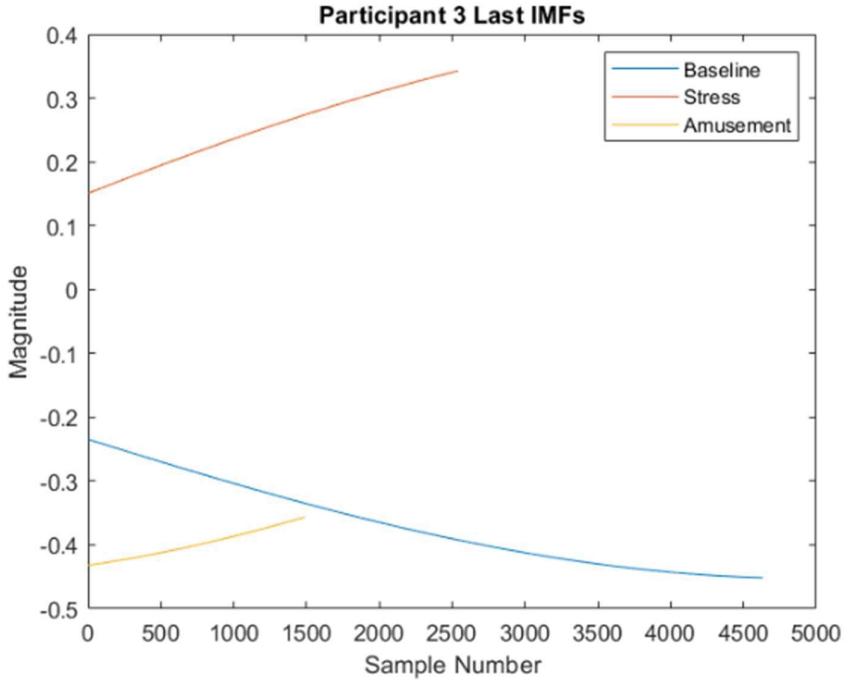

Figure 12 Comparison of last IMFs for stress and baseline conditions participant 3

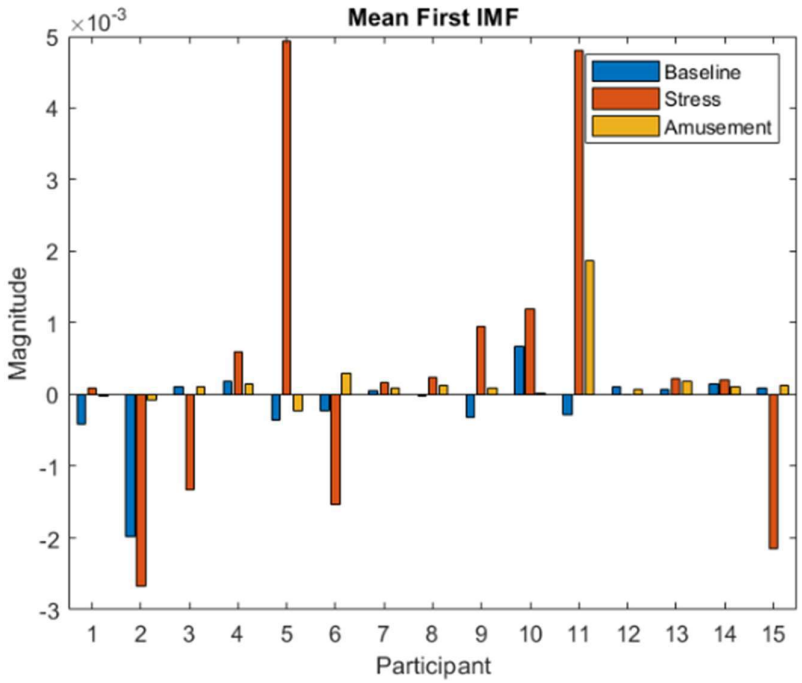

Figure 13 Comparison of first IMF mean for each condition of all participants

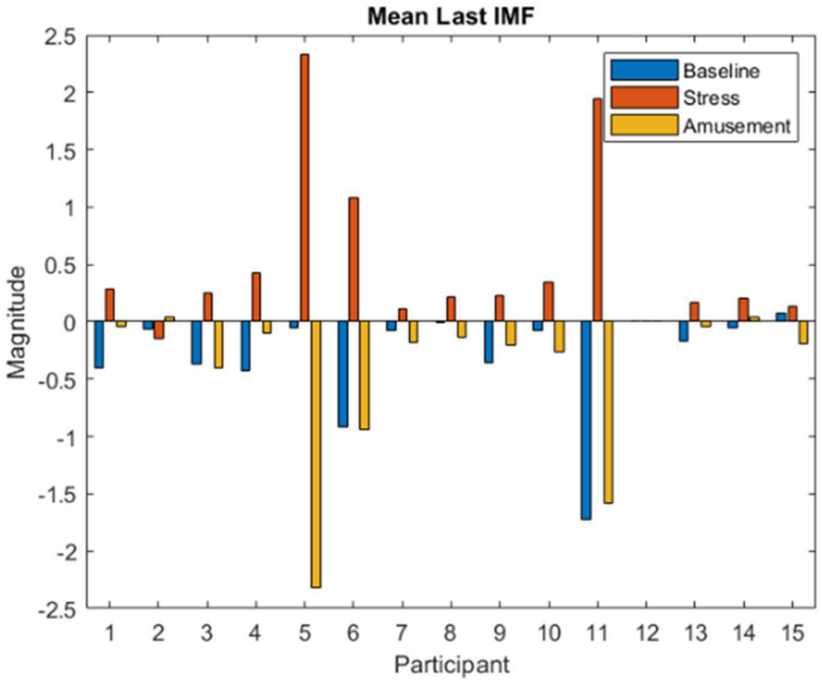

Figure 14 Comparison of last IMF mean for each condition of all participants

3.2.5.3. First Derivative

My research established that each condition contains different frequency components which suggests that the rate of change of the signal may also provide predictive power. This is supported when analysing the first derivative of the three conditions as shown in Figure 15. The stress signals show greatest rate of change since the sweat glands are innervated more frequently through the stress response, producing frequent changes in SC. Mean and standard deviation were extracted to describe central tendency and spread of the derivative signal. Peak frequency, peak mean, and peak RMS have been applied successfully in a previous stress detection study and thus were also used in this research project (Nath et al., 2020).

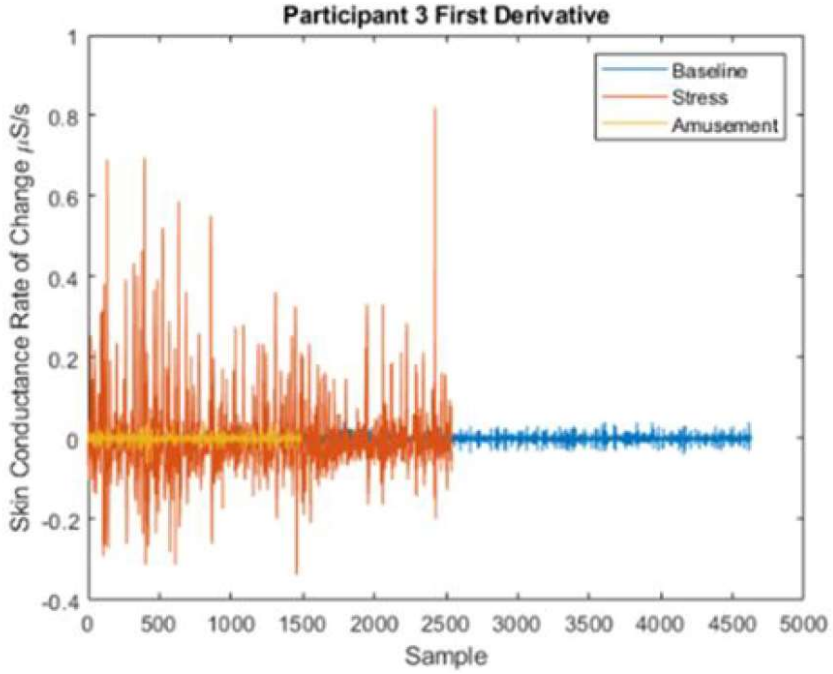

Figure 15 Example first derivative

3.2.5.4. Mel Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCCs) are frequently used in speech recognition and music processing systems. The cepstrum of a signal describes periodicity in the log power spectrum, also known as quefrencies and is calculated using Equation *1*, where $C(\,)$ is the cepstrum of the time domain signal $x(t)$, F is the Fourier transform, and F^{-1} is the inverse Fourier transform.

$$C\big(x(t)\big) = F^{-1}[\log(F[x(t)])]$$

Equation 1

The cepstral domain can be used to identify and separate convolved signals that contain considerably different frequency content, such as glottal pulses and vocal tract filtering, or in this case sweat gland innervation pulses x_t and the gland impulse responses h_{tonic} and h_{phasi} (Benedek & Kaernbach, 2010a).

$$EDA = x_t * [h_{tonic} + h_{phas}\ ]$$

Equation 2

The Mel scale is a representation of frequency based on the human perception of pitch and transforms frequency so that equal differences in Mel have the same perceptual difference. A filterbank of 40 evenly spaced triangular filters with centres ranging between 0 and 2Hz (Fs/2) was defined in the Mel domain and converted to frequency by rearranging Equation *3*, X is the frequency domain signal.

$$Mel = \ 2595log10(\frac{X}{700} + 1)$$

Equation 3

The EDA signal was segmented. A hamming window was applied to reduce spectral leakage and then converted to the frequency domain via FFT. Each segment was multiplied by the Mel filter bank in the frequency domain, and then a log was taken of the magnitude. The discrete Cosine transformation was then applied giving 40 MFCCs, of which only the first 13 were retained as done by Shukla et al.

The relevance of this scale is clear with speech recognition systems since the vocal signal is an acoustic signal, but questionable in this context since the data does not relate to psychoacoustics. The role of the Mel scale here is to create equally distributed coefficients in the cepstral domain that give a succinct time-frequency representation of the signal. MFCCs are not as commonly used as the other features in this set, but performed better than all other groups in a recent and comprehensive feature comparison study, and so were included to evaluate their performance with this dataset (Shukla et al., 2019). Figure *17* shows the MFCCs for participant three, interestingly each of the positive coefficients are zero for all conditions, an explanation for this could not be established but it meant that only the negative

coefficients were useful for classification.

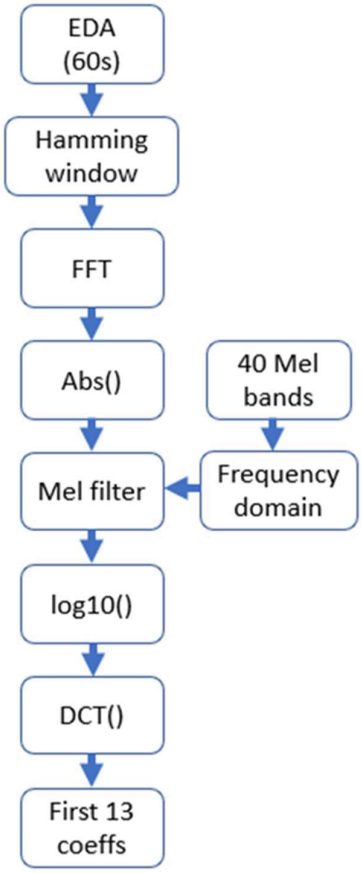

Figure 16 Calculation of Mel frequency cepstral coefficients

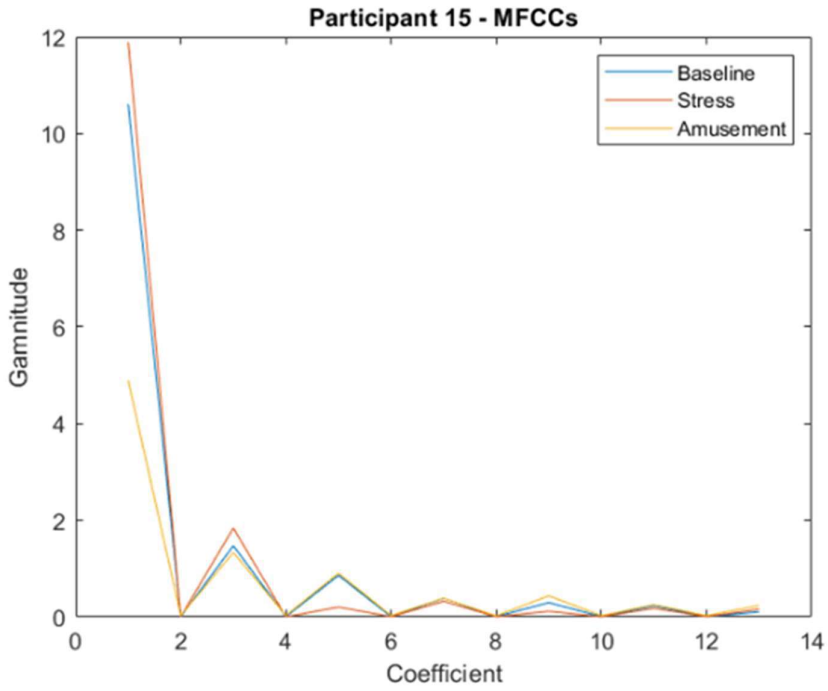

Figure 17 Example Mel Frequency Cepstral Coefficients for each condition

3.2.6. Classification Algorithm

3.2.6.1. Algorithm Choice

The role of the classification algorithm is to learn the mapping of features onto classes so that new data can be accurately classified, algorithm choice is therefore heavily dependent on the data and aim. For this reason, no algorithm is best for all problems and there are usually multiple candidates that could perform well meaning previous systems must be studied and compared to understand what has worked well with similar datasets before (Gaurav & Patel, 2020). It is also important to understand how each candidate algorithm works so implementation time and suitability can be determined.

Algorithms with many hyperparameters may offer flexibility but this also creates a complex fitness landscape that can be time consuming to search through to get the optimal solution (Gressling, 2020). If development time is limited, it can be more practical to use a simpler algorithm.

Thirteen stress detection systems (Table 2) were studied, and candidate algorithms were chosen based on how often and successfully they were implemented giving three candidate algorithms: Random Forest, K Nearest Neighbour (KNN), and Support Vector Machine (SVM). Advantages and disadvantages of each were listed to help decide which to use. The

most common algorithm used for affect classification is the SVM which has consistently performed well but is known to struggle with training time/computational cost for large datasets (Cervantes et al., 2020). KNN is the next most often used algorithm and has performed well in multiple studies however can be sensitive to noise and outliers. The Random Forest method has been used less often but has performed very well where applied, prevents the need for data standardisation and requires little hyperparameter tuning (Schmidt, Reiss, Duerichen, & van Laerhoven, 2018). Random forests are designed such that each decision tree splits on different features, this decreases the correlation of the predictions which makes random forest models resistant to overfitting. Schmidt et al. provide numerous studies that successfully used the Random Forest algorithm for affect detection. Fernandex-Delgado et al. even describe them as "clearly the best family of classifiers" (Fernández-Delgado et al., 2014). For these reasons the Random Forest was used for this project.

Paper	Metrics	Affects	Algorithms	Performance
Portable System for Real Time Detection of Stress Level	EEG, ECG, EMG and GSR	Stressed, relaxed and neutral	LDA	86% accuracy
An Enhanced Fuzzy Algorithm Based on Advanced Signal Processing for Identification of Stress	GSR, HR, and ECG	Stressed, medium stress and relaxed	Soft computing	-
Ustress Understanding College Student Subjective Stress Using Wrist Based Passive Sensing	GSR, HR	Intended stress and self reported stress	RF	88.8% F-Measure
Real Time Mental Stress Detection Based on Smartwatch	GSR, RR Interval and Body Temperature	Stress and relaxed	KNN Classifier	84.5% accuracy
Towards an Anxiety and Stress Recognition System for Academic Environments Based on Physiological Factors	HR, S02, ST, GSR and BR	Stress and non-stressed	KNN, SVM, LogR and RF	95.98% KNN, 94.44% Random forest
Validating Physiological Stress Detection Model Using Cortisol As Stress BioMarker	GSR + PPG	Stress and non-stressed	RF	92% accuracy
Realisation of Stress Detection Using Psychophysiological Signals for Improvement of Human-Computer Interaction	Blood Volume Pulse (BVP), GSR and Pupil Diameter	Stress and non-stressed	SVM - Sigmoid Kernel, RBF, linear	80% sigmoid kernel, 60% RBF, 57.14% linear

Non-Intrusive Physiological Monitoring for Automated Stress Detection in Human-Computer Interaction	Blood Volume, GSR, ST and PD	Stress and relaxed	SVM, Decision tree, Naive Bayes	SVM 90.10%, Decision tree 88.02%, Naive Bayes 78.65%
Stress Detection From Speech and Galvanic Skin Response	GSR, Speech	Calm vs heavy workload	SVM	80.7%+-0.6
Personalised Stress Detection from Physiological Measurements	ECG, GSR, Respiration RIP, ST - 26 features gathered total	Stress and non-stressed	SVM	0.67 precision with 80% recall
Emotion Recognition via Galvanic Skin Response: Comparison of Machine Learning Algorithms and Feature Extraction Methods	GSR	Valence and arousal (high and low)	RF	81.81% arousal and 89.29% valence
CStress: Towards a gold standard for continuous stress assessment in the mobile environment	ECG	Stress and non-stressed	SVM	72% accuracy
Combined analysis of GSR and EEG signals for emotion recognition	GSR, EEG	Valence and arousal	SVM and KNN	88% accuracy user dependent, 72% accuracy group
Electrodermal activity sensor for classification of calm/distress condition	EDA	Stressed and calm	Decision trees	89% accuracy
Human Emotion Recognition Using Deep Belief Network Architecture	EDA, PPG and zENG	Happy, Relaxed, Disgust, Sad and Neutral	DBN and SVM	89.53% accuracy
Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity	EDA	Valence and arousal	SVM, using three different feature selectin algorithms	85.75% accuracy and 0.63 F1 arousal classification, 83.9% accuracy and 0.61 F1 for valence classification
Discriminating Stress From Cognitive Load Using a Wearable EDA Device	EDA	Stress and cognitive load	LDA, SVM and NCC	82.8% accuracy

Table 2 Summary of methods used for affect classification

3.2.6.2. Random Forest Classifier

A Random Forest classifier is a supervised learning ensemble method made up of many decision trees (Breiman, 2001; Breiman & Cutler, 2001). Decision trees are rule-based

algorithms made up of a root node, internal nodes, and leaf nodes. The root node is the feature that offers maximum information gain as measured by the Gini impurity and each subsequent node offers the next most information gain until all features are used. The Random Forest algorithm creates N bootstrapped datasets with a random selection of features and trains a decision tree on each set. Decisions are then made by aggregating the output of each tree, the process of which is known as bootstrapped aggregation.

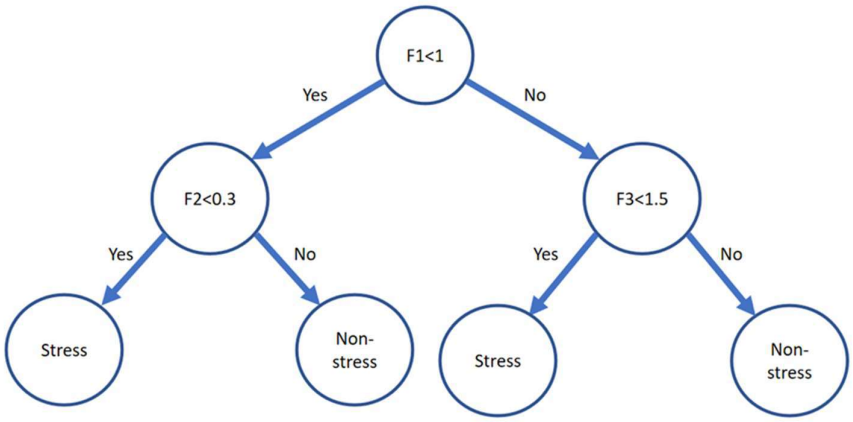

Figure 18 Decision tree schematic diagram

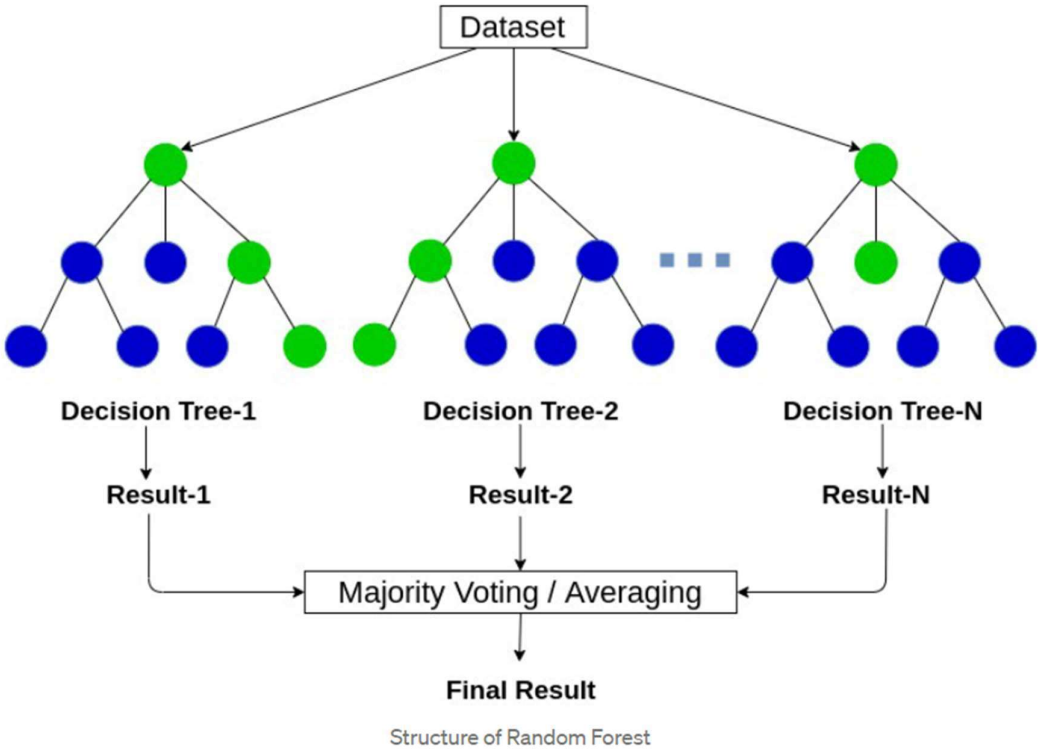

Figure 19 Random Forest diagram

3.2.7. Feature Selection

The best combination of features from the extracted set must be found to reduce computation time, increase accuracy, and increase interpretability. Shukla et al point out that feature selection is rarely applied in affect recognition using EDA, and when some form of feature selection is applied it is often merely dimensionality reduction techniques such as Principal Component Analysis. Such techniques are not best suited to the problem since they do not consider the target variables. Instead, they propose using information-based methods such as Joint Mutual Information, Conditional Mutual Information Maximization, and Double Input Symmetrical Relevance (Shukla et al., 2019). Unfortunately, there was difficulty in finding applications of these methods in the literature and very few examples of implementations in MATLAB, leading to the consideration of the more frequently used filter and wrapper methods. Filter based selection relies on using statistical techniques such as Pearson correlation to check for feature redundancy. If two variables are strongly correlated, they provide similar information to the model which can reduce performance. In such a case, one of the redundant features can be removed, or both can be merged into a new feature. Wrapper methods use the classification algorithm and some predefined evaluation criteria to find the subset that gives optimal performance. Filter methods are fast making them particularly useful when model runtime for training and testing a model is long, yet they don't consider performance criteria and in this context the training time was not time consuming so a wrapper method was used.

The process by which the feature set is searched through is known as the search procedure and the varying methods are described in detail by Kuhn & Johnson (Kuhn & Johnson, 2013). In summary, there are four main types of search procedure: backward elimination, forward selection, stepwise selection, and all possible subset selection (Chowdhury & Turin, 2020). Firstly, some performance criteria must be defined which is dependent on the problem but for classification tasks this often involves accuracy and f-measure. Backward elimination begins with the full feature set and iteratively removes redundant features until only those that significantly contribute to the model are left. Forward selection is the opposite, beginning with an empty feature set and adding those that contribute most until a threshold is achieved or a knee occurs. Stepwise selection combines the two, adding variables that contribute most and then rerunning the new feature set so that features that become redundant can be removed. All possible subset feature selection is a brute force technique, training and running the model with all feature combinations. Forward and backward selection both risk missing the optimal subset since, for forward selection, no check for redundancy is performed once a

new variable is added, and for backward selection once a variable is removed it cannot be readded.

Accuracy is defined as the correct prediction rate and appears to be the most cited metric in classification. A downfall of accuracy is that it does not sufficiently describe which groups the misclassification occur most in and can falsely represent performance in cases where there are class imbalances.

$$accuracy = \frac{correct\ predictions}{all\ predictions}$$

Equation 4

Precision and recall are used for a more complete description of classification performance. Precision describes the certainty of true positives, if there are many false positives then the model will lack precision. Recall describes how many positive cases are correctly classified compared to all positive cases, if the model misses positive cases, it lacks recall.

$$recision = \frac{TP}{TP + FP}$$

Equation 5

F score (also known as F-Measure or F_1) combines both precision and recall into one metric via Equation *6*. F_1 ranges from 0 to 1 with 1 being the best possible value and any reduction in either recall or precision produces a reduction in F_1 since it is a uniformly weighted combination of the two.

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall}$$

Equation 6

3.3. Algorithmic Composition

This chapter describes the design of the music generation system. Firstly, justification for using MIDI as the music representation method will be presented followed by a description of the bespoke dataset created for this project. The theory behind Markov models for music will then be explained and the rules used to manipulate musical parameters are then presented. Lastly, the method used to pass data from MATLAB to Analog Lab 4 for audio output is described.

3.3.1. Representation

Music can be represented in a variety of ways such as standard music notation, guitar

tablature, Musical Instrument Digital Interface (MIDI), and one-hot encoding. Each have advantages and disadvantages, so the aim of the project and availability of training data must be considered. MIDI is a technical standard used by many modern electronic audio devices. Information about note value, velocity, and duration (as well as value of modulation controls), are encoded as hexadecimal numbers which can be processed by computers. Below is an example MIDI message showing a note-on and note-off message.

Command	**Meaning**	**Number of Parameters**	**Parameter 1**	**Parameter 2**
0x80	Note-on	2	Key	Velocity
0x90	Note-off	2	Key	Velocity

Table 3 Example MIDI message

One-hot encoding transforms chords and melodies into a binary matrix so it can be processed by algorithms. In this system a bespoke training set was created (described in the following section) and was represented using MIDI during input and output stages, but one-hot encoded during the generation process.

3.3.2. Dataset

Many datasets created for training generative models use complex genres such as classical and jazz pieces (Magenta's MAESTRO dataset and the University of California's Bach Chorales dataset). This is because classical and jazz pieces have less defined structures more often than in other genres such as pop. Structure is particularly difficult to recreate using algorithms since it requires high level memory. That said, these genres are relatively complex harmonically, rhythmically, and melodically, which complexity can produce an unmusical output as the model struggles to learn the patterns brought about through complex and abstract music theory.

When searching for music for stress alleviation into search engines such as YouTube the results tend to be multiple hours' worth of rich, slowly changing chords with few percussive elements (Healing Soul, 2022; Yellow Brick Cinema - Relaxing Music, 2023). These pieces tend to be harmonically simple compared to jazz and classical music and so the existing datasets were not considered suitable. A dataset was made from 28 common British/American children's TV songs and nursery rhymes (full list given in Appendices).

These MIDI files were pre-processed for maximal compatibility with later stages. Firstly, all melodies were imported into Studio One 4 (a Digital Audio Workstation) and processed so that only the main melody remained. All note overlaps were removed, and all notes were quantised to a semi-quaver grid. Longer melodies were split into multiple 2 bar sections, meaning the 28 songs produced a dataset made up of 81 files. Lastly, all files were saved as MIDI in a single folder.

3.3.3. Generative Algorithm

3.3.3.1. Markov Model

First order Markov Models (MM) are some of the simplest forms of generative algorithm but can produce convincing music whilst being relatively straightforward to implement and quick to train. The Markov model builds upon the first order Markov assumption given below where $q_1, \dots q_t$ is a set of states (Schulze & Merwe, 2011).

$$P(q_{t-1}, q_{t-2}, \dots q_1) = P(q_{t-1})$$

Equation 7

Chains are built by connecting states using transition probabilities and the order of the MM determines the number of previous states included in the transition matrix. In this case the states are notes and the transition probabilities are the conditional probability of the next given note given the notes played before. Figure *20* shows a diagram of a Markov Model with two notes and all possible transitions have an associated probability forming the transition matrix. Notes were then be chosen by creating a pseudorandom value ranging from 0 to 1 and assigning a threshold based on the probabilities given in the transition matrix. If the note is currently on C and the pseudorandom number is below 0.6 then the next note will be C, if above 0.6 the next note will be D.

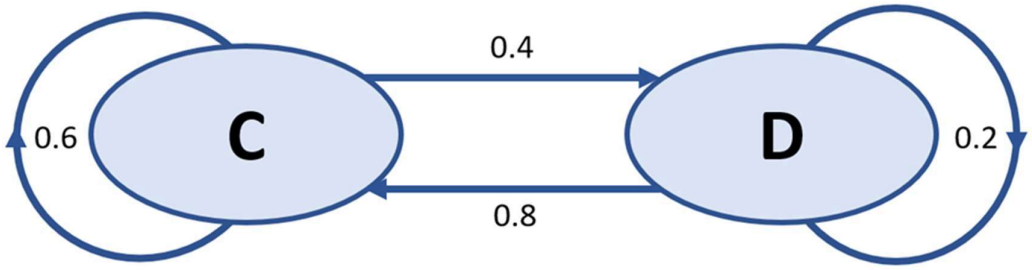

Figure 20 Example of Markov chain for generative music

Although Markov models are easily implemented and can create new, musical melodies without much data to train on, user control of musical parameters is limited, and melodies can excessively mimic the training data particularly when higher order models are used. This research only a first order Markov model was implemented.
Generation was achieved by importing the MIDI data and transposing to C major or A minor. The system created a one-hot encoding according to all note values in the dataset and then calculated a first order transition matrix, from which 2 bar melodies were produced by initially choosing a note from uniform probability distribution, then choosing the next note by sampling from the transition matrix.

3.3.4. Music Parameters and Modulation Implementation

This system was designed using ideas common in western societies; hence it uses the equal temperament tuning with A4 being 440Hz. When two notes are played simultaneously the resulting pressure is the sum of both at any moment in time. The human ear converts these changes in pressure into analogous electrochemical signals via the tympanic membrane, ossicles and cochlear. Real acoustic signals, such as piano notes, contain overtones and when there are multiple overtones at different frequencies but within a critical bandwidth the sound is perceived as being dissonant. This relates to the ratio between intervals since overtones often occur at roughly integer multiples of the fundamental frequency, if the ratio is simple such as 2:3 (a perfect fifth), the overtones align making it the most consonant interval excluding the octave. If the ratio is irrational the overtones align less frequently but still occur within a critical bandwidth making the resulting sound dissonant.
Scales are cyclic lists of intervals, for example the major scale is built by increasing the note value by T, T, ST, T, T, T, ST where T stands for tone and ST stands for semitone, this spans an octave and repeats on completion. Each scale type has a different combination of intervals which changes the degree of consonance and therefore the emotional associations usually made. Chords are built by simultaneously combining intervals producing a mixture of consonance and dissonance, which is also affected by the voicings used. Musicians take advantage of consonance and dissonance to create tension, release and surprise throughout a piece which relate to the emotional associations, but these characteristics are difficult to replicate using algorithms.
Tempo, modality and note density were modulated to guide a user to a less stressful mental state. The emotional associations of tempo is a well-studied area and tempo is known to

generally alter emotional arousal rather than valence. Low tempo correlates with low arousal and vice versa. Modality is another well studied musical parameter that has strong emotional associations. Modality describes the tonality of a musical piece and correlates more with valence rather than arousal. There are seven so called “church modes” each built from starting on a different degree of the major scale. The seven modes beginning on C are shown in Table 4.

Mode	**Scale Degree**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Ionian	C	D	E	F	G	A	B
Dorian	C	D	Eb	F	G	A	Bb
Phrygian	C	Db	Eb	F	G	Ab	Bb
Lydian	C	D	E	F#	G	A	B
Mixolydian	C	D	E	F	G	A	Bb
Aeolian	C	D	Eb	F	G	Ab	Bb
Locrian	C	Db	Eb	F	Gb	Ab	Bb

Table 4 Table of notes in all modes of C

The brightness theory of modality states that the emotional valence that each mode tends to be associated with is caused by the number of flattened or sharpened notes compared to the natural major scale. For example, pieces written in Phrygian are frequently voted as being the darkest mode (ignoring Locrian which is rarely used due to the flattened 5^{th}) whereas Ionian is judged as the happiest. A flaw in this theory is that the Lydian scale would be expected to be voted as the happiest since it has a sharpened fourth with no flats, yet it is usually second happiest (Ramos et al., 2011).

Combining the brightness theory of modality with the iso-principle two ideas provided a method for how the modality mapping could be implemented. Music that reflects the user’s emotional state is played initially (Phrygian mode, fast tempo, and high note density), and then parameters are gradually modulated in the direction of higher valence and lower arousal to calm the user. The heart rate of the user was also considered since when stressed the beats per minute (BPM) is normally higher than average. Heart rate is controlled by the ANS and so a reduced BPM often indicates a lower arousal emotional state, suggesting a reduced stress level. The tempo was modulated by the stress classification using the equation given below, where the minimum and range of the BPM are arbitrary. but 60 and 80 were used

respectively since they cover the common range.

$$BPM = minBPM + (stressModulation * rangeBPM)$$

Equation 8

It was initially thought multiple Markov models could be trained, each on a specific mode and during runtime the model in use could switch dependent on the stress level. Unfortunately, no MIDI datasets arranged according to modality. Instead, a note value quantisation was applied after each two-bar melody was produced, forcing all notes to the required mode beginning on C. The system was set to begin using the Phrygian mode with high tempo as per the iso-principle but to modulate as the stress level decreased, each of the six modes used were assigned to equally spaced thresholds.

When musicians perform, they don't play with quantised rhythm and uniform note velocity. Instead, they extend and contract sections of the piece, vary note articulation, and alter dynamics dependent on the emotion intended, this is known as expression. A common downfall of computer-generated music is that the performance sounds mechanical and thus loses emotional content. To add expressiveness, a gaussian curve with mean of 70 and standard deviation of 10 was made, this was sampled for each note in the melody so the notes wouldn't have uniform velocity. Unfortunately, the start and end time of each note could not be moved off integer multiples of 0.125 due to the incremental counter used as a clock in Pure Data, a more advanced system would have expression controls to add a human feel. This could be implemented using SuperCollider.

3.3.5. Pure Data

The Pure Data project resembled a sequencer with a clock controllable via MATLAB for tempo modulation. Each generation produced 2 bars of melody, so the sequencer had storage capacity for 16 notes with individual velocity, pitch, and timestamp. A counter increased in 0.125 increments (the duration of a quaver relative to a bar) on each click of the metronome, and when the timestamp of each note equalled the value of the counter the note would be activated and sent as MIDI to Analog Lab 4 for synthesis.

3.3.6. Synthesiser

The last part of the system was audio output, MIDI data was sent from Pure Data to Analog Lab 4 via the LoopBe1 Internal MIDI Port, any standalone synthesiser plugin could be used

in place of Analog Lab 4 highlighting the flexibility of the system.

4. Results

4.1. Stress Detection

4.1.1. Subject Dependent Full Model

Multiple random forest models were trained to understand how the number of trees, feature domains and feature selection methodology affect the resulting model performance. Initially, a model was trained using all the features extracted from the EDA signal. Training and testing sets were created using random sampling with an 80:20 training to testing ratio. This model had an accuracy of 99% and an F-measure of 99% suggesting it was overtrained. This was due to using random sampling which meant that observations from adjacent windows could exist in training and testing sets, hence the model was tested on very similar data to that which it was trained on, suggesting the methodology was flawed.

A training set was therefore created from the first 80% of each condition and the last 20% was used as the testing set. This meant that although some data was shared in the first 12 windows of the training set for each condition due to the overlap, the remaining windows would contain data that had not been seen before. This model had an accuracy of 82% and an F-measure of 89%. It was trained using 100, 250 and 500 trees and there was a negligible increase in accuracy and F-Measure, hence 100 trees were used for the remainder of the project. This testing methodology mimics the in-situ use of the system more realistically since the system would not have seen observations from that specific signal before and would have to apply patterns found through training to new data. Ideally, a validation set would be created using EDA data from a different dataset however this was not possible within the project timeframe.

4.1.2. Subject Dependent Model Feature Domains

Each feature domain was tested independently to understand which provided the most useful information as shown in Table 5. The results suggest that the features extracted from the first derivative offer the greatest predictive power whereas those extracted from the MFCCs offered the least.

	Accuracy	F Measure
First Derivative	80%	0.87
IMF	75%	0.84
Time Domain	74%	0.83
MFCC	72%	0.82

Table 5 Performance of subject dependent Random Forest model trained on each feature domain

4.1.3. Subject Dependent Forward Selection

Forward selection was used to iteratively select the best performing variables using F-Measure as the performance criterion. The optimal number of features was found by looking for the knee in accuracy and F-measure (Figure 21 and Figure 22). This method suggested the first 7-9 variables contained the most useful information since adding more after this point resulted in reduced performance. Since the first 7 variables provided similar performance to 9 with less information this is deemed to be the optimal feature set. The full feature selection list is given in Table 6. This optimal feature set gave an accuracy of 86% and an F-Measure of 0.94 which matches the performance of similar systems.

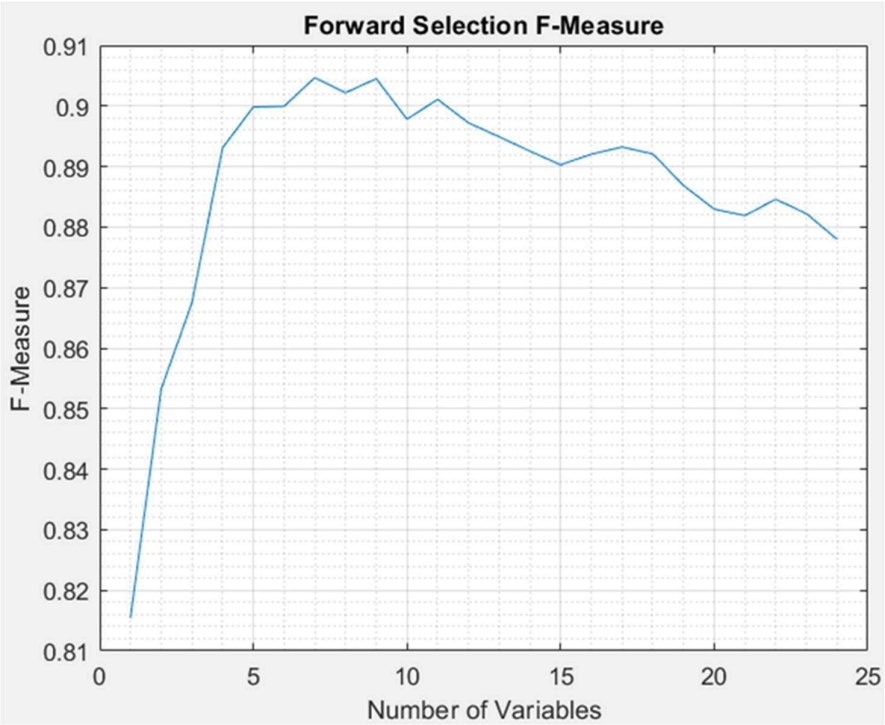

Figure 21 F-Measure per number of variables during forward selection for a subject dependent model

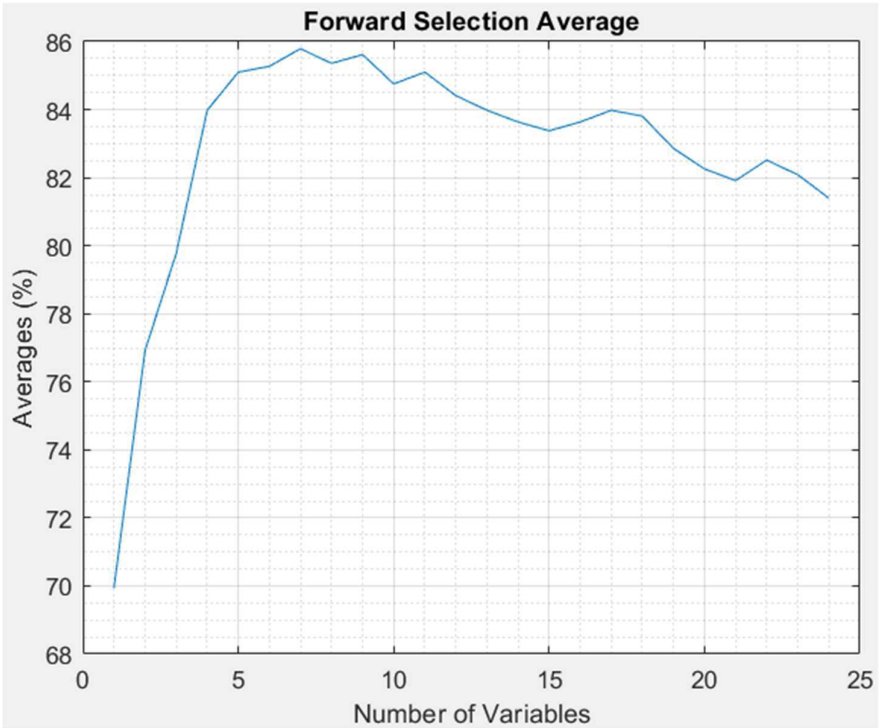

Figure 22 Average per number of variables during forward selection for a subject dependent model

Order Selected	Variable Name
1	First Derivative Peak Frequency
2	MFCC Skew
3	First Derivative Mean
4	First Derivative Peak RMS
5	Skew
6	MFCC Mean
7	IMF Kurtosis
8	IMF Skew
9	IMF Max
10	Kurtosis
11	MFCC Median
12	MFCC Standard Deviation
13	IMF Mean
14	Mean
15	AUC
16	IMF Standard Deviation
17	Minimum
18	First Derivative Peak Mean
19	First Derivative Standard Deviation
20	MFCC Kurtosis
21	Median
22	Maximum
23	Standard Deviation
24	IMF Minimum

Table 6 Order of features selected via forward selection

4.1.4. Subject Independent Full Model

Accurate subject independent emotion models are particularly difficult to create due to the individual variability in physiological signals. To understand how generalisable the features were, models were created with each participant held-out and used as the testing set. Perfect accuracy was achieved during training for all participants, however the average testing

accuracy was 74% with a minimum accuracy 40% and maximum of 92% (Figure *23*).
The variation between testing accuracies for each participant is due to the natural variation between how individuals' bodies react to stress. Those whose mapping of EDA data onto stress is different to the main group will have lower predication accuracy and therefore the generalisability of the model is limited. To increase generalisability more data could be gathered through using additional biosignals with an extended feature set, but considering it is easy to collect data from the individual it is best to create subject dependent models.

4.1.5. Subject Independent Feature Domains

The features extracted from the first derivative of the EDA signal performed best for both subject dependent and independent models suggesting those features were the most generalizable (Table *7*).

	Accuracy	F Measure
First Derivative	79%	0.85
MFCC	71%	0.78
IMF	68%	0.74
Time Domain	66%	0.72

Table 7 Performance of subject independent Random Forest model trained on each feature domain

The testing and training accuracies using each participant as a hold out set are given in Figure 24, Figure 25, and Figure 27. The testing accuracy is both participant dependent and feature domain dependent. This suggests that a sophisticated system should perform feature selection on an individual basis to find the optimal model.

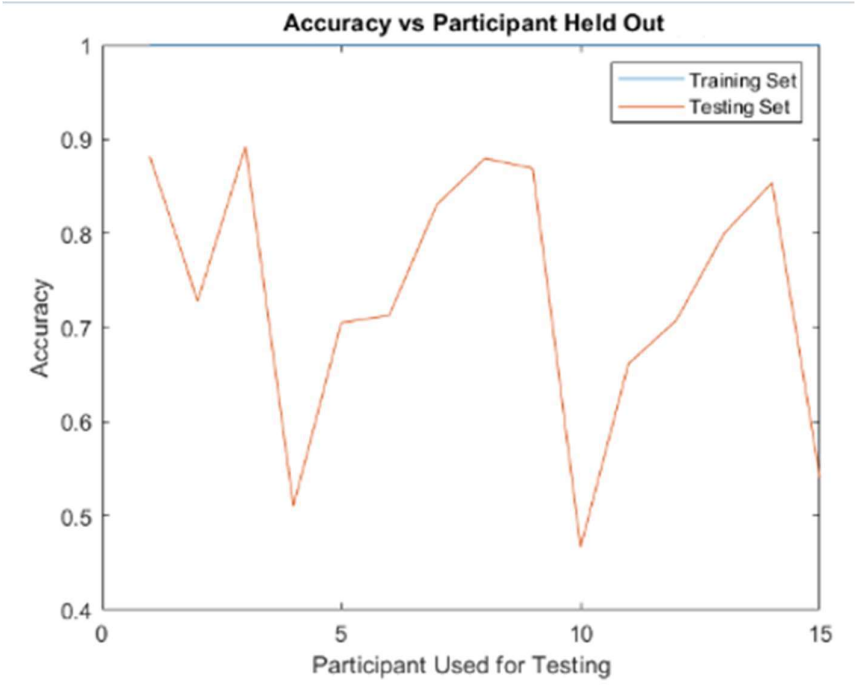

Figure 23 Full model approach using one participant used as a hold out set

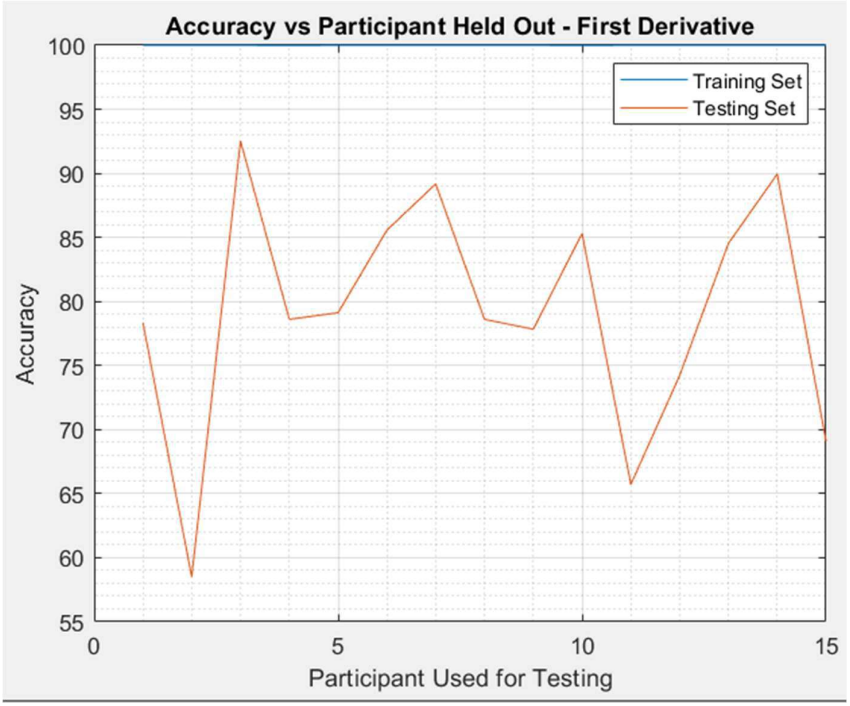

Figure 24 Testing and training accuracy using each participant as a holdout set and only features extracted from the first derivative of the EDA signal

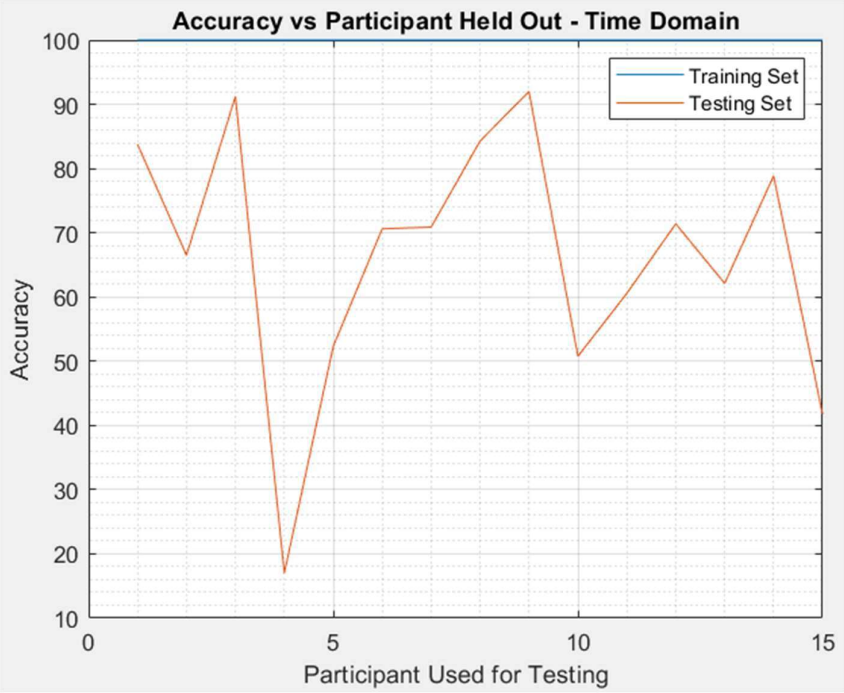

Figure 25 Testing and training accuracy using each participant as a holdout set and only features extracted from the time domain of the EDA signal

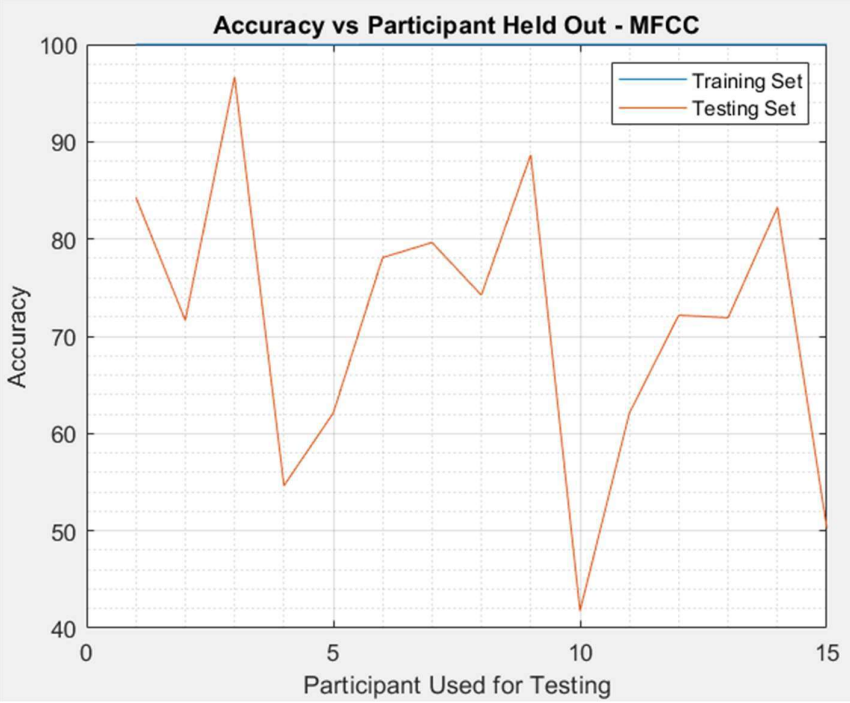

Figure 26 Testing and training accuracy using each participant as a hold out set and only features extracted from the MFCCs of the EDA signal

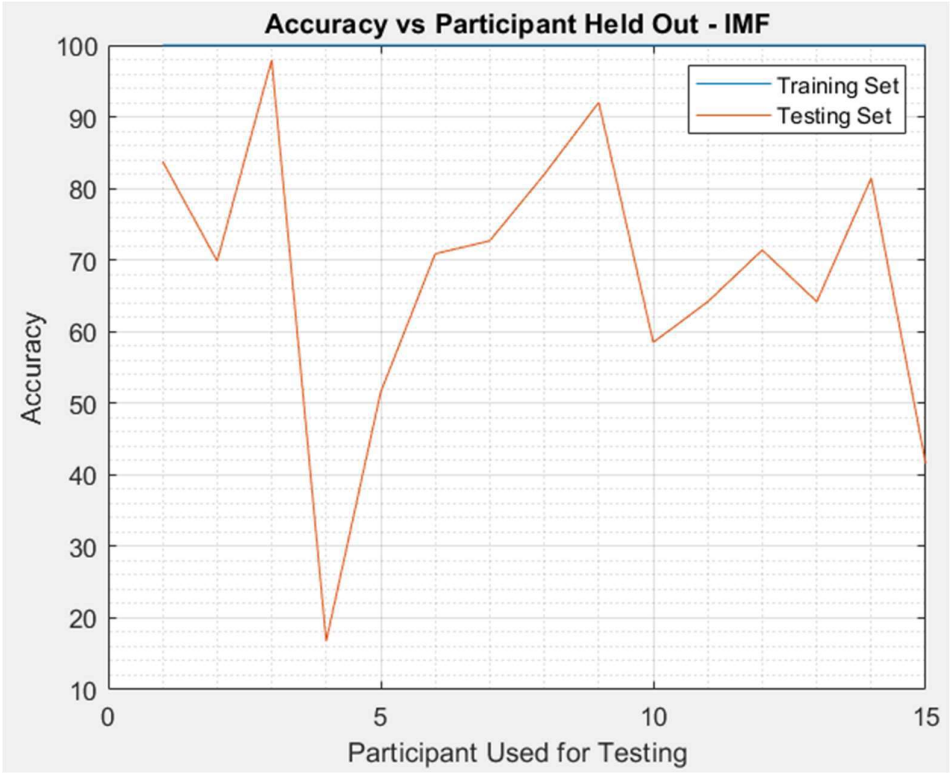

Figure 27 Testing and training accuracy using each participant as a hold out set and only features extracted from the intrinsic mode functions extracted from the EDA signal

4.2. Music Generation

4.2.1. Stress Modulation Signal

Initial tests were carried out by bypassing the stress classification algorithm and feeding a stress label into the music generation system. A step function was used for this purpose, where the stress label was 1 (stressed) for 90 samples and 0 (non-stressed) for 90 samples. This replicated someone going from a complete stress state to a complete non-stress state which would allow all modulation parameters to be checked. The audio output, MIDI and musical parameter values were recorded for analysis. As discussed, a moving average filter was used to transform the stress classification to on binary so that it could be used more effectively as a modulation parameter. This resulted with a form of time lag as shown in Figure 28.

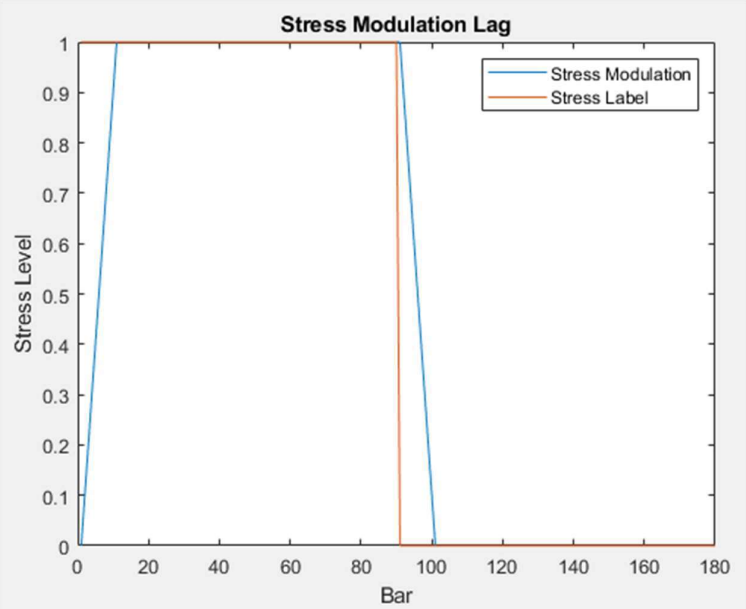

Figure 28 Classification lag due to moving average

Clearly, this method makes the stress classification less accurate, but it allows the musical parameters to be changed more gradually to induce the desired response and create more coherent music. To improve accuracy a non-binary classification algorithm must be used however this is dependent on the datasets available. In this case a binary classification algorithm was used since the training data was labelled as such.

4.2.2. Tempo Modulation

In accordance with the iso-principle the BPM rule was created such that each piece will always begin at the fastest possible setting to match the users' mental state. Figure *29* shows the rule was applied successfully since initially the tempo of the piece is elevated to match the elevated arousal, but this is followed by a gradual decrease in tempo designed to reduce arousal and guide the user towards a calmer state. This can also be heard in the audio output. The clear limitation of this is that the BPM (as well as other musical parameters) cannot be perpetually decreased, this is discussed further in later sections.

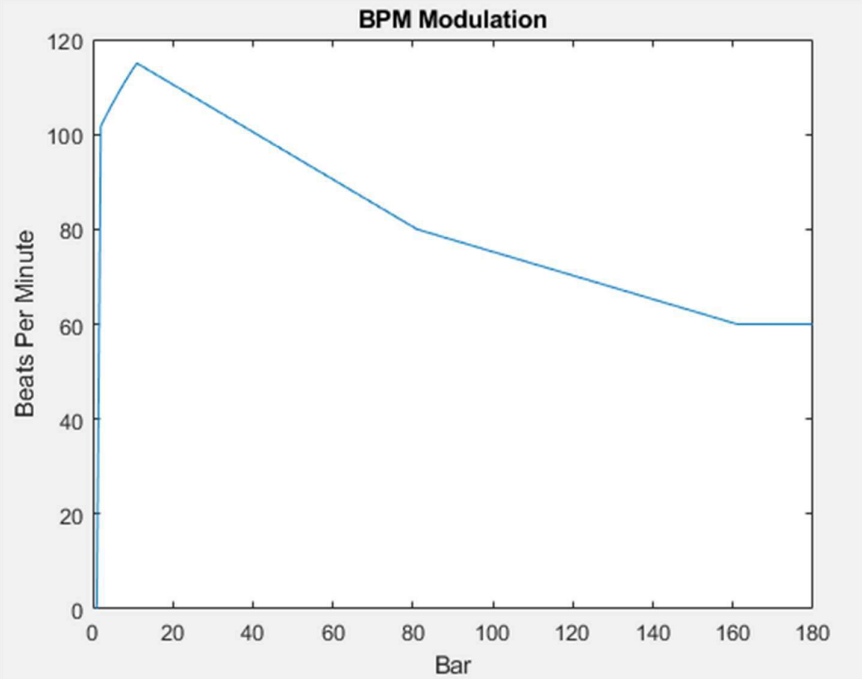

Figure 29 BPM used in each bar

4.2.3. Expression and Loudness Modulation

Loudness is known to correlate strongly with arousal hence the rule used to define loudness was like that of BPM. At the beginning of each piece the loudness was designed to be high to match the elevated arousal state.

Figure *30* shows that generally the mean velocity decreased over time with a maximum at the beginning and the gaussian sampling successfully created a greater dynamic range which

added a form of expression.

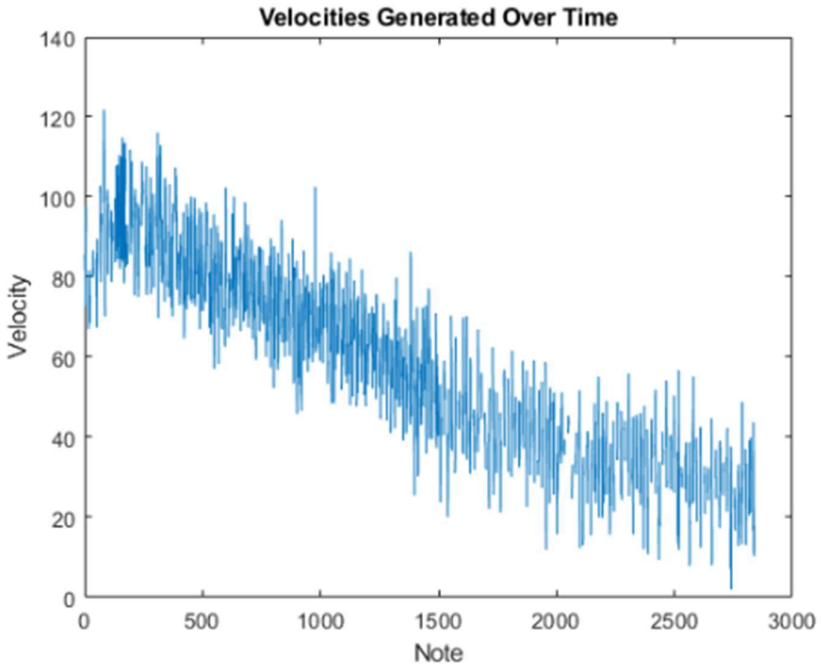

Figure 30 Note velocities generated

4.2.4. Modal Modulation

The variations in mode shown in Figure 31 shows that the iso-principle has been applied successfully but that the current system gets stuck in the Ionian mode. This emphasises the difficulty of creating explicit rules that can alter parameters to alleviate stress. Once the parameters have been altered so that they are at the last value possible, the current system will become stuck. The next step of this system will be to continuously change parameters without increasing stress. Distributions of notes generated for each mode confirm that the modal quantisation algorithm was implemented successfully.

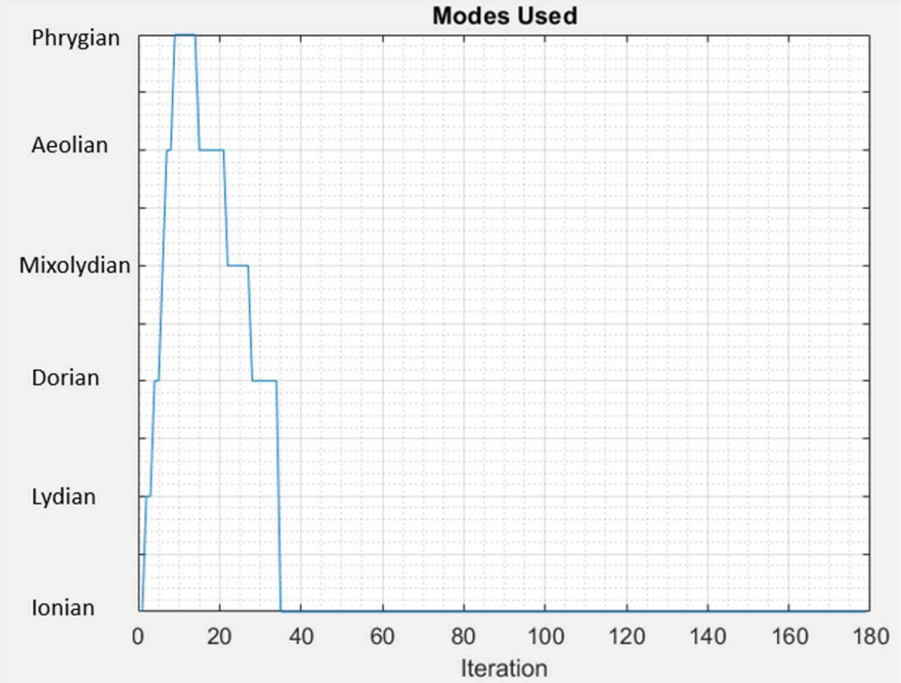

Figure 31 Mode contour throughout generated piece

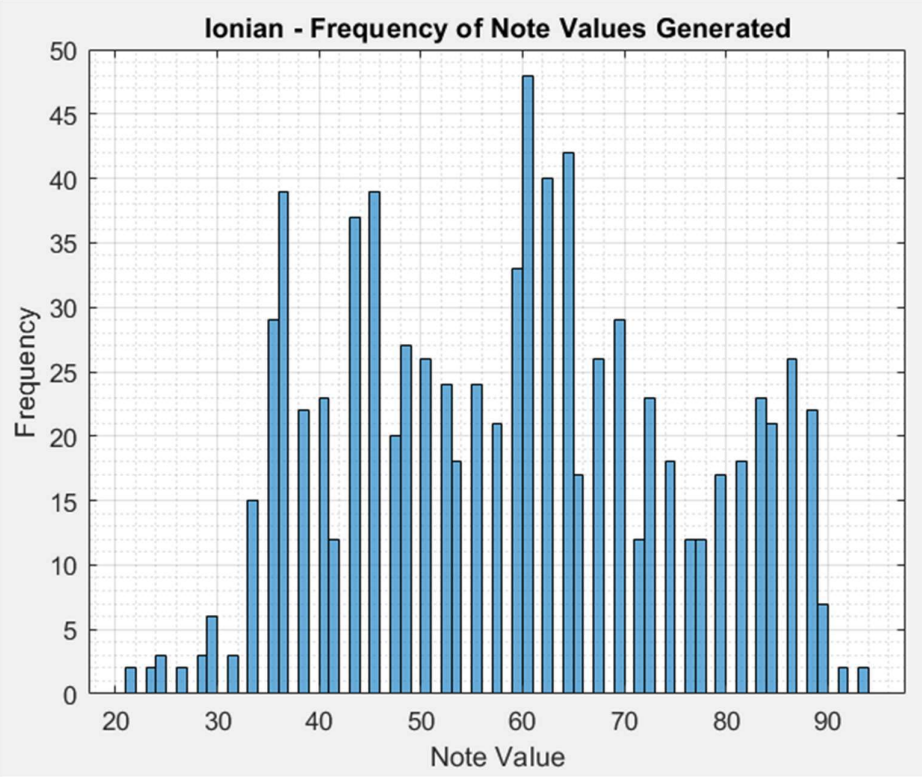

Figure 32 Distribution of Ionian melodies

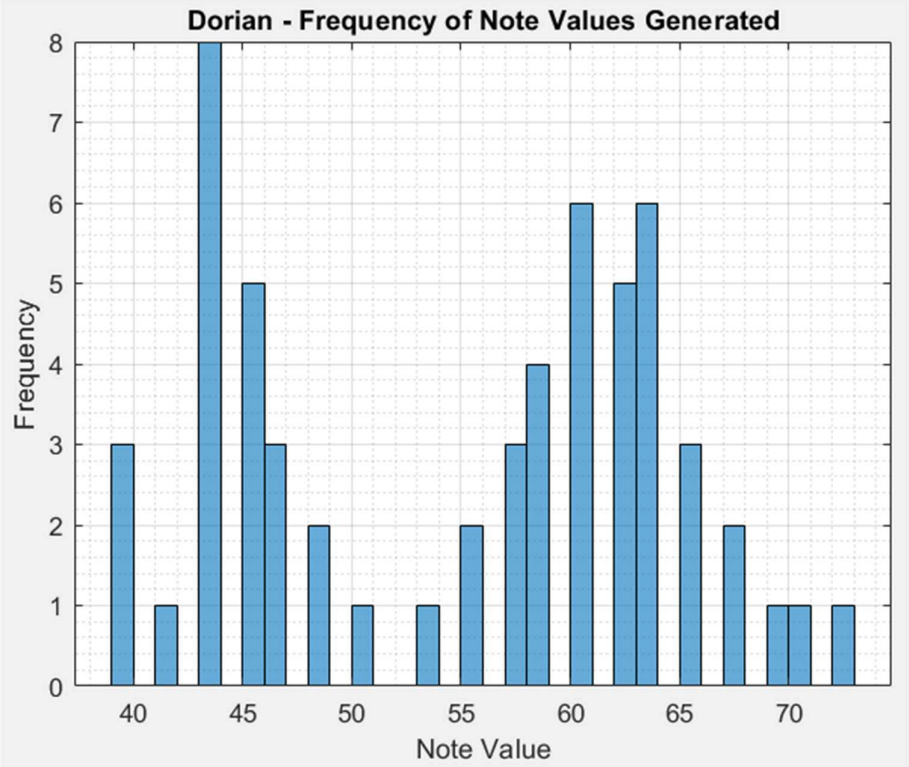

Figure 33 Distribution of Dorian melodies

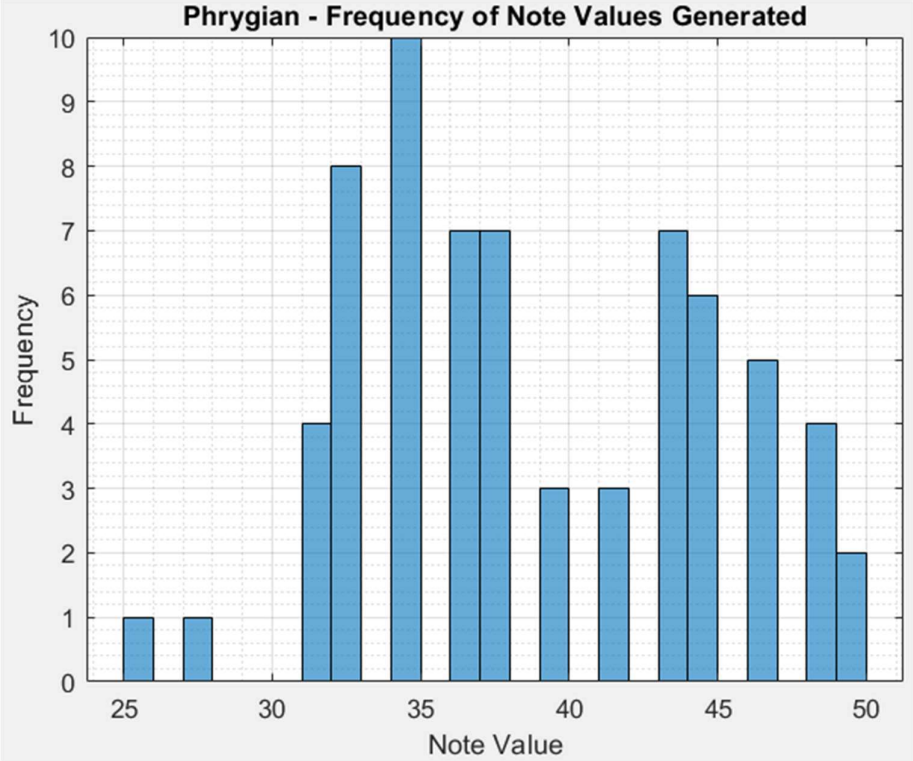

Figure 34 Distribution of Phrygian melodies

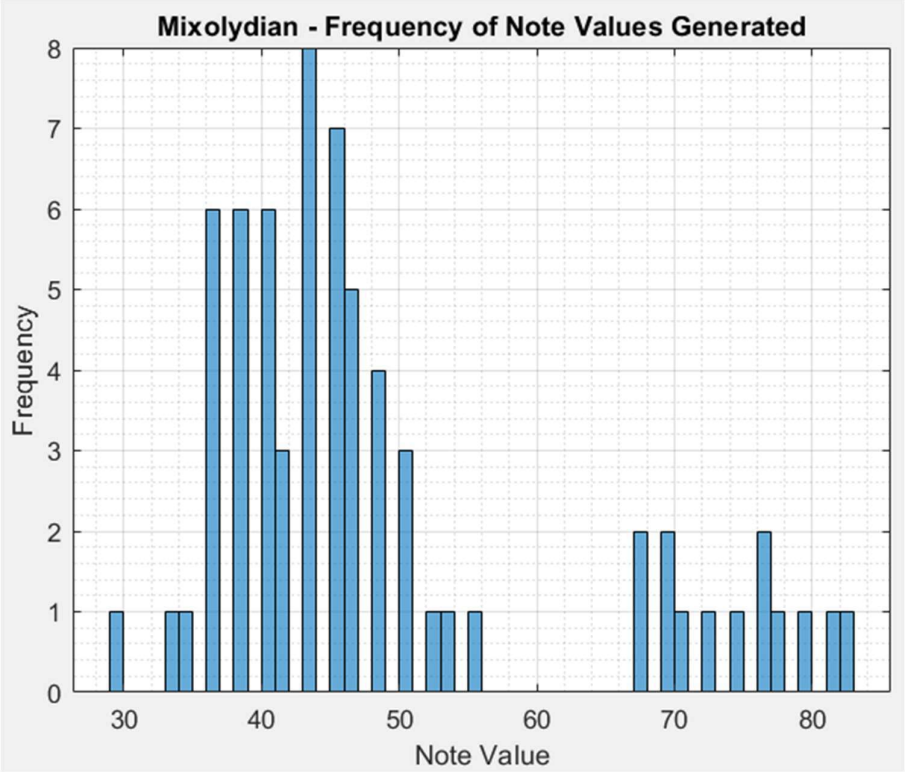

Figure 35 Distribution of Mixolydian melodies

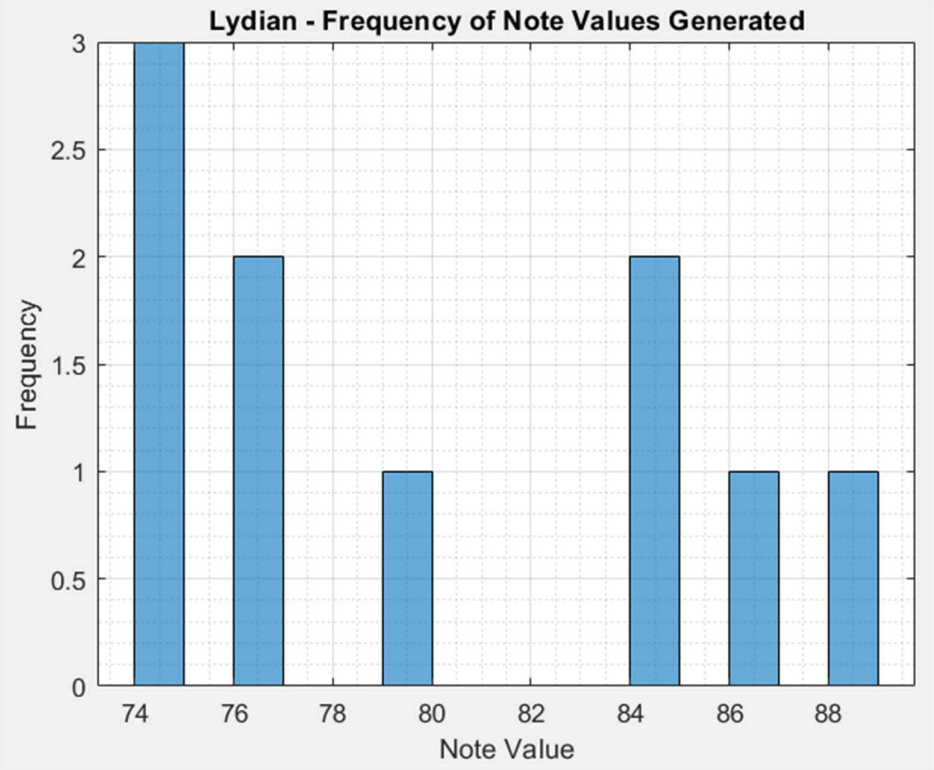

Figure 36 Distribution of Lydian melodies

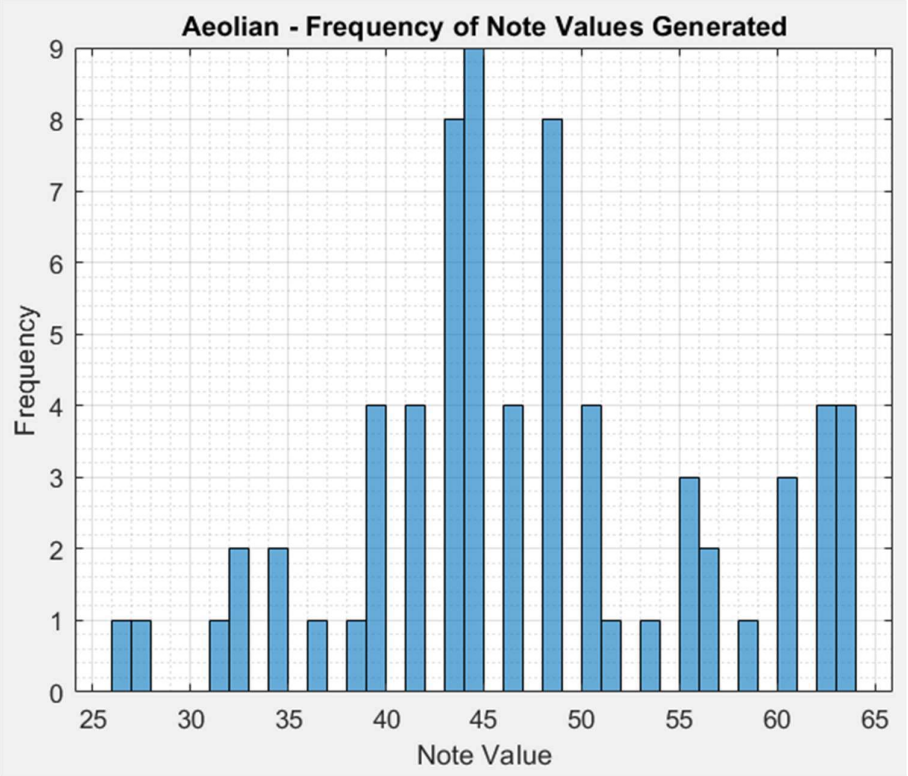

Figure 37 Distribution of Aeolian melodies

4.2.5. Melodic Contour

The melodic contour appears stochastic which does not mimic most music created by humans (Figure *38*). Note that this is plotted before the BPM adjustment is applied (since this is applied only in the Pure Data code) hence there is no visible change in tempo. There are occasional large intervals between consecutive notes of approximately an octave that likely occur when the stress modulation transitions causing the pitch register to update. This adds surprise which can increase the listeners attention but may also be jarring depending on the context. Judging the generated piece purely from the contour would suggest that the music would not take the listener on a journey and may become boring without a motif that is repeated occasionally and gradually altered. However, the issues outlined are known to occur with low order Markov models and this could be improved by using a more state of the art algorithm.

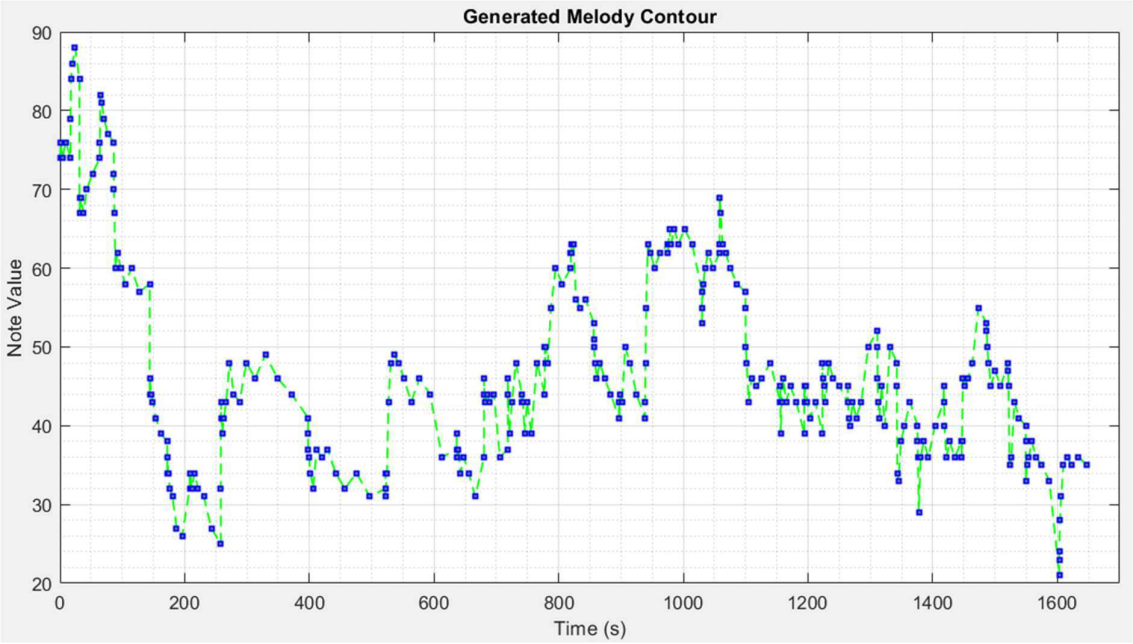

Figure 38 Plot of melody over time

5. Discussion

This body of work aimed to create a proof-of-concept stress detection and alleviation system using EDA and generative music. This section provides a comprehensive review of the system and the music created.

5.1. Stress Classification Comparison with Previous Work

Subject dependent and independent models were created to compare performance, understand how generalisable the patterns found in the feature set were, and determine how best to detect a users' stress level with minimal sensors. As expected, the subject dependent model achieved a considerably higher accuracy and F-Measure due to the variability of stress responses between subjects. This is a well-known phenomenon in affective computing and is the reason subject dependent models are more popular. It was shown that performance of feature domains varied between individuals suggesting where possible (and if sufficient data can be gathered), stress detection systems should involve feature selection on an individual basis, rather than attempting to generalise patterns found in populations.

The best performing classification model created in this study achieved an accuracy of 86% and an F-Measure of 0.94. This is an improvement to the benchmark Random Forest model trained on wrist EDA data provided by Schmidt et al. (date) which achieved an F_1-Score of 70.88% ± 0.20 and an accuracy of 76.29% ± 0.14. In fact, this new model approximately matches the prediction performance achieved when using all physiological signals recorded by the Empatica E4 (EDA, skin temperature, blood volume pulse). This enhanced performance is likely due to the extended feature set in combination with the feature extraction method. Forward selection successfully increased the classification performance whilst decreasing the feature set size from 25 to 7.

For subject dependent models the best performing feature domain was the first derivative, and the worst performing was the MFCCs. These results are in contradiction to those found by Shukla et al., who found that statistical features extracted from MFCCs outperformed those extracted from the time domain for both valence and arousal recognition (Shukla et al., 2019). Shukla et al. also found that 95-96 EDA features were required for optimal classification performance whereas in this study 7 features provided optimal performance. Similarly, Ayata et al. found that 14 features provided optimal classification performance, but the difference is likely due to differences in methodology (Ayata et al., 2017). This study

regarded binary stress vs nonstress classification whereas the two studies mentioned considered emotion recognition in general and aimed to classify emotion via arousal and valence. This is a more difficult classification task which would likely require more data than a stressed vs non-stressed task such as this.

Altering the number of trees from 100 to 500 made no significant difference to the system performance, however 100 trees was the minimum number tested. Fewer trees may provide similar classification accuracy with less processing. Data processing limitations meant that the smallest window size that could be adopted in this study was 60 seconds. Ayata et al. showed that decreasing the window size from 60 seconds to 3 seconds can increase the classification accuracy by 5% suggesting this could be a relatively simple way improve the performance of this model.

5.2. Increasing Stress Classification Resolution

A limitation of using the Random Forest algorithm with the WESAD dataset is that the classification was binary. A moving average was used to convert this value into a modulation signal, but this reduced the accuracy of the system and introduced a time lag. Non-binary stress classification models have been created such as that by Salazar-Ramirez et al. but this method relies on biosignals (or features of bio signals) that can be separated using membership functions, such as gaussian curves (Salazar-Ramirez et al., 2018). Therefore, this method is limited to defining probability distributions for each condition to form fuzzy boundaries. It may be possible to use the number of majority case decision trees to create a certainty metric that could be used to define a higher resolution of stress classification. Unfortunately, time limitations meant that this route was not explored. Alternatively, there is a requirement for experimental procedures that can provide a higher resolution of stress measurement.

5.3. Future Issues with Stress Classification

To create labelled affect datasets, all experimental procedures require self-report to act as a ground truth but there are well known inaccuracies associated with this. Hence one aim of affective computing is to create systems that allow researchers to bypass self-report for accurate and continuous psychology research. Creating a dataset with increased resolution would require participants to report their emotional state accurately and reliably, with high resolution. This has clear issues. Quantifying small differences in stress levels is difficult and

relies on accurate self-rating of emotional states, which raises unanswered questions such as what is the difference limen for an emotional state? If high resolution systems are created they must be validated rigorously.

The research area aims to create systems that can detect stress with equal or more certainty and resolution than self-report, but the creation and validation of such systems requires data of equal certainty and resolution. We do not yet have systems that we trust to provide stress classification to a higher level than the individuals themselves, yet the individual does not know their own current emotional state with complete accuracy. Frequently, people go through stressful periods of their lives without considering it at the time due to their appraisal of the situation. Therefore, the data provided to classification systems will always have inaccuracies that would be picked up in the model. A perfectly accurate model just means that it successfully maps the biosignals onto the labels provided. If the labels provided are not completely correct (due to issues associated with self report or with assumptions about the environment), a perfectly accurate model is not so when compared to ground truth. On the other hand, if a system were truly able to detect emotional state with perfect accuracy (not compared with measured data but with ground truth) the developer may treat it as though it is inaccurate and continue to develop until it matches the labels provided.

The system developed in this research is only trained to detect stress from two other states; amusement and a baseline. It may confuse the many other emotional states with the stress state since it hasn't observed the biosignals associated with such states during training. This is particularly likely for emotions that lie in a similar position in Russel's circumplex.

5.4. Melody Generation Appraisal

The change of mode is audible and successfully alters the feeling of the piece. However, training the model on C Ionian and applying mode quantisation after generation is not the optimal method since it applies the probability distribution from one mode to all others. In music theory the scale degree describes the position of a note in a scale compared to the tonic. Classical or jazz composers will accentuate the defining notes of a mode to emphasise the feeling each mode induces. There are two ways this could be included in this system; either separate Markov models could be trained on music of each specific mode, or the Markov model could be slightly adapted after training to emphasise intervals or notes that best define that mode.

The current system is limited to playing only two notes simultaneously, the bass drone and a

melody note. Previous work has shown that such droning successfully induces the emotional states best described by each mode, but it is a simplistic and stochastic model lacking music structure (Bostwick et al., 2018). Furthermore, chords were omitted completely from the system since it would have added considerable complexity. Chord types depend on previous chords as well as the melody, requiring a model that can take these interdependencies into account to create coherent music. If chords are to be included this would add variety to the music created which would likely increase flexibility and interest. No melodies in the training set were recreated meaning the system successfully composed new melodies but each lacked structure. Higher order Markov models could be used to improve structure but there is a risk of generating melodies that exist in the training set.

The Markov model created 2 bar melodies that did not consider the previous melodies generated meaning it did not have memory. Music is heavily time dependent so using an algorithm that can take this dependence into account (such as an LSTM) and generate one note at a time with the information provided (rather than two bar melodies) would improve results. However, more sophisticated models may require more processing power so it must be ensured that the generative algorithm is capable of creating melodies in the limited timeframe.

The use of MIDI meant the system is very versatile and can be used with any standalone soft synth, hence offers the user considerable flexibility. This allows genre specific sounds to be chosen allowing genre preferences to be considered. Acoustic features can also be considered and used to inform the modulation of synthesiser parameters.

5.5. Difficulties with Hard Coding the Iso-Principle

The iso-principle was applied via sets of rules for each musical parameter which successfully created music that adapted over time based on the stress data provided. However, musical parameters such as tempo and velocity cannot be perpetually increased or decreased. Once all parameters have been changed to their maximum or minimum values the current system becomes static. This could happen if the users stress level does not decrease even after the system has switched to Ionian.

There have only been a limited number of direct studies of the iso-principle however it has been applied in a variety of contexts with case studies supporting it's efficacy (Goldschmidt, 2020). It has been used to aid the treatment of severe mental disorders, but the choice of the music used is decided by experienced professionals who understand the patient and the effect

music can have. Although the iso-principle was considered the best framework with which to approach this task there are details in how it is applied automatically that need to be further developed. For example, in order to reflect the users' current mental state, music composed in minor keys were initially used but this may have the opposite effect to what is intended, making the stressful experience more intense. This is known as an affective feedback loop and the system could potentially emphasise a stressful state rather than alleviate it. The rules that define how the musical parameters change must be determined through detailed research of their effect on stress level.

Most the studies used to inform the musical parameter modulation in my research considered all emotions either via emotion categories or through the valence and arousal model. Therefore, the research is generally not specific to stress alleviation.

There is great difficulty defining rules that can generate continuously changing audio that alleviates stress, particularly when considering the variety of emotions that can be experienced. In this case it has been assumed the stress state is low valence and high arousal and all parameters have been modulated to alter each dimension. Rather than a stress vs non-stress detection system, a valence/arousal system may be more appropriate, allowing each parameter to be modulated as needed. However, since the relationship between musical parameters and emotions is complex and subjective, defining this relationship using rules is difficult.

5.6. System Architecture Evaluation

The connection between MATLAB, Pure Data and Analog Lab 5 worked sufficiently for prototyping but presented challenges. The use of a clock in Pure Data limited the notes to a strict grid which reduced the performance expression. Human performance rarely lands directly on the grid and so listeners expect that as part of the performance. Furthermore, there was difficulty synchronising MATLAB and Pure Data, and ensuring no notes were played when data was sent to Pure Data. The use of multiple languages meant the development became complex, this could be simplified by using a single programming language for all processing (such as MATLAB, or C++ if being implemented as an application) and communicating directly with a synthesiser. This was not done in this case since OSC and Pure Data was the only way considered to send MIDI data from MATLAB to a software

synthesiser.

5.7. Suitability for a Real-Time System

It was vital to use an algorithm capable of classifying windows of EDA data in the same amount of time 2 bars of music played. The Random Forest model successfully provided stress classification in this limited time frame. Each classification required 0.016 seconds including feature extraction, suggesting it would be capable of providing classifications every 5 seconds as would be the case if the current window shift were adopted in a real-time system.

The generative algorithm produced two bars of melody in 0.17 seconds meaning it could be implemented in a real-time system. The varying tempo means that the duration of each two bar melody also changes and so it cannot be synced with the classification window shift when present in the same loop as done in this case. Instead, the classifier and music generator must be separate but share information. Both should be defined as separate objects that can share data whilst maintaining separate calculation periods. This will ensure the classifier can use an optimal window shift size and the music generation can create coherent music that varies according to the BPM.

5.8. Ethical Considerations of Affective Computation and Algorithmic Composition

This thesis covers two controversial research topics; autonomous emotion recognition and computer-generated art. Care must be taken in developing both types of system due to their power and potential use. Some warn that governments currently cannot create legislation and regulations that keep up with the speed of technological innovation, suggesting ethics and best use is currently in the hands of the developers. It is therefore vital to consider not only the end goal of the tools built but also the implications they could have on society, good and bad. This is an ongoing area of debate that mustn't be shied away from. There are potentially extremely pernicious uses of these systems that arguably most people would not agree with, though that is similar with many powerful technologies that can also be used for good. If music generation algorithms become capable of creating music indistinguishable from that created by humans, could automation take hold of the music industry? If released as open software, will streaming platforms become saturated with music created with no human input? How will this affect the economy of the music industry? Is this a future that developers want to see? Potentially more concerning is social control through knowledge of emotional

state. Gross misconduct has already occurred multiple times at large companies concerning the handling of individuals data. It is now standard to sell personal data to advertising companies to create more precise targeting algorithms, which has already begun to affect voting tendencies and alter geopolitics. Adding emotional data to such systems creates potential for mass manipulation. Facial recognition systems are already implemented in some societies to enforce laws and automatically fine individuals. Emotion can also be recognised in real-time using videos of individuals faces with high accuracies reported (though it's worth noting the true accuracy of such systems is debated) which would allow emotion detection without formal consent. There are ongoing ethical debates around the implications for people whose data has been used in training such systems.

6. Conclusion

A proof-of-concept stress detection and alleviation system has been created utilising a Random Forest model for classifying windows of EDA data into stressed vs non stressed conditions. A first order Markov model was then used in combination with multiple rules that manipulated musical parameters based on the current stress classification to apply the iso-principle.

It was shown that using random sampling created a severely overtrained model due to the overlap of windows. This was because common data was present in both the training and testing sets. After revisiting the sampling methodology, a new model was created using all features extracted, which achieved an accuracy of 82% and an F-Measure of 0.89. Forward selection was then applied providing optimal performance using 7 features and an accuracy of 86% and an F-Measure of 0.94. The best 7 features found were: First Derivative Peak Frequency, MFCC Skew, First Derivative Mean, First Derivative Peak RMS, Skew, MFCC Mean, IMF Kurtosis, IMF Skew, and IMF Max. This is an improvement on the benchmark Random Forest model trained on the same data provided by Schmidt et al. and matches the performance achieved in that same study when using all physiological data measured by the Empatica E4.

Three values for tree number per forest were tested (100, 250 and 500) and this had a negligible effect on the model performance. However, less trees may be able to match this performance requiring less processing (this was not tested). A subject dependent model was created using each participant as a hold out set and training on the remaining 14. Perfect training accuracy was achieved for all cases but overall, an average testing accuracy of 74% with a minimum accuracy 40% and maximum of 92% was achieved which supports the notion that subject dependent models are generally more accurate. It was found that some feature domains performed best for different individuals suggesting that a sophisticated system should use feature selection on an individual basis, this could then be improved via in-situ continuous learning. The subject dependent model created ought to be validated by testing using labelled EDA data from a new dataset.

The output from the stress classification system was used as a modulation system to inform the alteration of musical parameters such as tempo, loudness, and mode. For each musical parameter, rules were defined to apply the iso principle and guide the user towards a calmer state. These rules were based on current literature. The output of the first order Markov model was a monophonic melody that was played over a C bass drone to set the tonal centre

and induce a mode in C. It was noted that a classification algorithm that offers higher resolution than binary would be better suited to this problem allowing the parameters to be changed gradually and synchronised with the users' mental state.

This approach is limited since the musical parameters cannot be perpetually decreased. More sophisticated rules need to be defined so that the system does not become static if the user stays in a state for a prolonged period after all parameters have been changed to their maximum or minimum levels. Additionally, applying the iso-principle in this context requires the beginning of the piece to reflect the users state which could emphasise their stress but this is necessary for the iso-principle to be applied. It is difficult to define rules due to the complex mapping between musical parameters and emotion. Some rules had unintended effects such as with pitch register which created occasional large intervals between consecutive notes. The classification algorithm took 0.016 seconds to process (including feature extraction) and the generative algorithm took 0.17 seconds to complete. This suggests both algorithms are ready to be implemented in real-time.

7. Further Work

A smartphone application could be created that uses physiological data from wrist worn devices to generate stress alleviating music. This can act as a cheap, unintrusive addition to other forms of therapy such as meditation and used to reduce preoperative stress for example. More sophisticated algorithms can be used for both stress classification and music generation though they must be capable of producing output quickly. Subjective tests could then be carried out to more clearly understand how rules can be defined to apply the iso-principle and prevent affective feedback.

There was no objective consideration for the synthesiser sound being used even though this has a very large impact on the response. Sound quality metrics can be used as an intermediate to find the correlation between synthesiser parameters and emotional state. These can then be modulated via the stress signal creating a comprehensive music therapy tool. Note that this project focuses heavily on ideas from western music theory, efficacy will likely rely on cultural learning.

A framework for collaboration with musicians could also be created, allowing composers to use physiological and emotional data in their music. This would allow for music that is never the same and adapts according to each listening experience. This could be achieved by creating premixed loops that are automatically mixed by an emotionally informed algorithm based on labels supplied by the musician. Although drones have been proven to affect felt emotion, few genres share this nature. For example, by:

1. MIDI - the user could be prompted to select their top genres or artists producing a library of audio. Isolation techniques could then be applied in combination with music information retrieval to suggest virtual instruments (both which instruments and presents to use). The generation section could then be trained using libraries of MIDI selected for each genre.
2. Sample by sample audio - with increasing processing power, using audio as the form of music representation may become more feasible. This would allow new music to be created from user chosen preferences and the rules outlined in this work (in combination with others that were not applied) could be used to target calmer mental states.

In conclusion, this work helps set the foundation for two exciting future projects. The first being an application capable of generating stress alleviation music in real-time based on the user's current stress level that would only require a single wrist worn EDA sensor. Such an

application could be used for early intervention stress relief such as before an operation. The second is an entirely new form of music composition and listening. Rather than saving songs as a single file, they could be saved as building blocks of a song. These building blocks could then be arranged by an algorithm in real time based on the user's stress (or other emotional state) as defined by the composer. This would mean that no two listens would be the same and would combine human and algorithmic composition in a single system.

8. References

Akmandor, A. O., & Jha, N. K. (2017). *Keep the Stress Away with SoDA : Stress Detection and Alleviation System*. *3*(4), 269–282.

Ayata, D., Yaslan, Y., & Kamasak, M. (2017). Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods. *Istanbul University - Journal of Electrical and Electronics Engineering*, *17*(1), 3129–3136.

Bach, D. R. (2014). A head-to-head comparison of SCRalyze and Ledalab, two model-based methods for skin conductance analysis. *Biological Psychology*, *103*(1), 63–68. https://doi.org/10.1016/j.biopsycho.2014.08.006

Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *Journal of Neuroscience Methods*, *184*(2), 224–234. https://doi.org/10.1016/j.jneumeth.2009.08.005

Benedek, M., & Kaernbach, C. (2010a). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Benedek, M., & Kaernbach, C. (2010b). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, *47*(4), 647–658. https://doi.org/10.1111/j.1469-8986.2009.00972.x

Bostwick, J., Seror, G. A., & Neill, W. T. (2018). Tonality without structure: Using drones to induce modes and convey moods. *Music Perception*, *36*(2), 243–249. https://doi.org/10.1525/MP.2018.36.2.243

Bota, P. J., Wang, C., Fred, A. L. N., & Placido Da Silva, H. (2019). A Review, Current Challenges, and Future Possibilities on Emotion Recognition Using Machine Learning and Physiological Signals. *IEEE Access*, *7*, 140990–141020. https://doi.org/10.1109/ACCESS.2019.2944001

Bradley, M., & Lang, P. J. (1994). Self-Assessment Manikin (SAM). *J.Behav.Ther. & Exp. Psychiat.*, *25*(1), 49–59.

Braithwaite, J. J., Derrick, D., Watson, G., Jones, R., Rowe, M., Watson, D., Robert, J., & Mickey, R. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. …, 1–42. http://www.bhamlive.bham.ac.uk/Documents/college-les/psych/saal/guide-electrodermal-activity.pdf%5Cnhttp://www.birmingham.ac.uk/documents/college-

les/psych/saal/guide-electrodermal-activity.pdf%0Ahttps://www.birmingham.ac.uk/Documents/college-les/psych/sa

Breiman, L. (2001). *Random Forest*.

Breiman, L., & Cutler, A. (2001). *Random Forests*. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Briot, J. P., & Pachet, F. (2020). Deep learning for music generation: challenges and directions. *Neural Computing and Applications*, *32*(4), 981–993. https://doi.org/10.1007/s00521-018-3813-6

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). *A practical guide to sentiment analysis*.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Chamorro-Premuzic, T., & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life? *British Journal of Psychology*, *98*(2), 175–185. https://doi.org/10.1348/000712606X111177

Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health*, *8*(1). https://doi.org/10.1136/fmch-2019-000262

Clark, M., Isaacks-Downton, G., Wells, N., Redlin-Frazier, S., Eck, C., Hepworth, J. T., & Chakravarthy, B. (2006). Use of preferred music to reduce emotional distress and symptom activity during radiation therapy. *Journal of Music Therapy*, *43*(3), 247–265. https://doi.org/10.1093/jmt/43.3.247

Eerola, T., & Toiviainen, P. (2004). MIDI toolbox: Matlab tools for music research. In *University of Jyväskylä: Kopijyvä, Jyväskylä, Finland. Retrieved from www. jyu. fi/musica/miditoolbox/in May*. http://scholar.google.fi/scholar?q=Eerola%2C+T&hl=fi&btnG=Haku#0

Eerola, Tuomas, & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Eerola, Tuomas, & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception*, *30*(3), 307–340.

https://doi.org/10.1525/MP.2012.30.3.307

Egilmez, B., Poyraz, E., Zhou, W., Memik, G., Dinda, P., & Alshurafa, N. (2017). UStress: Understanding college student subjective stress using wrist-based passive sensing. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2017*, 673–678. https://doi.org/10.1109/PERCOMW.2017.7917644

Ekman, P. (1972). Universals and Cultural Differences in Facial Expressions of Emotion. In *Nebraska Symposium on Motivation* (Vol. 19, pp. 207–282). papers3://publication/uuid/FDC5E29A-0E28-4DDF-B1A4-F53FEE0B4F70

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic Nervous System Activity Distinguishes Among Emotions. In *Science* (Vol. 221, pp. 1208–1210).

Ekman, P., Richard Sorenson, E., & Friesen, W. V. (1968). *Pan-Cultural Elements in Facial Displays of Emotion*. *164*(1967).

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*, 3133–3181.

Gaurav, K. A., & Patel, L. (2020). *Machine Learning With R*. https://doi.org/10.4018/978-1-7998-2718-4.ch015

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, 1–22. https://doi.org/10.1109/TAFFC.2019.2927337

Goldschmidt, D. (2020). *Investigating The Iso Principle: The Effect Of Music Tempo Manipulation On Arousal Shift*. https://minerva-access.unimelb.edu.au/handle/11343/56627%0Ahttp://www.academia.edu/download/39541120/performance_culture.doc

Google. (2021). *Overview of Debugging ML Models*. https://developers.google.com/machine-learning/testing-debugging/common/overview

Gressling, T. (2020). Automated machine learning. In *Data Science in Chemistry*. https://doi.org/10.1515/9783110629453-084

Hatta, T., & Nakamura, M. (1991). Can antistress music tapes reduce mental stress? *Stress Medicine*, *7*(3), 181–184. https://doi.org/10.1002/smi.2460070309

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2),

156–166. https://doi.org/10.1109/TITS.2005.848368

Healing Soul. (2022). *Beautiful Relaxing Music - Stop Overthinking, Stress Relief Music, Sleep Music, Calming Music*. https://www.youtube.com/watch?v=_kT38XB1YHo&ab_channel=HealingSoul

Heidersheit, A., & Madson, A. (2015). *Use of the Iso Principle as a Central Method in Mood Management: A Music Psychotherapy Clinical Case Study*.

Hevner, K. (1935). *The Affective Character of the Major and Minor Modes in Music*. *47*(1), 103–118.

Hsieh, C. P., Chen, Y. T., Beh, W. K., & Wu, A. Y. A. (2019). Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection. *IEEE Workshop on Signal Processing Systems, SiPS: Design and Implementation*, *2019-Octob*, 330–335. https://doi.org/10.1109/SiPS47522.2019.9020321

Jiang, J., Rickson, D., & Jiang, C. (2016). The mechanism of music for reducing psychological stress: Music preference as a mediator. *Arts in Psychotherapy*, *48*, 62–68. https://doi.org/10.1016/j.aip.2016.02.002

Jin, C., Tie, Y., Bai, Y., Lv, X., & Liu, S. (2020). A Style-Specific Music Composition Neural Network. *Neural Processing Letters*. https://doi.org/10.1007/s11063-020-10241-8

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jung, C. G. (1919). *Studies in word-association; experiments in the diagnosis of psychopathological conditions carried out at the Psychiatric clinic of the University of Zurich*. https://archive.org/details/studiesinwordass00jung/page/480/mode/2up?q=galvano

Kalia, M. (2002). Assessing the economic impact of stress - The modern day hidden epidemic. *Metabolism: Clinical and Experimental*, *51*(6 SUPPL. 1), 49–53. https://doi.org/10.1053/meta.2002.33193

Kalingeri, V., & Grandhe, S. (2016). *Music Generation with Deep Learning*. http://arxiv.org/abs/1612.04928

Khalfa, S., Bella, S. ., Roy, M., Peretz, I., & Lupien, S. . (2003). Effects of Relaxing Music on Salivary Cortisol Level After Psychological Stress. *Ann. N.Y. Acad. Sci*, *021*, 67–69.

Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of motivation definitions, with a suggestion for a consensual definition. *Motivation and Emotion*, *5*(3), 263–291.

https://doi.org/10.1007/BF00993889

Koolhaas, J. M., Bartolomucci, A., Buwalda, B., de Boer, S. F., Flügge, G., Korte, S. M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wöhr, M., & Fuchs, E. (2011). Stress revisited: A critical evaluation of the stress concept. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1291–1301. https://doi.org/10.1016/j.neubiorev.2011.02.003

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. In *Applied Predictive Modeling*. https://doi.org/10.1007/978-1-4614-6849-3

Kurniawan, H., Maslov, A. V., & Pechenizkiy, M. (2013). Stress detection from speech and Galvanic Skin Response signals. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 209–214.

Kwekkeboom, K. L. (2003). Music versus distraction for procedural pain and anxiety in patients with cancer. *Oncology Nursing Forum*, *30*(3), 433–440. https://doi.org/10.1188/03.ONF.433-440

Liu, Y., & Du, S. (2018). Psychological stress level detection based on electrodermal activity. *Behavioural Brain Research*, *341*(November 2017), 50–53. https://doi.org/10.1016/j.bbr.2017.12.021

Maheshwari, S., & Kumar, A. (2014). Empirical Mode Decomposition: Theory & Applications. *International Journal of Electronic and Electrical Engineering*, *7*(8), 873–878. http://www.irphouse.com

Mao, H. H. (2018). DeepJ: Style-Specific Music Generation. *Proceedings - 12th IEEE International Conference on Semantic Computing, ICSC 2018*, *2018-Janua*, 377–382. https://doi.org/10.1109/ICSC.2018.00077

Maurer, J. (1999). *The History of Algorithmic Composition*. https://ccrma.stanford.edu/~blackrse/algorithm.html

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468). https://doi.org/10.1126/science.aax0868

Midya, V., Valla, J., Balasubramanian, H., Mathur, A., & Singh, N. C. (2019). Cultural differences in the use of acoustic cues for musical emotion experience. *PLoS ONE*, *14*(9), 1–17. https://doi.org/10.1371/journal.pone.0222380

Minguillon, J., Perez, E., Lopez-Gordo, M. A., Pelayo, F., & Sanchez-Carrion, M. J. (2018).

Portable system for real-time detection of stress level. *Sensors (Switzerland)*, *18*(8), 1–15. https://doi.org/10.3390/s18082504

Mishra, P. (2019). Music Tune Generation based on Facial Emotion. *International Journal of Engineering Research and Technology (IJERT)*, *8*(11), 501–504. https://www.ijert.org

Motoda, H., & Liu, H. (2002). Feature selection, extraction and construction. *Communication of IICM*, *5*, 67–72.

Nath, R. K., Thapliyal, H., & Caban-Holt, A. (2020). Validating physiological stress detection model using cortisol as stress bio marker. *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, *2020-Janua*, 2–6. https://doi.org/10.1109/ICCE46568.2020.9042972

Neerincx, M. A., & Kraaij, W. (2014). The SWELL Knowledge Work Dataset for Stress and User Modeling Research Categories and Subject Descriptors. *Proceedings of the 16th International Conference on Multimodal InteractionNovember 2014 Pages 291–*, 291–298.

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, *5*(1), 8–15. https://doi.org/10.1177/1754073912457227

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). *WaveNet: A Generative Model for Raw Audio*. 1–15. http://arxiv.org/abs/1609.03499

Picard, R. W. (1999). Affective Computing for HCI. *Proceedings of the 8th HCI International on Human-Computer Interaction: Ergonomics and User Interfaces*, 829–833. http://dl.acm.org/citation.cfm?id=647943.742338

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(10), 1175–1191. https://doi.org/10.1109/34.954607

Ramos, D., Bueno, J. L. O., & Bigand, E. (2011). Manipulating Greek musical modes and tempo affects perceived musical emotion in musicians and nonmusicians. *Brazilian Journal of Medical and Biological Research*, *44*(2), 165–172. https://doi.org/10.1590/S0100-879X2010007500148

Richardson, M. M., Babiak-Vazquez, A. E., & Frenkel, M. A. (2008). Music therapy in a comprehensive cancer center. *Journal of the Society for Integrative Oncology*, *6*(2), 76–81. https://doi.org/10.2310/7200.2008.0006

Russel, J. (1980). *A Circumplex Model of Affect*.

Sagha, H., Coutinho, E., & Schuller, B. (2015). Exploring the importance of individual

differences to the automatic estimation of emotions induced by music. *AVEC 2015 - Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge, Co-Located with MM 2015*, 57–63. https://doi.org/10.1145/2808196.2811643

Salazar-Ramirez, A., Irigoyen, E., Martinez, R., & Zalabarria, U. (2018). An enhanced fuzzy algorithm based on advanced signal processing for identification of stress. *Neurocomputing*, *271*, 48–57. https://doi.org/10.1016/j.neucom.2016.08.153

Schmidt, P., Reiss, A., Duerichen, R., & van Laerhoven, K. (2018). Wearable affect and stress recognition: A review. *ArXiv*.

Schmidt, P., Reiss, A., Duerichen, R., & Van Laerhoven, K. (2018). Introducing WeSAD, a multimodal dataset for wearable stress and affect detection. *ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction*, 400–408. https://doi.org/10.1145/3242969.3242985

Schulze, W., & Merwe, B. van der. (2011). *Music Generation with Markov Models*. 78–85.

Setz, C., Arnrich, B., Schumm, J., Marca, R. La, Tr, G., & Ehlert, U. (2010). Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *Technology*, *14*(2), 410–417.

Shi, Y., Nguyen, M. H., Blitz, P., French, B., Fisk, S., Torre, F. D. La, Smailagic, A., & Siewiorek, D. P. (2010). Personalized Stress Detection from Physiological Measurements. *Second International Symposium on Quality of Life Technology*. http://www.shrs.pitt.edu/qolt/qolt.aspx?id=2212

Shukla, J., Barreda-Angeles, M., Oliver, J., Nandi, G. C., & Puig, D. (2019). Feature Extraction and Selection for Emotion Recognition from Electrodermal Activity. *IEEE Transactions on Affective Computing*, *3045*(c). https://doi.org/10.1109/TAFFC.2019.2901673

Sloboda, J., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In *Music and Emotion Theory and Research* (pp. 71–104). http://psycnet.apa.org/psycinfo/2001-05534-001

Smith, J. C., Bradley, M. M., Scott, R. P., & Lang, P. J. (2004). The Psychophysiology of Emotion. *Medicine & Science in Sports & Exercise*, *36*(Supplement), S91. https://doi.org/10.1249/00005768-200405001-00432

Temperley, D., & Tan, D. (1973). Emotional Connotations of Diatonic Modes. *Music Educators Journal*, *60*(1), 101–101. https://doi.org/10.2307/3394408

Tsai, H. F., Chen, Y. R., Chung, M. H., Liao, Y. M., Chi, M. J., Chang, C. C., & Chou, K. R. (2014). Effectiveness of music intervention in ameliorating cancer patients' anxiety,

depression, pain, and fatigue: A meta-analysis. *Cancer Nursing*, *37*(6), E35–E50. https://doi.org/10.1097/NCC.0000000000000116

van der Zwaag, M. D., Westerink, J. H. D. M., & van den Broek, E. L. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae*, *15*(2), 250–269. https://doi.org/10.1177/1029864911403364

Wallis, I., Ingalls, T., Campana, E., & Goodman, J. (2011). A rule-based generative music system controlled by desired valence and arousal. *Proceedings of the 8th Sound and Music Computing Conference, SMC 2011*.

Wierzbicka, A. (1986). *Human Emotions : Universal or Culture-Specific? 88*(3), 584–594.

Wiriyachaiporn, P., Chanasit, K., Suchato, A., Punyabukkana, P., & Chuangsuwanich, E. (2018). Algorithmic Music Composition Comparison. *Proceeding of 2018 15th International Joint Conference on Computer Science and Software Engineering, JCSSE 2018*. https://doi.org/10.1109/JCSSE.2018.8457397

Wu, J., Hu, C., Wang, Y., Hu, X., & Zhu, J. (2017). A hierarchical recurrent neural network for symbolic melody generation. *ArXiv*, *50*(6), 2749–2757.

Yehuda, N. (2011). Music and Stress. *Journal of Adult Development*, *18*(2), 85–94. https://doi.org/10.1007/s10804-010-9117-4

Yellow Brick Cinema - Relaxing Music. (2023). *Relaxing Music 24/7, Stress Relief Music, Sleep Music, Meditation Music, Study, Flowing River*. https://www.youtube.com/watch?v=xp07Z_3XY3E&ab_channel=YellowBrickCinema-RelaxingMusic

Yu, Y., & Canales, S. (2019). Conditional LSTM-GAN for Melody Generation from Lyrics. *ArXiv*.

Zangróniz, R., Martínez-Rodrigo, A., Pastor, J. M., López, M. T., & Fernández-Caballero, A. (2017). Electrodermal activity sensor for classification of calm/distress condition. *Sensors (Switzerland)*, *17*(10), 1–15. https://doi.org/10.3390/s17102324

Appendices

Songs used in Training Set

“Three Little Pigs”

“Alladin Theme”

“An American Tail”

“Animaniacs”

“Bare Necessities”

“Big (Heart & Soul)”

“Bingo”

“Brady Bunch Theme”

“Ducktale”

“Gilligan’s Island”

“Happy & You Know It”

“Happy Birthday”

“Hush Little Baby”

“London Bridge”

“Looney Tunes”

“Muppet’s Theme”

“Old Mcdonald”

“Pooh”

“Puff the Magic Dragon”

“Ren & Stimpy Happy Happy Joy Joy”

“Scooby Doo”

“Sesame Street”

“Spoonful of Sugar”

“Supercalifragilisticexpialidocious”

“This Old Man”

“Tiggers Song”

“Woody Woodpecker”

“Zipadee Do Da”