sciendo

JAISCR, 2021, Vol. 11, No. 1, pp. 5 – 20
10.2478/jaiscr-2021-0001

TOWARDS EFFECTIVE MUSIC THERAPY FOR MENTAL HEALTH CARE USING MACHINE LEARNING TOOLS: HUMAN AFFECTIVE REASONING AND MUSIC GENRES

Jessica Sharmin Rahman*, Tom Gedeon, Sabrina Caldwell, Richard Jones, Zi Jin

Research School of Computer Science, The Australian National University
Canberra, Australia

E-mail: jessica.rahman@anu.edu.au

Submitted: 4th December 2019; Accepted: 14th September 2020

Abstract

Music has the ability to evoke different emotions in people, which is reflected in their physiological signals. Advances in affective computing have introduced computational methods to analyse these signals and understand the relationship between music and emotion in greater detail. We analyse Electrodermal Activity (EDA), Blood Volume Pulse (BVP), Skin Temperature (ST) and Pupil Dilation (PD) collected from 24 participants while they listen to 12 pieces from 3 different genres of music. A set of 34 features were extracted from each signal and 6 different feature selection methods were applied to identify useful features. Empirical analysis shows that a neural network (NN) with a set of features extracted from the physiological signals can achieve 99.2% accuracy in differentiating among the 3 music genres. The model also reaches 98.5% accuracy in classification based on participants' subjective rating of emotion. The paper also identifies some useful features to improve accuracy of the classification models. Furthermore, we introduce a new technique called 'Gingerbread Animation' to visualise the physiological signals we record as a video, and to make these signals more comprehensible to the human eye, and also appropriate for computer vision techniques such as Convolutional Neural Networks (CNNs). Our results overall provide a strong motivation to investigate the relationship between physiological signals and music, which can lead to improvements in music therapy for mental health care and musicogenic epilepsy reduction (our long term goal).

Keywords: psychological signals, music genres' classification, music therapy

1 Introduction

Music is considered by many to be a universal language, that can elicit emotion from people all over the world. It is an art form that not only gives us pleasure but also works as a medicine for both mind and body. Listening to music has been shown to decrease heart and respiratory rate, as well as the level of stress hormones such as cortisol [1], thus it has been used to reduce stress and anxiety for many years. Music has also been shown to maintain the control of attention and strengthen focus in tasks [2]. Therefore, when it comes to analysing human mind and emotions, music has been a popular choice of stimuli among researchers.

Although music has been used quite frequently in therapy and biofeedback training, surprisingly little is known as to how it affects humans. Music therapy has been proved to reduce anxiety, improve sleep quality, reduce epileptic seizures etc. Harmat et al. [3] conducted an experiment on

94 students and showed that classical music improved their sleep quality, statistically. In Coppola et al., eleven patients with drug-resistant epileptic encephalopathy listened to a set of Mozart's compositions 2 hours per day for 15 days, and the results demonstrated that their frequency of seizures were reduced by half [4]. Classical music has frequently been shown to have positive effects in improving mental health and daily behavior. Furthermore, it is well known that increasing gamma waves in the brain can be beneficial as these waves are known to improve focus, cognition and memory formation. This is why many music therapy sessions use music or video stimuli to increase gamma waves in the brain to enhance cognitive ability. However, other types of music have not been compared extensively to verify whether they have different effects in comparison to classical music. In addition, these studies did not verify if they have any correlation with the participants' physiological signals. The wide range of applications of music in improving mental health are fascinating, and it is certainly worthwhile to explore how human physiological signals change pattern in response to music stimuli.

Physiological signals are strong measures found in human beings that demonstrate sensitivity to emotional changes. Identifying different physiological signal patterns caused by different types of music can help understanding which music should be used in responding to or even treating the above mentioned disorders. Physiological signals have been used by a number of researchers in the last few years to identify different emotions in humans. These signals were used with video stimuli to classify smiles [5] and anger [6] and reached an accuracy of 96.5% and 95% respectively. Physiological signals were also able to identify stress [7]. Combining a biofeedback training approach with physiological signals, a correlation between the signals and reduction in epileptic seizures was observed in an experiment by Nagai et.al [8]. These approaches provide a strong motivation to analyse a range of physiological signals and identify if they have a relation to different audio (music) stimuli.

Terms such as 'chills', 'thrills' and 'frissons' are often used by psychology researchers to describe the psychophysiological moments of musical experience [9]. A 'frisson' is 'a sudden strong feeling of excitement' and 'micro-frisson' is a sudden small feeling, which is too small to detect consciously, but is reflected by a person's physiological signals [16]. In particular, chills and micro-frissons are closely related and they reflect the emotional intensity induced by music [10]. These sensations are said to be highly reflected in physiological measures such as skin conductance response [11, 12]. Thus, physiological signals measured on the skin can be very useful in analysing the emotional effects of music.

In this paper, we explore the effects of four different physiological signals, Electrodermal Activity (EDA), Blood Volume Pulse (BVP), Skin Temperature (ST) and Pupil Dilation (PD) from subjects listening to music. All of these signals showed significant changes (reflecting the listener's reaction) to different stimuli [13, 14, 15]. Our previous study [16] had only explored the effects of EDA signals in differentiating different types of music based on genre and participants' subjective ratings. This study extends that work and investigates the effects of all four different signals using a broader set of features and also the combination of all of those signals. The paper is organized as follows: Following the introduction in Section 1, we discuss the materials and methods for the experiment in Section 2. Next in Section 3, we introduce our novel Gingerbread Animation technique and a graph based visualisation method. Then in Section 4, we display the results and discuss those in detail. Finally, we conclude the paper by highlighting some limitations and mention possible future work.

2 Preliminaries

2.1 Stimuli

A total of 12 music pieces were chosen as stimuli for this experiment. All the pieces were around 4 minutes in length. The music pieces were divided into three categories: classical, instrumental and pop. The complete list of the music pieces used in this experiment is given in Table 1.

As music pieces with long lasting periodicity (phrases spanning several bars of music) have been used for music therapy due to their beneficiary effects [17], we chose classical music pieces having this feature. Although music therapy research mostly includes the use of classical music pieces, it

is not ideal to only use one type of music because frissons may also occur while listening to other types of music. Limiting the stimuli to only one type also limits the ecological validity of the results [9]. Hence, we also chose some pieces from instrumental and popular music genres. When choosing instrumental music, we chose binaural beats which are shown to effectively synchronize brainwaves to enhance a specific brainwave pattern [19]. We selected two different types of binaural beats: a Gamma wave boosting piece to regain focus and awareness [20], and an alpha wave boosting piece for relaxation [21]. We also chose a jazz genre piece and a rock genre piece. Both of these pieces were used to analyse the effects of alpha and beta waves by Hurless et.al [22]. For the pop music category, we chose the No. 1 song of Billboard Hot 100 year-end charts from years 2014-2017 [23].

Table 1. Music Stimuli Used in the Experiment

Song Name	**Genre**
Mozart Sonatas K.448 [4]	Classical
Mozart Sonatas K.545 [18]	Classical
F. Chopin's "Funeral March" from Sonata in B flat minor Op. 35/2	Classical
J.S Bach's Suite for Orchestra No. 3 in D "Air" [17].	Classical
Gamma Brain Energizer	Instrumental
Serotonin Release Music with Alpha Waves	Instrumental
The Feeling of Jazz by Duke Ellington	Instrumental
YYZ by Rush	Instrumental
Happy by Pharrell Williams	Pop
Uptown Funk by Mark Ronson featuring Bruno Mars	Pop
Love Yourself by Justin Bieber	Pop
Shape of You by Ed Sheeran	Pop

2.2 Participants

Thirteen male and eleven female students (24 in total) participated voluntarily in this experiment.The mean age was 21 years old ($\pm$4.6). Among the participants 19 were undergraduate while 5 of them were postgraduate students. Some of the students had experience in playing different instruments, but none of them are professional musicians or music students. All of the participants signed a written consent form before participating in the experiment. The study was approved by the Australian National University's Human Research Ethics Committee.

2.3 Physiological Measures

2.3.1 Electrodermal Activity

Electrodermal activity (EDA) or Skin Conductance (SC) is a useful physiological signal which is known to be sensitive to emotional changes [24]. The EDA response fluctuates slowly but significantly, reflecting the current emotional state, and have been shown to have a strong correlation with cognitive load [25, 26]. The flow of electricity along the skin increases during stressful tasks, while it decreases during a relaxed state. Due to the reliability of data (less prone to noise) and easy analysis method, EDA has become one of the most used physiological signals to detect various affective states. The signal can be measured by placing electrodes on the surface of the skin. They are generally placed on the hands, some devices are placed on the wrist while others require electrodes to be placed on the fingers. Skin conductance signals can be divided into two categories based on their frequency. One is referred to as Skin Conductance Response (SCR) which shows the rapidly changing peaks in the signal. The other is called Skin Conductance Level (SCL) which is the slowly changing levels of the signal. Generally for affective computing, SCR signals are analysed.

2.3.2 Blood Volume Pulse

Blood Volume Pulse (BVP) refers to measurement of the volume of blood that is flowing through the tissues of a particular part of the body. It is usually measured on every pulse. BVP has been shown to have a correlation to emotional state change. For instance, higher stress is said to be reflected by low BVP level and vice versa [27]. Therefore this signal is often used in biofeedback training for reducing stress and anxiety. The sensors are also less complicated than for other signals, thus it is a popular choice for biofeedback based therapy. BVP is generally obtained by a photoplethysmography (PPG) sensor that detects the amount of light reflected from an infrared light source positioned on

the skin. This gives the amount of blood present in that certain area at a certain time. Some devices that record EDA can also record BVP.

2.3.3 Skin Temperature

Skin temperature (ST) is another commonly used physiological measure. Although it is a relatively sluggish indicator, it is still able to show correlation to different emotional states. ST tends to increase during the relaxed state while it decreases during increased stress or anxiety [28]. ST is measured normally on the surface on the skin, using the same sensor delivery platform as devices that measure skin conductance.

2.3.4 Pupil Dilation

Human eyes provide valuable information on their emotions and current mental state. There are various features that can be derived from the eye such as pupil dilation, blinking rates, eye gaze point, fixation point and saccade and fixation times etc. Among these, pupil dilation, which is the measurement of pupil size over time, is considered a very effective feature in emotion recognition [29]. Pupil diameter changes are said to reflect changes in brain state [30]. Typically in lab based experiments an eye tracking device is aligned with a computer screen so it can track where a person is looking on the screen. Pupil dilation has been used as an indication of emotional arousal such as stress [31]. Changes in pupil dilation have been seen while listening to familiar music and vocals [32]. Therefore it can be a useful signal to identify the effects of different music.

2.4 Experimental Design

All participants were recruited through a volunteer research participation website of the University. After arrival at our lab, they were briefed about the experiment procedure and handed a participation information sheet with detailed instructions. After they understood the procedure and agreed to participate in the experiment by signing a written consent form, they were asked to sit comfortably in a chair in front of a 17.1 inch monitor. The participants were fitted with an Empatica E4 device [33] on their left wrist which recorded their EDA, BVP and ST signals.The sampling rate of EDA, BVP and ST were 4, 64 and 4 Hz respectively. Figure 1 shows the Empatica E4 device. PD data was collected using The Eye Tribe device at a sampling rate of 60 Hz [34].

Figure 1. Empatica E4 device [33]

The procedure started with the calibration process of the physiological sensors. Due to the device being sensitive to external movements, all participants were asked to limit any unnecessary movement during the experiment in order to avoid adding artefacts to the signals. They were also asked to wear noise cancelling headphones (Bose QuietComfort 20) which helped remove any effects from outside noise during the experiment. The entire experiment was conducted through an interactive website prepared for this purpose. Participants answered some basic demographic questions at the beginning of the experiment. Then the participants listened to each piece of music and gave a series of ratings to the music based on 6 different emotion scales. These scales are i) $sad \rightarrow happy$ ii) $disturbing \rightarrow comforting$ iii) $depressing \rightarrow exciting$ iv) $unpleasant \rightarrow pleasant$ v) $irritating \rightarrow soothing$ vi) $tensing \rightarrow relaxing$. The first 4 ratings are to find participants' general impression about the music itself, and the other 2 ask about the participants' feelings while listening to that piece of music. The metrics are described in detail in [35]. The subjective ratings are based on a 7-point Likert scale, chosen as this is considered the most appropriate number for Likert [36]. At the end of the experiment, participants provided some general comments about the music pieces in a post-experiment questionnaire.

In order to analyse the emotional ratings provided by the participants, we have visualised the ratings based on their valence-arousal level in a two-dimensional emotion model. The original model was proposed by Russell [37] which contained a wider list of emotions. Based on that model we have created our model with the 6 emotion scales used in the study (Figure 2). This is a more ef-

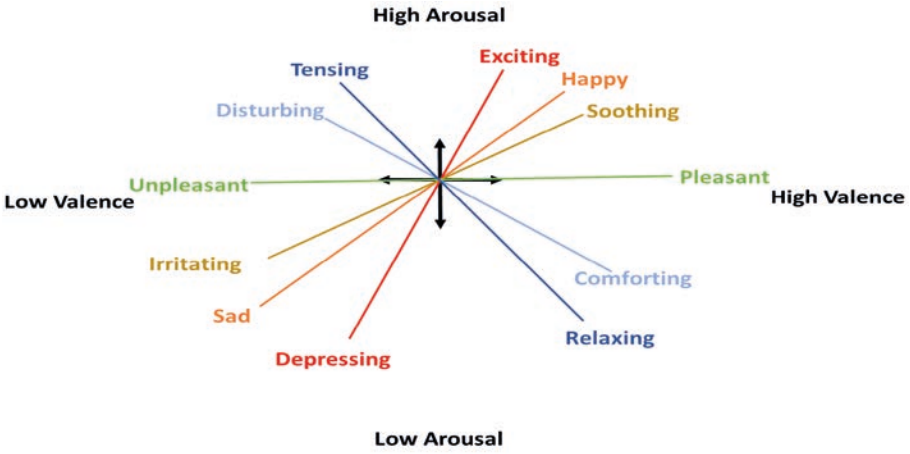


Figure 2. Two Dimensional Emotion Model by Valence and Arousal

fective approach than modeling the emotions with discrete labels because real world stimuli induce blended emotions, and they can be expressed better in multidimensional space [38].

After collecting the raw physiological signals, they are analysed through multiple steps. The complete process is shown in Figure 3.

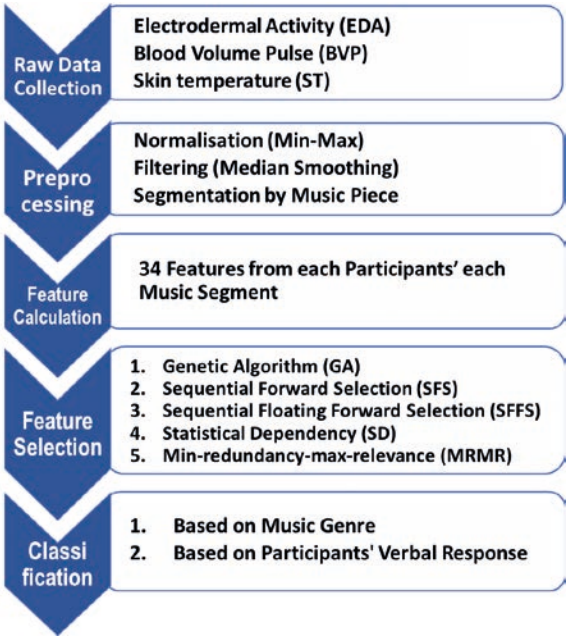


Figure 3. Overall Steps of the Experiment

2.5 Preprocessing

Performing some preprocessing techniques on the collected signals is a crucial part of data analysis. These physiological signals vary widely in range due to the values of the signals being subject-dependent. Therefore, normalizing the data is necessary. Min-Max normalization technique is used to normalize the collected EDA, BVP, ST and PD signals. The equation for min-max normalization is

$$v\prime = \\ = \left(\frac{v-min\_v}{max\_v-min\_v}\right) * (new\_max - new\_min) + (new\_min), \tag{1}$$

where, $v\prime$ corresponds to min-max normalized data and v is the full range of raw data, max_v and min_v are the maximum and minimum value of v respectively. Here we chose $new\_min = 0$ and $new\_max = 1$. Thus, all values were normalized to have a value within the range of 0 to 1. Data from each participant were normalized individually, across all their pieces of music.

Collected physiological signals are also prone to noise artefacts due to participants' movement, blinking and so on. So after normalization, we filtered the data to remove any potential artefacts. We chose median smoothing filter process for this step [39]. In keeping with our previous study [16], we chose a 10 point median filter as the optimum number which results in minimum loss of data. For pupil dilation data, we needed to perform an additional preprocessing step because several data points were empty due to blinking by the participant. In this case, we applied linear interpolation to generate those data points.

2.6 Feature Extraction

After preprocessing, a number of features were extracted from the physiological signals as the recorded features are very large in size and therefore computationally expensive to analyse. Based on a number of papers in the literature [40, 41, 42, 43, 44], we extracted a total of 34 different features (linear and nonlinear) from each of the 4 physiological signals. These include features in both time and frequency domains. Some features were also extracted from both normalized and filtered signals. The first 14 features follow some statistical features used in the analysis of our previous study [16]: we expanded the list with more features which are listed below:

- Root mean square, average value of the power of signals, integrated signals, simple square integral - These features provide basic statistical information of the physiological signal changes that were seen in the duration of music listening.
- Average Amplitude Change - Calculated by first finding the difference between two consecutive samples, then averaging them over the length of the piece of music.
- Log Detector - This feature represents a non-linear characteristic which is calculated by the average logarithm of the signal value over the music length.
- Difference Absolute Standard Deviation Value - Calculated by the standard deviation between two consecutive signals.
- Non Linear Features to Measure Complexity and Correlation - We calculated a number of non-linear features to identify the complexity and correlation properties of the physiological signals. Some of these take self-similarity into account, while others do not. The features are detrended fluctuation analysis (DFA), Hjorth parameters and Hurst exponent. There are 3 Hjorth parameters, we consider extracting mobility for this study.
- Non Linear Features to Measure Randomness - Entropy represents the randomness in physiological signals. Similar to the other non-linear features we extracted, some of these are based on the self-similarity of signals while others are not. We calculate 5 types of entropy for the features. They are, sample entropy, approximate entropy, Shannon's entropy, permutation entropy, fuzzy entropy.

2.7 Feature Selection

Feature selection is often done in order to find useful features and remove any redundant features. Finding an optimum number of features can result in a speedy classification process. Having irrelevant features has been shown to significantly decrease the performance of a classification model [45]. The feature selection process can be done in two ways. One is to rank each of the features and select a fixed number of top ranked features to build the feature set; the other way is to select different subsets of features and classify in order to find the optimal set. We chose two feature ranking algorithms (Statistical Dependency (SD), Minimal-redundancy-maximal-relevance (MRMR)) and four feature subset selection methods (Genetic Algorithm (GA), Random Subset (RSFS), Sequential (SFS) and Sequential Floating (SFFS)) explained in [46, 47].

2.8 Evaluation Measures

While classification accuracy is necessary to show how good the model is, it does not give complete information about the benefit or values of a model. Sometimes, models with a lower accuracy can have a higher predictive power compared to models with higher accuracy [48]. Therefore, some additional measures also need to be reported along with classification accuracy. We also calculate the F-measure (also known as F-Score or F1 Score), which is a commonly used evaluation measure represented by the harmonic mean of precision and recall. Precision refers to the fraction of the predicted labels matched while recall refers to the fraction of reference labels matched. We also reported the precision and recall (also referred as sensitivity), specificity (true negative rate) and geometric mean values (measure of central tendency).

3 Visualisation of the Physiological Signals - Gingerbread Animation

We were interested to investigate the effects of analysing the physiological signals using different visualisation techniques. As a preliminary study, we visualised the physiological signals in a 2D graph. Each participant's data were segmented according to the audio stimuli length. EDA, BVP and ST values were represented in red, blue and green colour respectively. We did not use PD values in this preliminary study. Figure 4 shows a sample graph image used in this analysis.

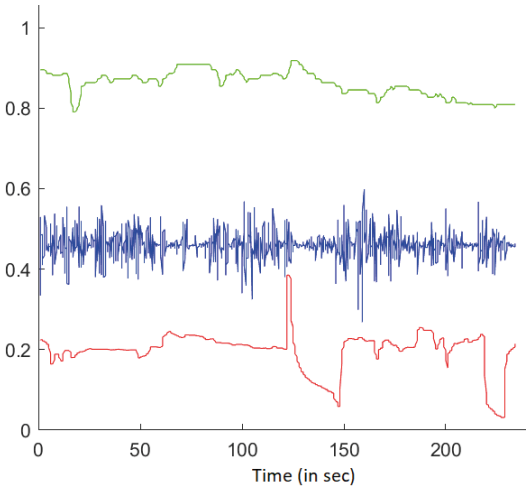


Figure 4. Physiological Signals Representation as a Graph (Blue = BVP, Red = EDA, Green = ST)

When using physiological signals, in particular many physiological signals over a longer period, it is difficult to visualise the data. We have devised an approach we call *Gingerbread Animation* which uses a stylized 2D representation of a human body and visually represents the time series of physiological signals propagating on that 2D surface, which can be presented as a video. In the gingerbread animation, we have used BVP, PD, EDA and ST signals, which can be represented by red, green, blue and gray colours respectively. The locations of data representation also reflect the locations where these signals are generated where possible. The PD, BVP, EDA and ST signals are displayed in right eye, heart, left wrist and right foot area respectively. These colours can combine and create mixed colours on the surface. Thus, we produce a sequence of images (forming a video) representing each experimental trial, and retaining a representation of each signal – the colours mix, but the R, G, B values are not lost. This representation leads to an additional benefit, that we can make use of the highly advanced computer vision techniques available for images to classify and predict based on our new video data.

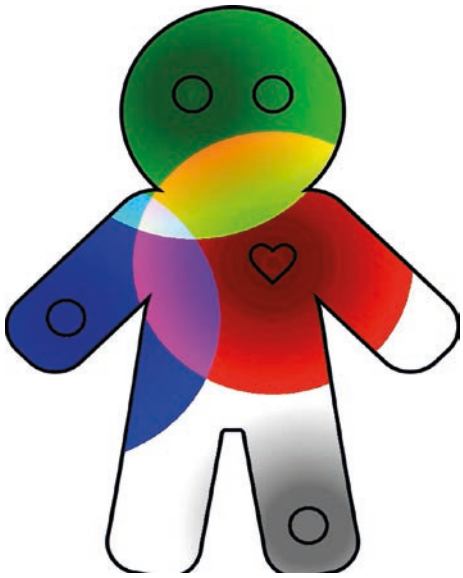


Figure 5. Physiological Signals Representation in an Animation (Red = BVP, Blue = EDA, Green = PD, Gray = ST

Figure 5 shows a representative image, being a few seconds into an audio stimulus. Each datum is represented as a ring with a fixed maximum width in the animation. The latest data appears in the middle of the circle, for each type of signal, up to 40 time steps of data are showing at the same time. These 40 rings constitute an entire circle. The older the data, the closer to the outside edge the circle it moves to, which simulates the effect of data rippling out.

Physiological data is mapped to the RGB model in the animation, in which (0,0,0) is black and (255,255,255) is white. To make the visualisation more in line with human intuition, the background is set as white, so as to highlight stronger signals that appear darker due to lower RGB values.

As the data spreads, the intensity continues to decay until it drops to 0, which is represented by 255 in the RGB model. For example, when a BVP datum is 0.8, it is represented as (51.2,0,0) in the RGB model in the middle of the circle when it first appears and then after spreading out, it begins to decay slowly and ends up as (255,0,0) which is seen as a bright red color in the animation.

In areas where multiple types of signal overlap, the overlapped RGB value is added by the RGB values of each signal. For example, a BVP datum of 0.8 (i.e. (51.2,0,0) in RGB) meets a wrist datum of 0.5 (i.e. (0,0,128) in RGB), and the resulting output is (51.2,0,128) in RGB.

We can see that in some parts of the image the amplitudes of the original signals can still be easily seen as the colours have not yet begun to mix. It is noticed that the BVP signals vary rhythmically, while the ST varies in a much smoothed manner, while the EDA is not rhythmic in this fashion. We can also see some regions where the colour has begun to mix, and producing visually pleasing complex patterns related to the data.

4 Results and Discussion

4.1 Classification Results

Two types of classification were performed on the experiment data. The first was to classify the data into 3 music genres, the other was to classify based on the subjective rating of emotions given by each participant. To classify the signals, 3 different classification techniques were used for comparison. They are: Neural Network (NN), K-Nearest Neighbor (KNN) and Support Vector Machine (SVM). Using these methods we performed the classification in 5 different conditions. We used EDA, BVP, ST and PD features individually for classification, and also combined features from all 4 signals. The entire process was done using all the features and also features selected by the 6 feature selection methods. For the 2 feature ranking methods SD and MRMR, we have chosen the top 12 features to use in the classification process. A leave-one-observer-out process was performed as the validation approach. Classification was performed using MATLAB R2018a software with an Intel(R) Core(TM) i7-5200U processor with 3.60 GHz, 16.00 GB of RAM and Microsoft Windows 10 Enterprise 64-bit operating system.

For classification using neural networks, a pattern recognition network was constructed with one input layer, one hidden layer and one output layer. The hidden layer consisted of 30 nodes. This was chosen based on the comparison of different hidden layer sizes done in our previous study [16]. Other parameters of the network were: Levenberg—Marquardt method as network training function and mean squared normalized error as performance function. The classification process was done 20 times and the average of those results were selected. For KNN, we performed the process using node size 3-30 and chose the best results. K= 5 or 7 resulted in best outputs for all cases. We used Minkowski as the distance metric. The multiclass SVM chosen for this study uses tree learner and one-versus-all coding design.

For classification using the graphs constructed from physiological signals, we used a pre-trained convolutional neural network (CNN) resnet18 and modified the final layer in order to train (fine-tune) the model using our graph images. Resnet introduced skip connections which help resolve the vanishing gradient issue [49]. Figure 6 shows the resnet18 architecture.

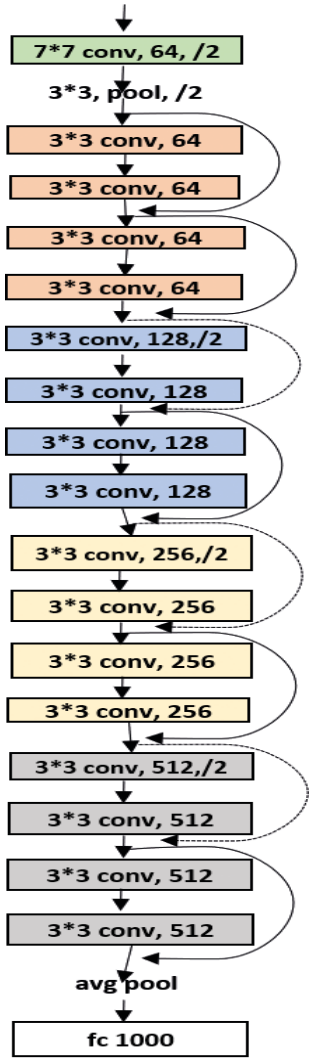


Figure 6. Resnet18 Architecture

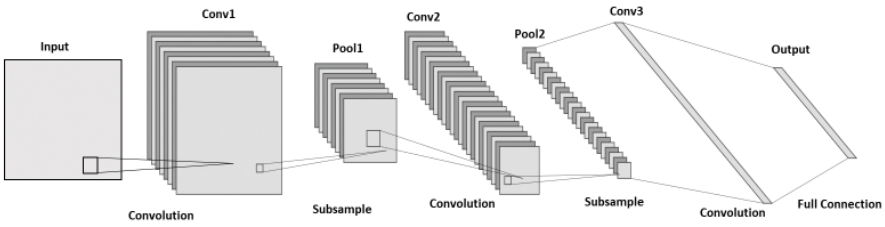


Figure 7. CNN Architecture for Gingerbread Animation

To classify the images obtained from our Gingerbread Animation, we constructed a CNN using stochastic gradient descent with momentum (SGDM) with an initial learning rate of 0.005, mini batch size of 32. We chose a version of the classic Lenet-5 architecture containing three convolutional layers, two max pooling layers, a fully-connected layer and a softmax classifier. Figure 7 shows the CNN architecture.

Certain patterns are observed from the classification results which are described below.

4.1.1 Neural network performs best among all classifiers

We compared the results of all 5 classification approaches and in every case NN performed significantly better than KNN and SVM. Figure 8 shows the accuracy results based on the participants' subjective rating based on the emotion scale *tensing* $\rightarrow$ *relaxing*, using all features.

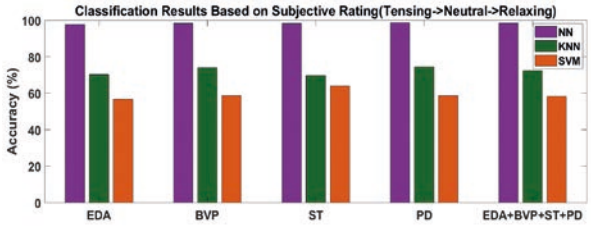


Figure 8. Classification Result Based on Subjective Rating (*tensing* $\rightarrow$ *relaxing*)

From Figure 8 we can see that NN performs best in terms of accuracy in all 5 combination of features. Neural network gives the accuracy of 97.7%, 98.3%, 98.3%, 98.7% and 98.5% accuracy using EDA, BVP, ST, PD and EDA+BVP+ST+PD features respectively. In comparison, KNN gives 70.3%, 73.9%, 69.8%, 74.5%, 77.6% and SVM gives 56.8%, 58,9%, 64.1%, 58.9%, 62.5% accuracy. This pattern prevails in classification using features from the feature selection methods as well. This study solidifies the results from our previous study in showing that a simple NN can be a strong system in classifying physiological signals.

4.1.2 Feature selection produces best results for music genre Classification

We compared the accuracy results of different feature selection methods for music genre classification and the results are shown in figure 9.

It can be seen that for the EDA, BVP, ST, PD and EDA+BVP+ST+PD feature combinations, the RSFS, MRMR and SD methods result in the best NN accuracy. Both KNN and SVM also produce their best results using feature selection methods. The results show similar patterns across all evaluation measures. Table 2 shows the results of all 6 evaluation measures for NN classification of music genres. The table does not include results using GA and SFFS as feature selection methods because they do not achieve the highest values in any of the evaluation measures.

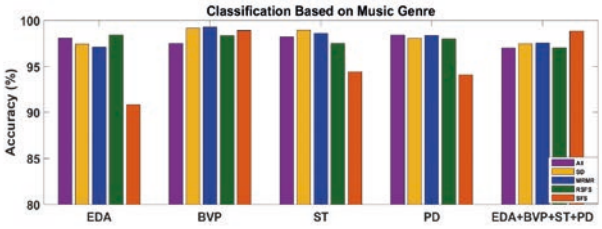


Figure 9. Classification Result Based on Music Genre

From the table we can observe that the high score for all evaluation measures is reached by a feature selection method. A few exceptions can be

Table 2. Classification Results Based on Music Genre

		All	SD	MRMR	RSFS	SFS
EDA	Accuracy	0.981	0.975	0.971	**0.984**	0.908
	Precision	0.982	0.976	0.979	**0.983**	0.875
	Recall	0.959	0.947	0.934	**0.969**	0.847
	Specificity	0.991	0.988	0.989	**0.992**	0.939
	F-Measure	0.971	0.961	0.956	**0.976**	0.860
	G-mean	0.975	0.967	0.961	**0.980**	0.892
		All	SD	MRMR	RSFS	SFS
BVP	Accuracy	0.975	0.992	**0.993**	0.983	0.989
	Precision	0.957	0.992	**0.994**	0.974	0.984
	Recall	0.969	0.983	**0.984**	0.977	0.983
	Specificity	0.978	0.996	**0.997**	0.987	0.992
	F-Measure	0.963	0.988	**0.989**	0.975	0.984
	G-mean	0.973	0.989	**0.991**	0.982	0.988
		All	SD	MRMR	RSFS	SFS
ST	Accuracy	0.982	**0.989**	0.986	0.975	0.944
	Precision	0.959	**0.983**	0.979	0.962	0.925
	Recall	**0.989**	0.985	0.979	0.963	0.906
	Specificity	0.979	**0.991**	0.989	0.981	0.963
	F-Measure	0.974	**0.984**	0.979	0.963	0.915
	G-mean	0.984	**0.988**	0.984	0.972	0.934
		All	SD	MRMR	RSFS	SFS
PD	Accuracy	0.983	0.981	**0.984**	0.979	0.941
	Precision	0.984	0.979	**0.985**	0.981	0.924
	Recall	**0.968**	0.963	0.966	0.959	0.898
	Specificity	**0.992**	0.989	0.993	0.991	0.963
	F-Measure	0.975	0.971	**0.975**	0.969	0.91
	G-mean	0.979	0.976	**0.979**	0.974	0.929
		All	SD	MRMR	RSFS	SFS
EDA+ BVP+ ST+ PD	Accuracy	0.97	**0.978**	0.977	0.972	0.958
	Precision	0.952	0.951	0.948	**0.957**	0.937
	Recall	0.96	0.984	**0.986**	0.96	0.936
	Specificity	0.975	0.975	0.973	**0.979**	0.969
	F-Measure	0.956	**0.967**	0.966	0.959	0.936
	G-mean	0.968	**0.979**	0.979	0.969	0.952

observed such as, the condition using ST features where the highest recall score is achieved by using all features. Furthermore, using PD features, the highest recall and specificity is achieved using the full sets of features. However, we can see that in those combinations the highest F-measure is reached by SD and MRMR method respectively, aligning with the other measure values. F-measure is the harmonic mean of precision and recall, which takes both false positive and false negative values into account. Recall does not consider false positive values, therefore the F-measure is a stronger measure for evaluating a model, compared to just precision or recall. Thus we can say that using a smaller subset of the features not only reduces the computational time, but also increases accuracy of our model in classifying different music genres.

We also compare this result to our previous study [16], where we analysed only the EDA signals with 14 extracted features. The study reached the highest accuracy of 96.8% for music genre classification. In our current study, the accuracy using EDA features has increased to 98.4%, and overall the highest accuracy of our classification model is 99.3%, using BVP features. A better set of features has contributed to the improved accuracy of our classification models.

4.2 Statistical analysis on all evaluation measures

Results of the 6 evaluation measures for NN classification across all 5 conditions were analysed using analysis of variance (ANOVA) test. A one-way ANOVA test showed high statistical significance ($p < 0.01$) for all of the evaluation measures. We also compared the accuracy results for all pairs of feature selection methods for statistical significance. The results are shown in Table 3.

In Table 3, the numbers in colour and bold are the pairs that show meaningful differences. Red colour shows a significance of $p < 0.05$, while blue colour shows significance $p < 0.01$ and green colour shows significance at threshold $p < 0.001$. We further observe that both SFS and SFFS reach high significance values in comparison with other selection methods. This is reflected in the number and type of features chosen by these methods as well. It can be clearly seen from the table that different combinations of features in the model result in significant differences in model accuracy. Therefore, in the section below we discuss some of the features that were shown to be useful for our classification models.

Table 3. Significance Values for All Pairs of Feature Selection Methods

All						
SD	**0.00003**					
MRMR	**0.00009**	0.386				
GA	**0.0002**	**0.029**	0.085			
RSFS	**0.006**	0.139	0.119	0.782		
SFS	**0.000002**	**0.00002**	**0.000009**	**0.000006**	**0.00004**	
SFFS	**0.000009**	**0.00008**	**0.00006**	**0.00003**	**0.0003**	0.335
	All	SD	MRMR	GA	RSFS	SFS

4.3 Top features selected by Feature Selection Methods

We counted the number of times each feature was chosen by the 6 feature selection methods in all of our classification models. Based on that, we report the top 12 features from the 34 features we extracted in Table 4. Unless specifically mentioned, most of the features were extracted from filtered signals.

Table 4. Top 12 features selected by all methods

Feature Names	Feature Type
Number of peaks (both normalized and filtered), Variance, Sum, Absolute Sum, Simple Square Integral	Linear features from time domain
Mean, Minimum, Maximum of the first 16 data points from Welch Power Spectrum Density Analysis	Linear features from frequency domain
Sample and Approximate entropy, Hjorth paremeters (Mobility)	Non-linear features from time domain

The list gives us some interesting insights into what types of features are best to represent the 4 physiological signals' changes. The number of peaks for both normalized and filtered values were selected the most times by all methods. These peaks are thus the most valuable feature that reflects the SCR occurrences (rapidly changing states). SCR occurrences are considered to be most useful in reflecting autonomic arousal [50]. Although the normalized and filtered signal features are quite similar, they clearly do not add redundancy to the sys-

tem. With some signals, useful peaks might be removed due to the filtering process. In those cases, peaks in the normalized signals proved to be more useful. We also notice that the 3 features extracted from the Welch power spectrum density analysis appeared in the top features list. This shows that the frequency domain features can be very useful to identify patterns in this signals. Future work will include extracting more frequency domain features.

Some of the other interesting features are entropies and mobility. All of these features represent the level of complexity of the signals. Features like entropies can effectively capture short range correlations and thus, they are effective in identifying transient emotional state changes [51].

4.4 Genetic Algorithm produces best results among all feature selection methods

We further analysed the NN accuracy results for all feature selection methods and ranked the methods based on how many times that method achieved highest accuracy. The list below shows the rank of the feature selection methods and their frequency of achieving the highest accuracy.

- GA - 11
- MRMR - 7
- RSFS - 5
- SD - 4
- SFS - 1
- SFFS - 0

It should be noted that GA was not able to achieve the highest accuracy for music genre classification in any combination. But it was able to achieve the highest accuracy in most combinations for the 6 emotion based classifications. For the cases where GA was not able to reach the highest accuracy, it was still able to achieve close to reaching the highest. In our previous study [16], we reported that for the 3 emotions that have a negative slope (*depressing* $\rightarrow$ *neutral* $\rightarrow$ *exciting*, *sad* $\rightarrow$ *neutral* $\rightarrow$ *happy* and *irritating* $\rightarrow$ *neutral* $\rightarrow$ *soothing*) in the emotion model (shown in figure 2) GA feature selection methods performed the best. For the emotions that have a slope of 0 or a positive value (*disturbing* $\rightarrow$ *neutral* $\rightarrow$ *comforting*, *relaxing* $\rightarrow$ *neutral* $\rightarrow$ *tensing* and *unpleasant* $\rightarrow$ *neutral* $\rightarrow$ *pleasant*) SD/MRMR methods work the best. However, further analysis using more physiological signals and a wider set of features showed that GA is able to select a robust set of features for all 6 emotions based classifications. Therefore, we recommend using GA feature selection method for classification of music based on different emotion ratings.

4.5 Gingerbread Animation is an effective method for classifying emotions

We used two different subjective rating of emotions (*sad* $\rightarrow$ *happy* and *tensing* $\rightarrow$ *relaxing*) for classification using the graph and animation images. We used a 3-fold cross validation approach to validate the accuracy of the network. We randomly selected 16 participants' data for training and 8 participants' data for testing. The graph images trained using a pre-trained CNN achieved 61.9% accuracy for the emotions *sad* $\rightarrow$ *happy* and 73.4% accuracy for *tensing* $\rightarrow$ *relaxing*. In comparison, the animation images reach 68.1% and 74.8% for the same emotion pairs. We note that the comparison is not exact, the graph visualisations show 250 sec of data with 3 physiological signals, while the gingerbread animation shows 10 sec of data (40 time steps at 4 Hz) in each frame for 4 physiological signals. These results suggest that our gingerbread animation can be both a visually attractive and effective approach to identify emotion from human physiology using state-of-the-art machine learning methods.

Another observation is that humans are often incorrect in giving subjective ratings to their emotional response to the audio stimuli. To initially label the emotions according to subjective rating, we used the majority voting approach to label each audio stimuli. Afterwards we labelled each audio stimulus based on each participants' individual subjective response. This resulted in the accuracy dropping from 62% to 50% for *sad* $\rightarrow$ *happy* and from 73.4% to 47.4% for *tensing* $\rightarrow$ *relaxing*. Therefore, we can see that some participants are incorrect (compared to the population) in rating their emotions listening to the audio stimuli. However, on average the participants' response correlate with their

physiological response. Thus, each person reacted as expected from the population view as to the emotional content of that piece of music, while their own conscious view was incorrect.

There is scope for improvement in both the gingerbread animation and our computational models using that data. In particular, we note that our results with the simple neural network are based on substantial work in preprocessing, which could not be done by the simple neural network model. Using a pre-trained CNN, we achieved notable results using just the raw data. In the future, the overlapping of the colours need to be re-defined and improved to make the color mixture more natural and the edges more blurred. In addition, further preprocessing needs to be implemented in the input physiological signals so that some slight signals changes can be more clearly displayed in the animation. As the current labels are based on participants' subjective response so the dataset can be biased due to the small sample size. More data and tuning the network options can improve the accuracy of the model. However, the preliminary results are encouraging and we aim to improve the animations and fine tuning the network in order to build a robust model, with an expectation of surpassing the results from our current work.

5 Conclusions

In this paper, we conducted a study to collect participants' EDA, BVP, ST & PD activity while they listened to different genre of music. These collected physiological signals were first normalized, then smoothed. Then a range of features were extracted and a set of feature selection methods were applied to find the best features. Analysis using 3 different classification methods (NN, KNN and SVM) were performed and evaluated using 6 different measures. All the results were compared using features from a specific signal and also the combination of all signals. Neural networks achieved the highest accuracy across all different conditions with the highest accuracy of 99.2% and 98.5% in classifying music based on genre type and human emotions respectively. Furthermore, GA feature selection method has shown to be best for classifying music based on subjective emotion ratings by participants'

We have introduced a novel animation technique to both visualise physiological signals and to make them accessible to computer vision classifiers. Preliminary results using a CNN achieved upto 74.8% accuracy in identifying different music based on the subjective rating of participants' emotion. Our approach will allow us to leverage state-of-the-art computer vision approaches in analysing multiple physiological signals collected during affective experiments with high effectiveness.

There are certain limitations to our work. Due to the difficulty of finding participants' and collecting data, the number of samples is not very large and therefore we could not explore applying higher power deep learning models. In addition, due to labelling the stimuli according to participants' subjective rating of emotion, the dataset may have introduced bias and thus weaken the predictive power of our models. Future work will involve collecting more data to be able to build a more robust system. We also want to compare the results of our techniques using physiological signals such as electroencephalogram (EEG), as brain activity has also shown to be a strong indicator to understand effects of music [52]. In addition, further analysis on useful features will be conducted to identify if the features correlate with certain patterns in music. This will be beneficial in identifying which music pieces are good for music therapy. It can also reveal which pieces could trigger epileptic seizures for each participant, and thus should be avoided. Finally, more comparisons with our models could be made using publicly available physiological datasets of patients having mental disorders. Research studies such as ours will strengthen the motivation to use physiological signals in the area of medical and affective computing and music therapy treatments in improving mental health.

Acknowledgement

The authors would like to thank the participants who took part in this research. Data relating to this study will be made publicly available upon completion and publication of the complete study.

References

[1] A. Bardekar and A. A. Gurjar, Study of Indian Classical Ragas Structure and its Influence on Human Body for Music Therapy, in 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016, pp. 119-123: IEEE.

[2] C. L. Baldwin and B. A. Lewis, Positive valence music restores executive control over sustained attention, PLOS ONE, vol. 12, no. 11, p. e0186231, 2017.

[3] L. Harmat, J. Takács, and R. Bodizs, Music improves sleep quality in students, Journal of advanced nursing, vol. 62, no. 3, pp. 327-335, 2008.

[4] G. Coppola et al., Mozart's music in children with drug-refractory epileptic encephalopathies, Epilepsy & Behavior, vol. 50, pp. 18-22, 2015.

[5] M. Z. Hossain, Observer's galvanic skin response for discriminating real from fake smiles, 2016.

[6] L. Chen, T. Gedeon, M. Z. Hossain, and S. Caldwell, Are you really angry?: detecting emotion veracity as a proposed tool for interaction, presented at the Proceedings of the 29th Australian Conference on Computer-Human Interaction, Brisbane, Queensland, Australia, 2017.

[7] J. A. Healey and R. W. Picard, Detecting stress during real-world driving tasks using physiological sensors, IEEE Transactions on intelligent transportation systems, vol. 6, no. 2, pp. 156-166, 2005.

[8] Y. Nagai, L. H. Goldstein, P. B. Fenwick, and M. R. Trimble, Clinical efficacy of galvanic skin response biofeedback training in reducing seizures in adult epilepsy: a preliminary randomized controlled study, Epilepsy & Behavior, vol. 5, no. 2, pp. 216-223, 2004.

[9] L. Harrison and P. Loui, Thrills, chills, frissons, and skin orgasms: toward an integrative model of transcendent psychophysiological experiences in music, Frontiers in psychology, vol. 5, p. 790, 2014.

[10] D. Huron and E. Margulis, Musical Expectancy and Thrills, Handbook of Music and Emotion: Theory, Research, Applications, pp. 575-604, 07/29 2011.

[11] M. Guhn, A. Hamm, and M. Zentner, Physiological and musico-acoustic correlates of the chill response, Music Perception: An Interdisciplinary Journal, vol. 24, no. 5, pp. 473-484, 2007.

[12] D. G. Craig, An exploratory study of physiological changes during "chills" induced by music, Musicae scientiae, vol. 9, no. 2, pp. 273-287, 2005.

[13] K. H. Kim, S. W. Bang, and S. R. Kim, Emotion recognition system using short-term monitoring of physiological signals, Medical and biological engineering and computing, vol. 42, no. 3, pp. 419-427, 2004.

[14] M. Z. Hossain, T. Gedeon, and R. Sankaranarayana, Using temporal features of observers' physiological measures to distinguish between genuine and fake smiles, IEEE Transactions on Affective Computing, pp. 1-1, 2018.

[15] A. Haag, S. Goronzy, P. Schaich, and J. Williams, Emotion recognition using bio-sensors: First steps towards an automatic system, in Tutorial and research workshop on affective dialogue systems, 2004, pp. 36-48: Springer.

[16] J. S. Rahman, T. Gedeon, S. Caldwell, R. Jones, M. Z. Hossain, and X. Zhu, Melodious Micro-frissons: Detecting Music Genres from Skin Response, in International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019: IEEE.

[17] J. R. Hughes and J. J. Fino, The Mozart effect: distinctive aspects of the music—a clue to brain coding?, Clinical Electroencephalography, vol. 31, no. 2, pp. 94-103, 2000.

[18] L. C. Lin et al., Parasympathetic activation is involved in reducing epileptiform discharges when listening to Mozart music, Clin Neurophysiol, vol. 124, no. 8, pp. 1528-35, Aug 2013.

[19] R. McCraty, The effects of different types of music on mood, tension, and mental clarity."

[20] Youtube. (2016). Gamma Brain Energizer - 40 Hz - Clean Mental Energy - Focus Music - Binaural Beats. Available: https://www.youtube.com/watch?v=9wrFk5vuOsk

[21] Youtube. (2017). Serotonin Release Music with Alpha Waves - Binaural Beats Relaxing Music, Happiness Frequency. Available: https://www.youtube.com/watch?v=9TPSs16DwbA

[22] N. Hurless, A. Mekic, S. Pena, E. Humphries, H. Gentry, and D. Nichols, Music genre preference and tempo alter alpha and beta waves in human non-musicians.

[23] Billboard Year End Chart. Available: https://www.billboard.com/charts/year-end

[24] D. J. Thurman et al., Standards for epidemiologic studies and surveillance of epilepsy, Epilepsia, vol. 52, pp. 2-26, 2011.

[25] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen, Galvanic skin response (GSR) as an index of cognitive load, in CHI'07 extended abstracts on Human factors in computing systems, 2007, pp. 2651-2656: ACM.

[26] T. Lin, M. Omata, W. Hu, and A. Imamiya, Do physiological data relate to traditional usability indexes?, in Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future, 2005, pp. 1-10: Computer-Human Interaction Special Interest Group (CHISIG) of Australia.

[27] S. Reisman, Measurement of physiological stress, in Bioengineering Conference, 1997., Proceedings of the IEEE 1997 23rd Northeast, 1997, pp. 21-23: IEEE.

[28] R. A. McFarland, Relationship of skin temperature changes to the emotions accompanying music, Biofeedback and Self-regulation, vol. 10, no. 3, pp. 255-267, 1985.

[29] T. Partala and V. Surakka, Pupil size variation as an indication of affective processing, International journal of human-computer studies, vol. 59, no. 1-2, pp. 185-198, 2003.

[30] R. S. Larsen and J. Waters, Neuromodulatory correlates of pupil dilation, Frontiers in neural circuits, vol. 12, p. 21, 2018.

[31] J. Zhai and A. Barreto, Stress Recognition Using Non-invasive Technology, in FLAIRS Conference, pp. 395-401, 2006.

[32] M. W. Weiss, S. E. Trehub, E. G. Schellenberg, and P. Habashi, Pupils dilate for vocal or familiar music, Journal of Experimental Psychology: Human Perception and Performance, vol. 42, no. 8, p. 1061, 2016.

[33] E4 wristband from empatica. Available: https://www.empatica.com/research/e4/

[34] The Eye Tribe. Available: http://theeyetribe.com/about/index.html

[35] J. L. Walker, Subjective reactions to music and brainwave rhythms, Physiological Psychology, vol. 5, no. 4, pp. 483-489, 1977.

[36] D. F. Alwin, Feeling thermometers versus 7-point scales: Which are better?, Sociological Methods & Research, vol. 25, no. 3, pp. 318-340, 1997.

[37] J. A. Russell, A circumplex model of affect, Journal of personality and social psychology, vol. 39, no. 6, p. 1161, 1980.

[38] J. Kim and E. Andre, Emotion recognition based on physiological changes in music listening, IEEE Trans Pattern Anal Mach Intell, vol. 30, no. 12, pp. 2067-83, Dec 2008.

[39] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, Physiological signals based human emotion recognition: a review, in 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, 2011, pp. 410-415: IEEE.

[40] R. W. Picard, E. Vyzas, and J. Healey, Toward machine emotional intelligence: Analysis of affective physiological state, IEEE transactions on pattern analysis and machine intelligence, vol. 23, no. 10, pp. 1175-1191, 2001.

[41] U. R. Acharya et al., Characterization of focal EEG signals: a review, Future Generation Computer Systems, vol. 91, pp. 290-299, 2019.

[42] R. Chowdhury, M. Reaz, M. Ali, A. Bakar, K. Chellappan, and T. Chang, Surface electromyography signal processing and classification techniques, Sensors, vol. 13, no. 9, pp. 12431-12466, 2013.

[43] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 38, no. 3, pp. 502-512, 2008.

[44] T. Triwiyanto, O. Wahyunggoro, H. A. Nugroho, and H. Herianto, An investigation into time domain features of surface electromyography to estimate the elbow joint angle, Advances in Electrical and Electronic Engineering, vol. 15, no. 3, pp. 448-458, 2017.

[45] R. Kohavi and G. H. John, Wrappers for feature subset selection, Artificial intelligence, vol. 97, no. 1-2, pp. 273-324, 1997.

[46] J. Pohjalainen, O. Räsänen, and S. Kadioglu, Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits, Computer Speech & Language, vol. 29, no. 1, pp. 145-171, 2015.

[47] J. Yang and V. Honavar, Feature subset selection using a genetic algorithm, in Feature extraction, construction and selection: Springer, 1998, pp. 117-136.

[48] F. J. Valverde-Albacete and C. Peláez-Moreno, 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox, PloS one, vol. 9, no. 1, p. e84217, 2014.

[49] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[50] M. G. N. Bos, P. Jentgens, T. Beckers, and M. Kindt, Psychophysiological response patterns to affective film stimuli, (in eng), PloS one, vol. 8, no. 4, pp. e62661-e62661, 2013.

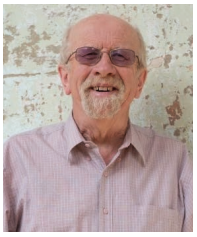
[51] S. Jerritta, M. Murugappan, K. Wan, and S. Yaacob, Emotion Detection from QRS Complex of ECG Signals Using Hurst Exponent for Different Age Groups, in 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, 2013, pp. 849-854.

[52] J. S. Rahman, T. Gedeon, S. Caldwell and R. Jones, Brain Melody Informatics: Analysing Effects of Music on Brainwave Patterns, in International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom, 2020: IEEE.

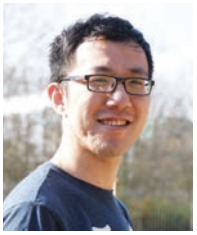
Jessica Sharmin Rahman received her B.Sc. (Hons) degree from the University of Dhaka, Bangladesh. She is now a Ph.D. Student in the Human-Centred Computing (HCC) group of the Research School of Computer Science at the Australian National University(ANU). She also works as an academic tutor at ANU. Her research interests include affective computing, emotion recognition, brain informatics. Her talk on her PhD thesis won the people's choice award in the ANU 3 Minute Thesis (3MT) competition 2020.

Tom Gedeon received the B.Sc. (Hons) and Ph.D. degrees from the University of Western Australia. He is currently chair Professor of Computer Science at the Australian National University, Canberra, Australia, and leads the Human Centred Computing Group at the Research School of Computer Science. His research interests are in bio-inspired computing and in human-centred computing. He is a former president of the Asia-Pacific Neural Network Assembly and a former President of the Computing Research and Education Association of Australasia. He serves on journal advisory boards as member or editor. He is a senior member of the IEEE.

Sabrina Caldwell holds a Ph.D. in Computer Science (ANU2016) and a Ph.D. in the Arts and Social Sciences (ANU 2008), as well as B.Sc./BA(Hons) (ANU2003). Her research interests centre around biometric signal processing and artificial intelligence to investigate how humans respond to deception and credibility, with the goal of introducing innovative solutions for bolstering online image and knowledge credibility. She teaches software development management and has an extensive background in the Information Technology industry as a project manager.



Richard Jones is an adjunct Professor in the Research School of Computer Science in ANU after a long career in commercial applied RD for mainly small software companies. Educated at Trinity College Dublin (B.A. Maths, Ph.D. mathematical physics) he has been involved in software R&D since 1969 when he wrote an algebraic manipulation system to support his Ph.D. work. He was heavily involved in developing an early text retrieval program (STATUS), and has since worked in a range of text analysis software that became commercially viable. His work in ANU has spanned affective computing to software engineering.



Zi Jin received his B.Sc. (Hons) and master degrees from the Australian National University. He is currently a programmer, working in Prof. Tom Gedeon's Human-Centred Computing (HCC) group at the Research School of Computer Science. His research interest includes natural language processing and emotion recognition.