EMOTION BASED MUSIC RECOMMENDATION SYSTEM USING WEARABLE PHYSIOLOGICAL SENSORS

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

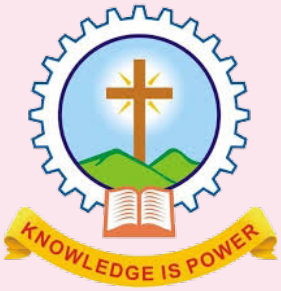
COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

SHILPA KV



Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

EMOTION BASED MUSIC RECOMMENDATION SYSTEM USING WEARABLE PHYSIOLOGICAL SENSORS

Seminar Report

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

of

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

Submitted By

SHILPA KV

Department of Computer Science & Engineering
Mar Athanasius College Of Engineering Kothamangalam

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MAR ATHANASIUS COLLEGE OF ENGINEERING
KOTHAMANGALAM

CERTIFICATE

This is to certify that the report entitled **Emotion Based Music Recommendation System Using Wearable Physiological Sensors** *submitted by* **Ms.SHILPA KV**, *Reg. No.* **MAC15CS053** *towards partial fulfillment of the requirement for the award of Degree of Bachelor of Technology in Computer science and Engineering from APJ Abdul Kalam Technological University for December 2018 is a bonafide record of the seminar carried out by her under our supervision and guidance.*

.....	.....	.....
Prof. Joby George	**Prof. Neethu Subash**	**Dr. Surekha Mariam Varghese**
Faculty Guide	***Faculty Guide***	***Head of the Department***

Date: Dept. Seal

ACKNOWLEDGEMENT

First and foremost, I sincerely thank the God Almighty for his grace for the successful and timely completion of the seminar.

I express my sincere gratitude and thanks to Dr. Solly George, Principal and Dr. Surekha Mariam Varghese, Head Of the Department for providing the necessary facilities and their encouragement and support.

I owe special thanks to the staff-in-charge Prof. Joby George , Prof. Neethu Subash and Prof. Joby Anu Mathew for their corrections, suggestions and sincere efforts to co-ordinate the seminar under a tight schedule.

I express my sincere thanks to staff members in the Department of Computer Science and Engineering who have taken sincere efforts in helping me to conduct this seminar.

Finally, I would like to acknowledge the heartfelt efforts, comments, criticisms, co-operation and tremendous support given to me by my dear friends during the preparation of the seminar and also during the presentation without whose support this work would have been all the more difficult to accomplish.

ABSTRACT

Most of the existing music recommendation systems use collaborative or content based recommendation engines. However, the music choice of a user is not only dependent to the historical preferences or music contents. But also dependent to the mood of that user. The proposed approach use an emotion based music recommendation frame work that learns the emotion of a user from the signals obtained via wearable physiological sensors. In particular, the emotion of a user is classified by a wearable computing device which is integrated with a galvanic skin response (GSR) and photo plethysmography (PPG) physiological sensors. This emotion information is fed to any collaborative or content based recommendation engine as a supplementary data. Thus existing recommendation engine performances can be increased using these data. The results of comprehensive experiments on real data confirm the accuracy of the proposed emotion classification system that can be integrated to any recommendation engine.

Contents

List of Figures

List of Abbreviations

GSR	Galvanic Skin Response
PPG	Photo Plethysmography Signals
EVADM	Emotion Valence-arousal Dimensional Model
EDA	Electro Dermal Activity
HRV	Heart Rate Variability
FS	Feature Sets
FLF	Feature Level Fusion

Introduction

Wearable computing is the study or practice of inventing, designing, building or using body-worn computational and sensory devices that leverages a new type of human-computer interaction with a body-attached component that is always up and running. As the number of wearable computing device users are growing every year, their areas of utilization are also rapidly increasing. They have influenced medical care, fitness, aging, disabilities, education, transportation,finance, gaming, and music industries.

Recommendation engines are algorithms which aim to provide the most relevant items to the user by filtering useful information from a huge pool of data. Recommendation engines may discover data patterns in the data set by learning users choices and produce the outcomes that co-relates to their needs and interests . Most of the recommender systems do not consider human emotions or expressions. However, emotions have noticeable influence on daily life of people. For a rich set of applications including human-robot interaction, computer aided tutoring, emotion aware interactive games, neuro marketing, socially intelligent software apps, computers should consider the emotions of their human conversation partners. Speech analytics and facial expressions have been used for emotion detection. However, in case of human beings prefer to camouflage their expressions, using only speech signals or facial expression signals may not be enough to detect emotions reliably. Compared with facial expressions, using physiological signals is a more reliable method to track and recognize emotions and internal cognitive processes of people.

Our motivation in this work is to use emotion recognition techniques with wearable computing devices to generate additional inputs for music recommender systems algorithm, and to enhance the accuracy of the resulting music recommendations. In our previous works, we have studied emotion recognition from only GSR signals. In this study we are enriching signals with PPG and propose a data fusion based emotion recognition method for music recommendation

engines . The proposed wearable attached music recommendation framework utilizes not only the users demographics but also his/her emotion state at the time of recommendation. Using GSR and PPG signals we have obtained promising results for emotion prediction.

Existing system

2.1 Related works

Ekman et al. have stated that facial expressions can be categorized into seven main categories including angry, disgust, happy, fearful, surprise, sad and neutral . In other words, these facial expressions were globally same for all races, social strata and age brackets and were recognized same among distinct cultures.

Emotion-specific activity in the autonomic nervous system was generated by constructing facial prototypes of emotion muscle by muscle and by reliving past emotional experiences. The autonomic activity produced distinguished not only between positive and negative emotions, but also among negative emotions. This finding challenges emotion theories that have proposed autonomic activity to be undifferentiated or that have failed to address the implications of autonomic differentiation in emotion.

However, in case of human beings prefer to camouflage their expressions, using only facial expression signals may not be enough to detect emotions reliably. Compared with facial expressions, using physiological signals is a more reliable method to track and recognize emotions and internal cognitive processes of people. Physiological signals, including respiration, heart rate, galvanic skin resistance or conductivity etc have been used ,to overcome this disadvantage in emotion recognition and tracking tasks.

Traditional recommendation engines use content based or collaborative filtering methods and do not consider user emotion state. However, using human emotion state with recommendation engines may increase recommendation engines performance. Shin et al., presented an automatic stress-relieving music recommendation system. System used wireless and portable finger-type PPG sensor. Nirjon et al., proposed a context-aware, biosensor based, music recommender system for mobile phones. Liu et al., presented a music recommendation

system which is aware of user heartbeat and preference. Yoon et al. implemented personalized music recommendation system using selected features, context information and listening history.

Rosa et al. presented a music recommendation system based on a sentiment intensity metric, named enhanced Sentiment Metric (eSM) that is the association of a lexicon-based sentiment metric with a correction factor based on the users profile. The users sentiments are extracted from sentences posted on social networks and the music recommendation system is performed through a framework of low complexity for mobile devices, which suggests songs based on the current users sentiment intensity. We propose a music recommendation system which considers users emotional state in its recommendations. Systems recommendations are mostly based on two factors: users past preferences, and the possible effects of recommended songs on the user emotion. The system detects user emotion and evaluates the emotional effects feedback before and after a song is recommended. The framework uses GSR and PPG to continuously track users emotional state changes. Before current work, we have proposed an emotion recognition system based on only GSR signals. In this study we are enriching signals with PPG and propose a data fusion base emotion recognition method for music recommendation engines. Our proposed framework aims to enhance music recommendation engines performance by considering users emotion states.

2.2 Traditional recommendation engines

A recommender systems main task is to propose right products or items to a cluster of users that might be accepted by them. The design of such recommendation engines depends on the domain and the particular characteristics of the data available. Figure 2.1 describes recommender system components, working order and data flow in recommendation engine. Data collection and processing unit (DCPU, Figure 2.1, step 1), provides a suitable tool for data collection which involves users and items. DCPU sends data to Recommender Model (step 2) where recommendation algorithms are executed. Recommendation Post Processing unit (step 3), makes the recommendations ready to be shown to users after filtering out and ranking. Feedback module (step 4) used to track usage and the user interface (step 5) com-

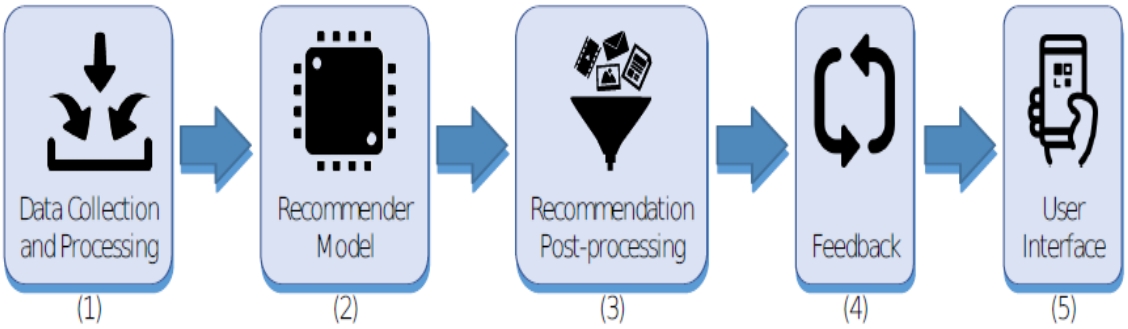


Fig. 2.1: Recommender system components and data flow.

ponent defines what users see and how they can interact with the recommender . Traditional recommendation engines use a number of different technologies to generate recommendations.

Collaborative filtering, also referred to as social filtering, filters information by using the recommendations of other people. It is based on the idea that people who agreed in their evaluation of certain items in the past are likely to agree again in the future. A person who wants to see a movie for example, might ask for recommendations from friends. The recommendations of some friends who have similar interests are trusted more than recommendations from others. This information is used in the decision on which movie to see. Collaborative filtering (Fig. 2.2A) is an approach to making recommendations by finding similarity and relation among users of a recommendation system. It presents an approach to find items of potential interest, which are not seen by the current user but have been rated by other users, and to predict the rating that the current users would give to an item. Collaborative filtering systems recommend items based on similarity between users and their history. The items that are preferred by similar users are recommended to current user.

Content-based filtering, also referred to as cognitive filtering, recommends items based on a comparison between the content of the items and a user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user.Several issues have to be considered when implementing a content-based filtering system. First, terms can either be assigned automatically or manually.

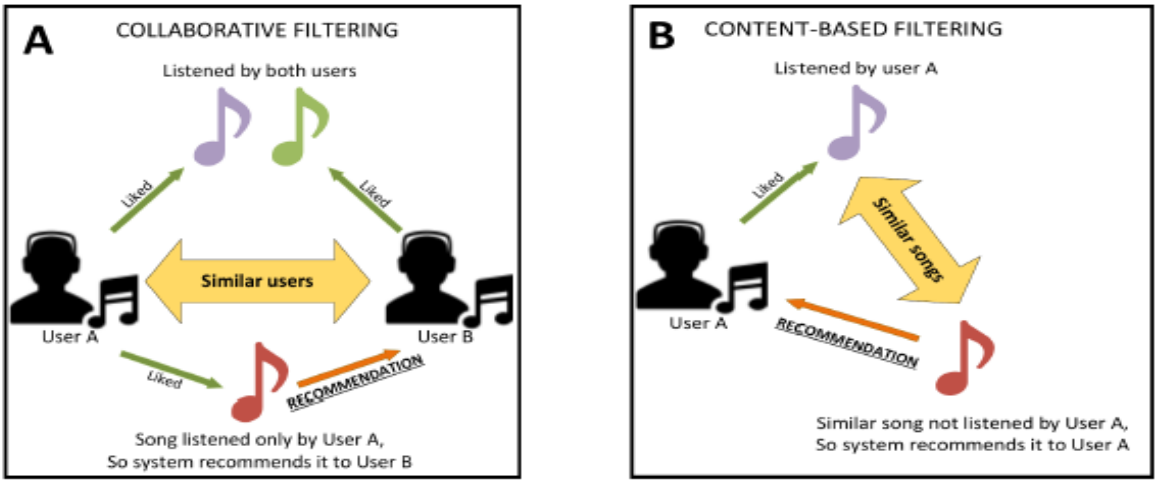


Fig. 2.2: Collaborative and Content-based filtering methods(A and B scenarios).

When terms are assigned automatically a method has to be chosen that can extract these terms from items. Second, the terms have to be represented such that both the user profile and the items can be compared in a meaningful way. Third, a learning algorithm has to be chosen that is able to learn the user profile based on seen items and can make recommendations based on this user profile.Content-based filtering methods (Fig. 2.2B) make recommendations by analyzing the properties and tags of the items that have been rated by the user and the description of items to be recommended.

The Proposed Method

Proposed framework involves using GSR and PPG to capture physiological signals from the user via a wearable computing device, and using these signals to enhance the accuracy of the recommendations made by the recommender system by tracking the users emotional state through these signals. Emotional effects of the past recommendations on the user are stored in the systems database and used in future recommendations, as the same musical tracks effects can be varied between different users. System architecture of this framework is given in Fig. 3.1. In Fig. 3.1, the green colored input elements named context, collaborative data, item profile, and user profile are traditional inputs for a music recommender system, and the blue colored input elements named plethysmography, galvanic skin response and past physiological effects are our proposed input elements to increase the accuracy of recommendations and enhance the recommendation system. Our general system flow is summarized in Algorithm 1.

Algorithm 1: Flow of the proposed algorithm.

Input: First Source Data Signal Photo plethysmography SPPG
Input: Second Source Data Signal Galvanic Skin Response SGSR
Output: Predicted Target emotion labels LE based on Arousal and Valence values
Output: Recommended Song RS

1. Get signal data from Photo plethsymograhpy SPPG and Galvanic Skin Response sensors SGSR
2. Sample and Extract features from SPPG and SGSR
3. Predict Target emotion labels LE based on Arousal and Valence values in Machine Learning Pipeline

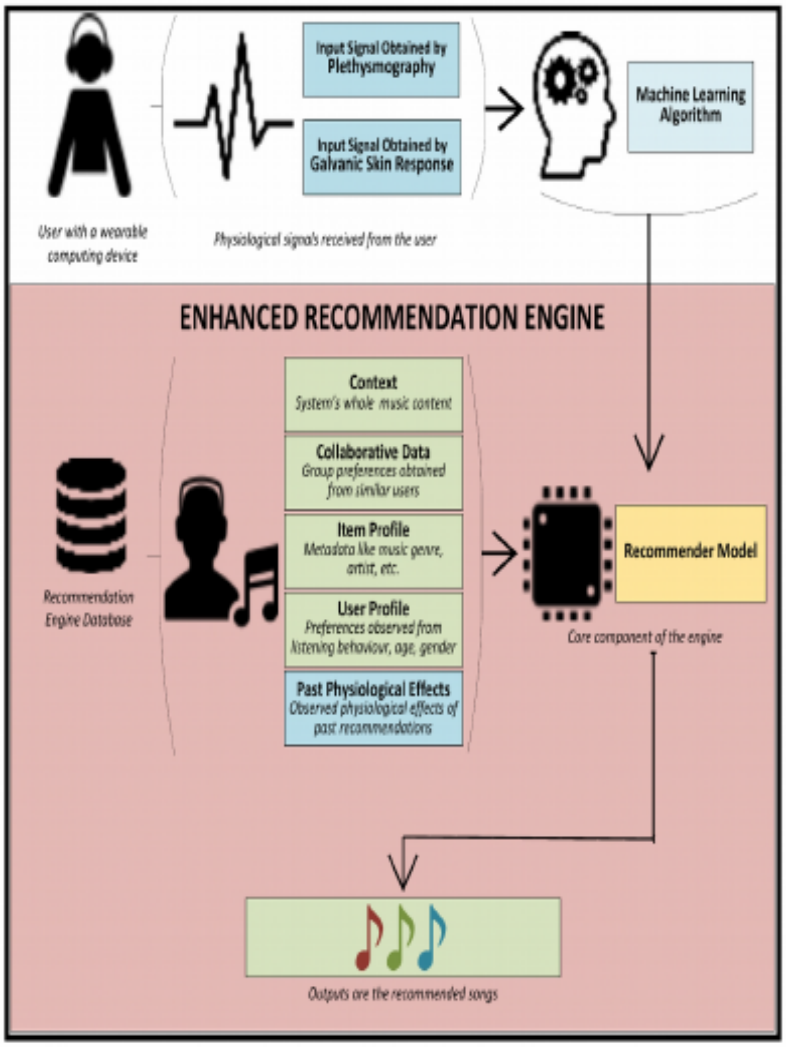


Fig. 3.1: System architecture.

4. Feed LE to enrich Recommendation Engine Decision Algorithm

5. Combine LE with User Profile, Item Profile and feed Recommendation Engine

6. Get RS and send it to player

3.1 Materials and methods

3.1.1 Emotion Recognition Using Physiological Signals

Nowadays affective computing has become the hotspot in computer science. Recording and recognizing physiologic signatures of emotion has become an increasingly important field of research in affective computing and human-computer interface . Traditional investigation, which has made considerable achievements, is based on the recording and statistical analysis of physiological signals from Autonomic nervous system . Some researchers have been doing their best to develop wearable devices, while others devoting themselves to implementing a physiological signal-based emotion recognition system . In 1999, researchers at IBM developed an emotion mouse about 75 percent successful in determining a user's emotional state . In 2001, Picard and colleagues at MIT Media Laboratory developed pattern recognition algorithms which attained 81% classification accuracy . Because of data acquired from only one subject, these emotion recognition methods can only measure one subjects emotion. In 2004, Kim and his group developed a multiple-users emotion recognition system using short-term monitoring of physiological signals. A support vector machine (SVM) was adopted as a pattern classifier, and correct-classification ratio for 50 subjects is 78.4% ˙can monitor physiological attributes of the human body that are controlled directly by autonomous nervous system. These sensors can collect signals including skin conductance, blood volume, temperature, heart rate. In this study, we have used GSR and PPG signals.

1) GSR Signals: GSR, which is also known as electro dermal activity (EDA) is a resistance or conductance based, easily captured, low cost physiological signal technique. Sometimes, one needs to control different emotional situations which can lead the person suffering them to dangerous situations, in both the medium and short term. There are studies which indicate that stress increases the risk of cardiac problems. In this study we have designed and built a stress sensor based on Galvanic Skin Response (GSR), and controlled by ZigBee. In order to check the devices performance, we have used 16 adults (eight women and eight men) who completed different tests requiring a certain degree of effort, such as mathematical operations

or breathing deeply. On completion, we appreciated that GSR is able to detect the different states of each user with a success rate of 76.56%. In GSR, hand or foot attached sensors are used to measure the electrical conductance of the skin. Experiencing emotions like stress or surprise causes changes in skin resistance. GSR is used to capture physiological reactions that generate excitement. When people get excited, body sweats, the amount of salt on the skin and skins electrical conductance changes.

Central feedback of peripheral states of arousal influences motivational behavior and decision making. The sympathetic skin conductance response (SCR) is one index of autonomic arousal. The precise functional neuroanatomy underlying generation and representation of SCR during motivational behavior is undetermined, although it is impaired by discrete brain lesions to ventromedial prefrontal cortex, anterior cingulate, and parietal lobe. We used functional magnetic resonance imaging to study brain activity associated with spontaneous fluctuations in amplitude of SCR, and activity corresponding to generation and afferent representation of discrete SCR events. Regions that covaried with increased SCR included right orbitofrontal cortex, right anterior insula, left lingual gyrus, right fusiform gyrus, and left cerebellum. At a less stringent level of significance, predicted areas in bilateral medial prefrontal cortex and right inferior parietal lobule covaried with SCR. Generation of discrete SCR events was associated with significant activity in left medial prefrontal cortex, bilateral extrastriate visual cortices, and cerebellum. Activity in right medial prefrontal cortex related to afferent representation of SCR events. Activity in bilateral medial prefrontal lobe, right orbitofrontal cortex, and bilateral extrastriate visual cortices was common to both generation and afferent representation of discrete SCR events identified in a conjunction analysis. Our results suggest that areas implicated in emotion and attention are differentially involved in generation and representation of peripheral SCR responses. We propose that this functional arrangement enables integration of adaptive bodily responses with ongoing emotional and attentional states of the organism.

2) Photo Plethysmography Signals: Plethysmography is a measurement technique that can be used to measure the volume changes in different parts of the body. In our study, change in blood volume has been measured via PPG sensor that is attached to participants thumb. Heart rate variability (HRV) and inter-beat periods measurements also can be done using PPG sensors.

Since emotions like stress may increase blood pressure, emotions have correlation with HRV and blood pressure. PPG sensors detect optical blood volume variations in the tissues microvascular bed.

Photoplethysmography (PPG) is a simple and low-cost optical technique that can be used to detect blood volume changes in the microvascular bed of tissue. It is often used non-invasively to make measurements at the skin surface. The PPG waveform comprises a pulsatile ('AC') physiological waveform attributed to cardiac synchronous changes in the blood volume with each heart beat, and is superimposed on a slowly varying ('DC') baseline with various lower frequency components attributed to respiration, sympathetic nervous system activity and thermoregulation. Although the origins of the components of the PPG signal are not fully understood, it is generally accepted that they can provide valuable information about the cardiovascular system. There has been a resurgence of interest in the technique in recent years, driven by the demand for low cost, simple and portable technology for the primary care and community based clinical settings, the wide availability of low cost and small semiconductor components, and the advancement of computer-based pulse wave analysis techniques. The PPG technology has been used in a wide range of commercially available medical devices for measuring oxygen saturation, blood pressure and cardiac output, assessing autonomic function and also detecting peripheral vascular disease. The introductory sections of the topical review describe the basic principle of operation and interaction of light with tissue, early and recent history of PPG, instrumentation, measurement protocol, and pulse wave analysis. The review then focuses on the applications of PPG in clinical physiological measurements, including clinical physiological monitoring, vascular assessment and autonomic function.

PPG system consists a detector and red/infrared light-emitting diodes used as the light source to monitor tissue light intensity variations through transmission or from reflection.Galvanic Skin Response (GSR) has recently attracted researchers' attention as a prospective physiological indicator of cognitive load and emotions. However, it has commonly been investigated through single or few measures and in one experimental scenario. In this research, aiming to perform a comprehensive study, we have assessed GSR data captured from two different experiments, one including text reading tasks and the other using arithmetic tasks, each impos-

ing multiple cognitive load levels. We have examined temporal and spectral features of GSR against different task difficulty levels. ANOVA test was applied for the statistical evaluation. Obtained results show the strong significance of the explored features, especially the spectral ones, in cognitive workload measurement in the two studied experiments.

3.1.2 Emotion Representation

Factor-analytic evidence has led most psychologists to describe affect as a set of dimensions, such as displeasure, distress, depression, excitement, and so on, with each dimension varying independently of the others. However, there is other evidence that rather than being independent, these affective dimensions are interrelated in a highly systematic fashion. The evidence suggests that these interrelationships can be represented by a spatial model in which affective concepts fall in a circle in the following order: pleasure (0), excitement (45), arousal (90), distress (135), displeasure (180), depression (225), sleepiness (270), and relaxation (315). This model was offered both as a way psychologists can represent the structure of affective experience, as assessed through self-report, and as a representation of the cognitive structure that laymen utilize in conceptualizing affect. Supportive evidence was obtained by scaling 28 emotion-denoting adjectives in 4 different ways: R. T. Ross's (1938) technique for a circular ordering of variables, a multidimensional scaling procedure based on perceived similarity among the terms, a unidimensional scaling on hypothesized pleasuredispleasure and degree-of-arousal dimensions, and a principal-components analysis of 343 Ss' self-reports of their current affective states.

Two models are common among proposed models for emotion representation by psychologists: the dimensional model and the categorical (discrete) model. According to the dimensional model, people emotions can be represented with a limited number of independent affective dimensions. Two important dimensions are Arousal, pointing out intensity of an emotion and Valence, marking the polarity of an emotion as either negative or positive. The emotion valence-arousal dimensional model (EVADM) is depicted in Figure 3.2. The pleasure

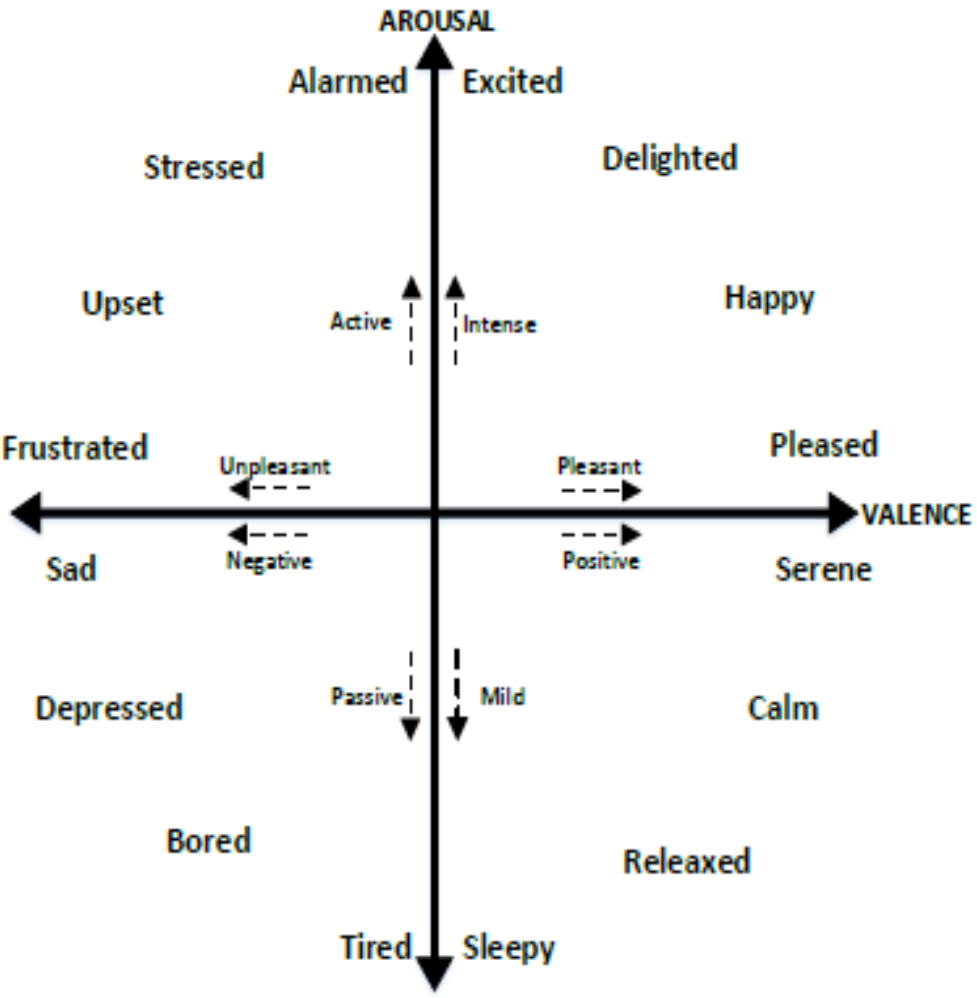


Fig. 3.2: Valence - Arousal Model..

(valence) - displeasure scale measures pleasantness degree of an human as an reaction to an external stimuli, and basically it reflects the degree of attraction of a person toward an object, event or stimuli. Emotion intensity is shown via arousal - nonarousal scale. Arousal range changes from active to passive and defined as psychological or physiological state of being awaken or passive reaction to an external stimuli.

3.1.3 Feature extraction

Feature extraction process is crucial in a machine learning pipeline and it is related to representing signals to machine learning algorithms via vectors.In machine learning, pattern recognition and in image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is a dimensionality reduction process, where an initial set of raw variables is reduced to more manageable groups (features) for processing, while still accurately and completely describing the original data set.When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named a feature vector). Determining a subset of the initial features is called feature selection. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power, also it may cause a classification algorithm to overfit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Many machine learning practitioners believe that properly optimized feature extraction is the key to effective model construction.[3]

Results can be improved using constructed sets of application-dependent features, typically built by an expert. One such process is called feature engineering. Alternatively, general dimensionality reduction techniques are used such as:

Independent component analysis, Isomap, Kernel PCA, Latent semantic analysis, Partial least squares, Principal component analysis, Multifactor dimensionality reduction, Nonlinear dimensionality reduction, Multilinear Principal Component Analysis, Multilinear subspace

learning, Semidefinite embedding, Autoencoder. In order to represent physiological signals, each signal has been divided to various length moving windows and features have been extracted in the time domain and based on statistics. After capturing GSR and PPG sensor signal data, each signal has been divided into subsignals to capture local signal information. Subsignal length has been changed and tested between one and 60 seconds to have the right resolution. Features from signals have been extracted in the time domain and based on statistics. At fist step, features have been extracted from each sub-signal, and then all sub-signal extracted features have been concatenated for each subject. This process is repeated for all subjects and for every individual video.

Various attributes have been selected as feature set and relationship between arousal and valence has been studied. Table I shows studied feature sets and their attributes. We have worked with four different feature sets FS-10, FS-14, FS-18, and FS-22. FS-14 contains FS-10 features with additional four features, FS-18 contains FS-14 features with additional four features, and FS- 22 is the richest feature set and contains FS-18 features with additional four features. Table II shows extracted feature list and their formulas. Arithmetic mean, maximum, minimum, variance, standard deviation, kurtosis coefficient, skewness coefficient, moments, median, mean energy, number of zero crossings, change in signal values have been considered as features as depicted in Table I.

TABLE I
FEATURE SETS AND ATTRIBUTES.

Feature set	Attributes
(FS-10)	Minimum, maximum, average, standard deviation, variance, skewness, kurtosis, median, zero crossings, mean energy
(FS-14)	Feature 10 set, 3rd, 4th, 5th, 6th moments
(FS-18)	Feature 14 set, mean absolute value, maximum scatter difference root mean square, mean absolute deviation
(FS-22)	Feature 18 set, 1st degree difference, 2nd degree difference 1st degree difference divided by standard deviation, 2nd degree difference divided by standard deviation

3.1.4 Data Fusion

Lahat et al. defines data fusion as the analysis of several datasets such that different datasets can interact and inform each other . Fusing various sensor data together facilitates detailed, reliable and efficient information representation.

Information about a phenomenon or a system of interest can be obtained from different types of instruments, measurement techniques, experimental setups, and other types of sources. Due to the rich characteristics of natural processes and environments, it is rare that a single acquisition method provides complete understanding thereof. The increasing availability of multiple datasets that contain information, obtained using different acquisition methods, about the same system, introduces new degrees of freedom that raise questions beyond those related to analysing each dataset separately. The foundations of modern data fusion have been

laid in the first half of the 20th century . Joint analysis of multiple datasets has since been the topic of extensive research, and earned a significant leap forward in the late 1960searly 1970s with the formulation of concepts and techniques such as multi-set canonical correlation analysis (CCA) , parallel factor analysis (PARAFAC) , and other tensor decompositions . However, until rather recently, in most cases, these data fusion methodologies were confined within the limits of psychometrics and chemometrics, the communities in which they evolved. With recent technological advances, in a growing number of domains, the availability of datasets that correspond to the same phenomenon has increased, leading to increased interest in exploiting them efficiently. Many of the providers of multi-view, multirelational, and multimodal data are associated with high-impact commercial, social, biomedical, environmental, and military applications, and thus the drive to develop new and efficient analytical methodologies is high and reaches far beyond pure academic interest. Motivations for data fusion are numerous. They include obtaining a more unified picture and global view of the system at hand; improving decision making; exploratory research; answering specific questions about the system, such as identifying common vs. distinctive elements across modalities or time; and in general, extracting knowledge from data for various purposes. However, despite the evident potential benefit, and massive work that has already been done in the field (see, for example [8]references therein), the knowledge of how to actually exploit the additional diversity that multiple datasets offer is still at its very preliminary stages. Data fusion is a challenging task for several reasons. First, the data are generated by very complex systems: biological, environmental, sociological, and psychological, to name a few, driven by numerous underlying processes that depend on a large number of variables to which we have no access. Second, due to the augmented diversity, the number, type and scope of new research questions that can be posed is potentially very large. Third, working with heterogeneous datasets such that the respective advantages of each dataset are maximally exploited, and drawbacks suppressed, is not an evident task. We elaborate on these matters in the following sections. Most of these questions have been devised only in the very recent years, and, as we show in the sequel, only a fraction of their potential has already been exploited.

Data fusion is the process of integrating multiple data sources to produce more consis-

tent, accurate, and useful information than that provided by any individual data source.Fusion of the data from two sources can yield a classifier superior to any classifiers based on dimension 1 or dimension 2 alone Data fusion processes are often categorized as low, intermediate, or high, depending on the processing stage at which fusion takes place. Low-level data fusion combines several sources of raw data to produce new raw data. The expectation is that fused data is more informative and synthetic than the original inputs.For example, sensor fusion is also known as (multi-sensor) data fusion and is a subset of information fusion.

Humans are a prime example of Data Fusion. As humans, we rely heavily on our senses such as our Vision, Smell, Taste, Voice and Physical Movement. A combination of all these senses combine on a daily basis to help us in performing most if not all tasks in our day to day lives. That in itself is a prime example of data fusion. We rely on a fusion of smelling, tasting and touching food to ensure it is edible or not. Similarly, we rely on our sight and our ability to hear and control movement of our body to walk or drive and perform most tasks in our lives. In all these cases, the Brain performs the fusion processing and controls what we need to do next. Our brain relies on a fusion of data gathered from the aforementioned senses.

Data fusion is classified as decision level and feature level. The main purpose of decision level approach is to use a set of independent classifiers to achieve higher accuracy and robustness by combining each individual classifiers results. Decision level fusion consists of fusion of classifiers or processing the classification results of prior classification stages. In feature level fusion (FLF), the feature extraction is done for each sensor data independently and then features are concatenated together. The fused feature vector is used in learning process. FLF approach facilitates to take advantage of mutual information from common sensor data [23]. We have used FLF in our study. In FLF process, we have obtained the feature vectors from both modalities (PPG and GSR).

Then classification proceeds the same as for the single modalities as depicted in Fig. 3.3. GSR and PPG sensors are used to collect physiological signals (2, 3) from the user (1). Signals are normalized (4,5) to make it possible to fuse them together in later steps. Feature extraction methods are applied to both GSR and PPG signal data (6, 7), and results are two different feature vector sets. Then, two different feature sets are fused together to form a

single feature (8). Classifier takes this single feature vector as input (9) and made a prediction about the emotional state of the user (10) by estimating arousal and valence values.

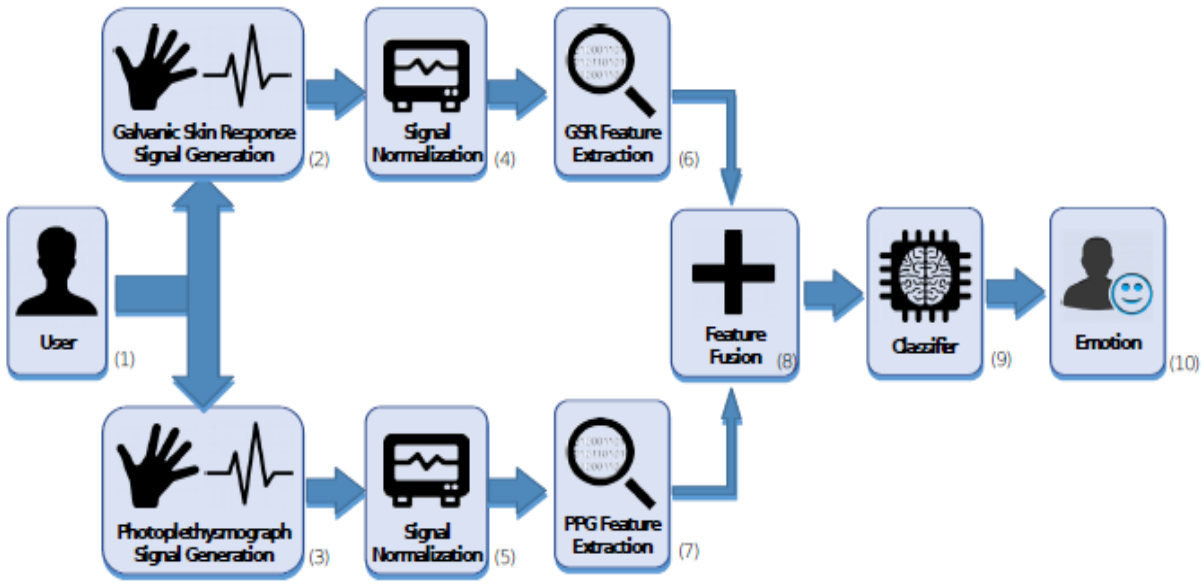


Fig. 3.3: Emotion recognition framework with GSR and PPG.

3.1.5 Classification

Subjects enter valence and arousal values which is between 1 and 9 during learning phase. These records are used during training and model creation phase. Since we have labeled data our study turns to supervised classification problem. For each subject and for each watched video we assign class value based on each videos rating value (low: 4.5, high: 4.5). Signal data captured from 32 subjects have been used for training and test steps. In order to find the appropriate classifier three classification algorithms including random forest, kNN and decision tree have been used after feature extraction phase. Random forests algorithm is ensemble based approach using decision tree forests [8] . Random forests may achieve high accuracy in a variety of problems, making them versatile choice for many applications. Since only a subset of the features used, random forests capable of handling high dimensional data. In addition, a trained model can be used to determine the pairwise proximity between samples.

TABLE II
BASIC FEATURES AND FORMULAS USED.

Attribute	Formula	Attribute	Formula
Min	$\min\{X_n\}$	Skewness	$\sum_{n=1}^{N}(X_n - AM)\frac{3}{(N-1)SD^3}$
Max	$\max\{X_n\}$	Kurtosis	$\sum_{n=1}^{N}(X_n - AM)\frac{4}{(N-1)SD^4}$
Arithmetic mean (μ)	$\frac{1}{N}\sum_{n=1}^{N}X_n$	Median	$\frac{(\frac{N}{2})^{th}+(\frac{N}{2}+1)^{th}}{2}$ or $(\frac{N+1}{2})^{th}$
Mean Absolute	$\frac{1}{N}\sum_{n=1}^{N}\lvert X_n\rvert$	Moment (kth order)	$\frac{1}{N}\sum_{n=1}^{N}X_n^k$
Root Mean Square	$\sqrt{\frac{1}{N}\sum_{n=1}^{N}X_n^2}$	First Degree Difference	$\frac{1}{N-1}\sum_{n=1}^{N}\lvert X_{n+1} - X_n\rvert$
Standard Deviation (SD)	$\sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n - AM)^2}$	Second Degree Difference	$\frac{1}{N-2}\sum_{n=1}^{N}\lvert X_{n+2} - X_n\rvert$

3.2 Experiments and results

3.2.1 Dataset

mExperiments have been performed on the multimodal DEAP emotion database which consists PPG, GSR and electroencephalogram (EEG) signals of 32 participants during video watching. Each participant watches 40 one-minute length videos and label an arousal and valence value at the end of video between 1 and 9. The dataset was first presented by Kolestra et al. [7] . The data was down sampled to 128Hz, EOG artefacts were removed, a band pass frequency filter from 4.0 - 45.0Hz was applied.

3.2.2 System Configuration Parameters Selection

In this section, we have evaluated the effect of three important system parameters (window duration size, feature set size, convolution / non-convolution) and classifiers on emotion prediction accuracy rate. We first tried various duration of window sizes to evaluate their on accuracy. Each physiological signal has been divided into windows with different window duration W 2 1;3;5;8;10;12;15;30;60 seconds. Afterwards features extracted from signals have been studied and various feature set size F 2 10;14;18;22 have been tested. We have also eval-

uated the effect of convolution. Hyperparameter tests have been conducted with 10 - fold cross validation.

1) Effects of Window Duration Size and Convolution: Window duration affects accuracy rate. Various window size duration between 1 seconds and 60 seconds have been selected. Tests with 3 seconds window duration performed better than other window duration sized for GSR and 8 seconds duration sizes performed better than the rest for PPG.Windows were slided by collapse (convolution) or not collapse (nonconvolution) manner. Overlapped and one second slide duration performed better compared to non-overlapping window sliding. Fig. 3.4 confirms that convolution is a better approach to increase accuracy rate (see Fig. 3.4).

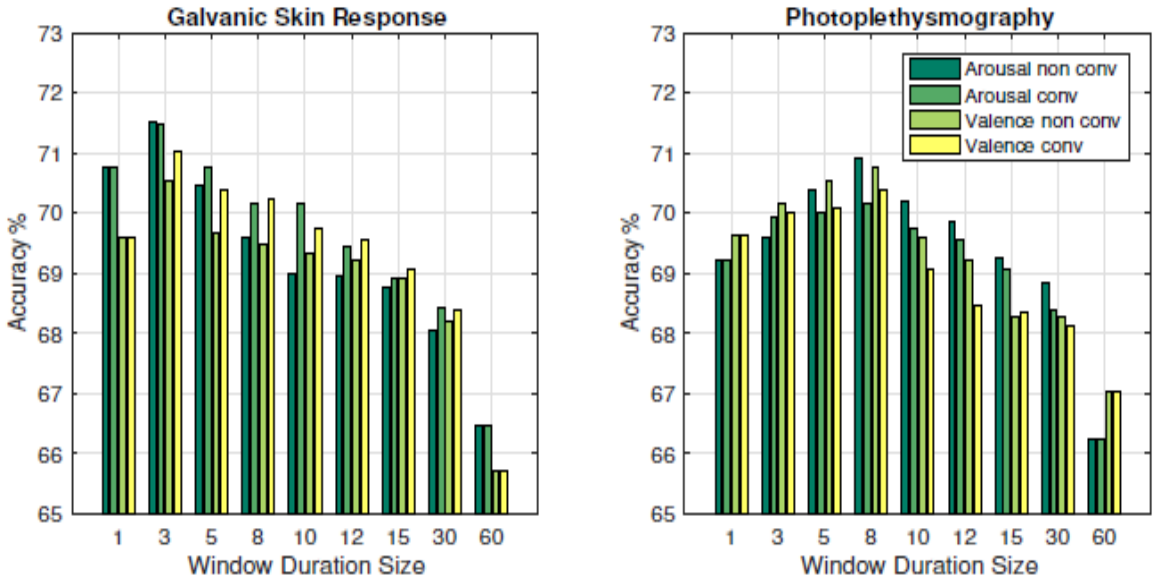


Fig. 3.4: Window Duration and Convolution effect for GSR and PPG.

2) Feature Set Tests: Feature extraction also affects system accuracy. Various feature sets (FS) have been selected. Tests with FS 10, FS 14, FS 18 and FS 22 were conducted. For GSR FS 14, and for PPG FS-10 set performed better than the other feature sets . Results are depicted in Figure 3.5.

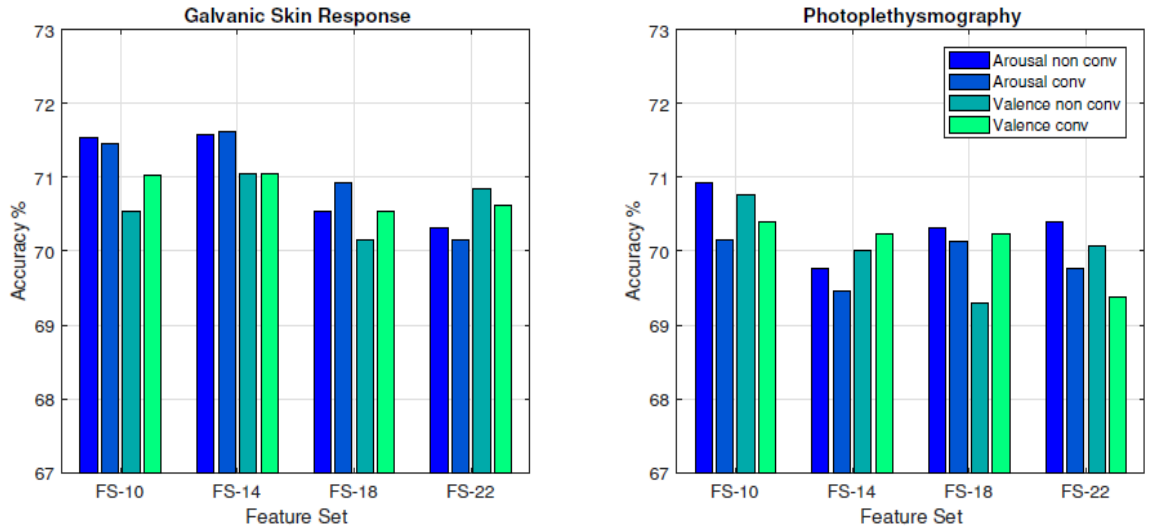


Fig. 3.5: Feature Set and Convolution effect for GSR and PPG

3.2.3 Data Fusion Based Test Results

We fused GSR and PPG feature vectors. Tests have been done with 10-fold cross validation (10F-CV) by using various classifiers C 2 DecisionTree(J48), Random Forests (RF), k - Nearest Neighbor(kNN), Support Vector Machine (SVM) and feature set size F 2 10;14;18;22 configurations. For the arousal, best accuracy rate was obtained with Feature Set-22 and Classifier RF 72.0618 also gave very close results for arousal as shown in Figure 3.6. For the valence, best accuracy rate was obtained with Feature Set-22 and Classifier RF with 71.053.6).

3.2.4 Discussion

In this section, test case results and potential consumer use cases are discussed. Generally, recognizing arousal and valence values directly from bio sensors individually is a challenge task. We have showed that there is relationship between GSR, PPG signals and arousal and valence.

A. Test Results Discussion Results encourage us to use these signals in emotion recognition pipeline. There is high correlation between GSR, PPG signals and emotion. Results also

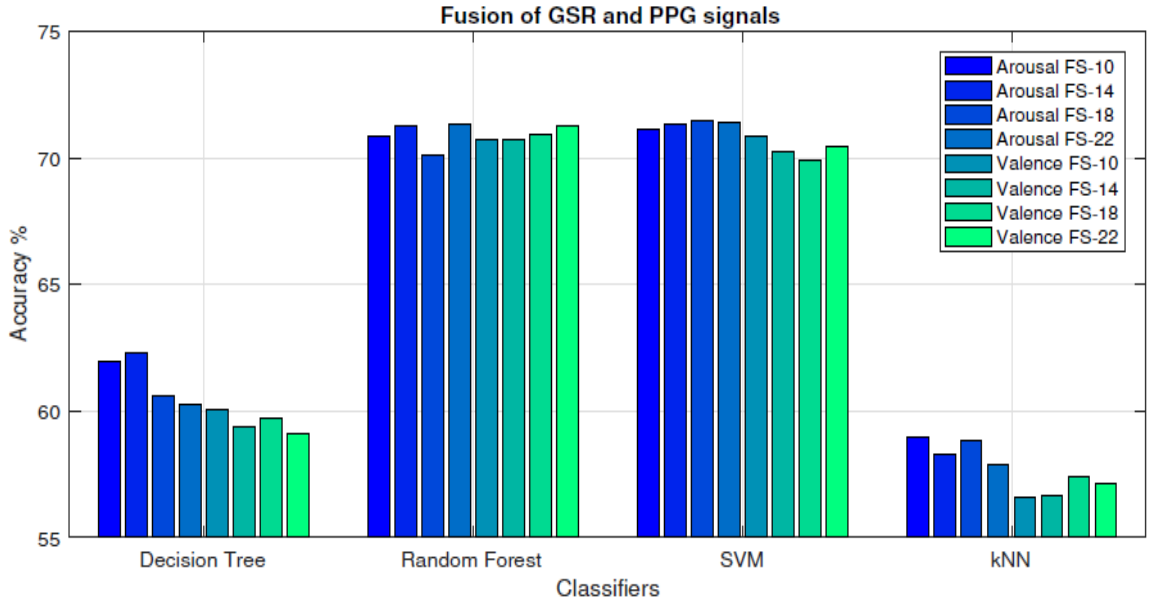


Fig. 3.6: Fusion of GSR and PPG signals.

revealed that multi modality may help to increase accuracy rate slightly compared to single modality as depicted in Figure 3.7. For GSR only, we obtained 71.53prediction respectively. For PPG we have obtained 70.92and 70.76respectively. Fusing GSR and PPG signals we have obtained slightly better results especially for arousal. Feature fusion of multi modalities slightly increased in the accuracy rate. To the best of our knowledge, Galvanic Skin Response and Photo Plethysmography signals have never been fused as we proposed in this work for the used dataset. Using more than one sensor in proposed fusion manner has potential to increase accuracy performance and robustness.

B. Use Cases Discussion Music recommender systems are recently seeing a sharp increase in popularity due to many new commercial music streaming services. Most systems, however, do not decently take their listeners into account when recommending music items. We have proposed a framework using machine-learning scheme in conjunction with wearable sensors for translating end-consumers subjective experience of music into arousal and valence scores that can be used in popular on-demand music streaming services. Proposed framework can be used by a mobile device for music recommendation to the consumer of the mobile device.

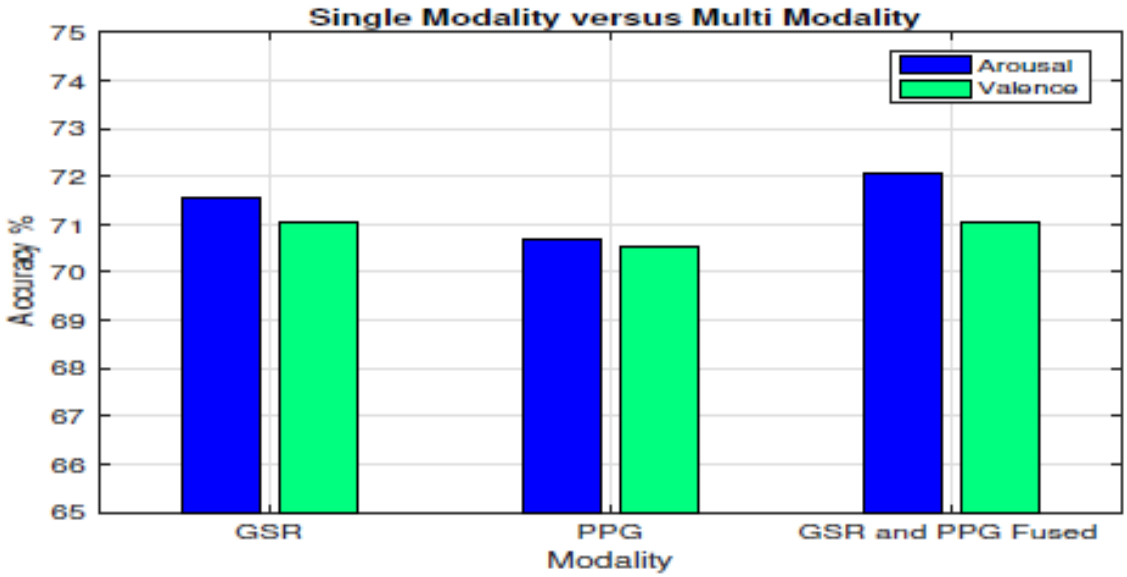


Fig. 3.7: Comparison of Accuracy for Single and Multi Modality Approaches.

Streaming service and/or mobile device can use framework for automatic play list generation and next best song offer based on demographics as well as physiological signals.

Conclusion

In this study, a framework for enhancing music recommendation engines performance via physiological signals has been introduced. Emotion recognition from multi-channel physiological signals was performed, data fusion techniques were applied to combine data from GSR and PPG sensors and FLF has been implemented. Considering emotion state of the listener improves the performance of recommendations. Recognizing arousal and valence values directly from only GSR and PPG signals is a challenging task. We have showed that there is relationship between GSR and PPG signals and emotional arousal and valence dimensions. For GSR only signal, we have obtained 71.53% and 71.04% accuracy rate for arousal and valence prediction respectively. For photoplehysmography only signal, we have obtained 70.93% and 70.76% accuracy rate for arousal and valence prediction respectively. Fusing GSR and PPG signals we have obtained the results, 72.06% and 71.05% accuracy rate for arousal and valence prediction respectively. Although there is only slight improvement using fusion in emotion recognition accuracy, the proposed framework is promising for music recommendation engines in terms of adding multi modal emotion phenomenon into music recommendation logic. Performance can be improved with the advancement of wearable sensor technologies and using different type of sensors. Using more than one sensor may also help for failure management. As future work, we will consider different combination of sensors that handle the failures of wearable sensors and additional sensors usage to increase performance. The results of this study can be used to increase user experience of multimedia tools and music recommendation engines. Since there is high correlation between physiological GSR and PPG data and affective state and cognitive state of a person multimedia recommendation engines can benefit from physiological computing systems.

REFERENCES

[1] R. D. Blumofe and C. E. Leiserson, Scheduling multithreaded computations by work stealing, J. ACM, vol. 46, no. 5, pp. 720748, Sep. 1999.

[2] D. Eager, J. Zahorjan, and E. Lazowska, Speedup versus efficiency in parallel systems, IEEE Transactions on Computers, vol. 38, no. 3, pp.408423, Mar. 1989.

[3] S. Jhajharia, S. Pal, and S. Verma, Wearable computing and its application, Int. J. Comp. Sci. and Inf. Tech., vol. 5, no. 4, pp. 5700 5704, 2014.

[4] K. Popat and P. Sharma, Wearable computer applications: A feature perspective, Int. J. Eng. and Innov. Tech., vol. 3, no. 1, 2013.

[5] I.-h. Shin, J. Cha, G. W. Cheon, C. Lee, S. Y. Lee, H.-J. Yoon, and H. C. Kim, Automatic stress-relieving music recommendation system based on photoplethysmography-derived heart rate variability analysis, in IEEE Int. Conf. on Eng. in Med. and Bio. Soc. IEEE, 2014, pp. 64026405..

[6] H. Liu, J. Hu, and M. Rauterberg, Music playlist recommendation based on user heartbeat and music preference, in Int. Conf. on Comp. Tech. and Dev., vol. 1. IEEE, 2009, pp. 545549.

[7] Y. Wang, F. Wu, J. Song, X. Li, Y. Zhuang, Multi-modal mutual topic reinforce modeling for cross-media retrieval, in: Proceedings of the ACM International Conference on Multimedia, 2014, pp. 307316.

[8] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo, Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks, in Proc. of Aust. Comp. Hum. Inter. ACM, 2012, pp. 420423.

[9] A. Nakasone, H. Prendinger, and M. Ishizuka, Emotion recognition from electromyography and skin conductance, in Proc. of Int. Work. on Biosignal Interp., 2005, pp. 219222.

[10] P. Ekman, R. W. Levenson, and W. V. Friesen, Autonomic nervous system activity distinguishes among emotions. Am. Assoc. for Adv. of Sci., 1983.

[11] S. Planet and I. Iriondo, Comparison between decision-level and feature-level fusion of acoustic and linguistic features for spontaneous emotion recognition, in Iberian Conf. on Inf. Sys. and Tech. IEEE, 2012, pp. 16.

[12] T. K. Ho, Random decision forests, in Proc. of Int. Conf. on Doc. Analy. and Recog., vol. 1. IEEE, 1995, pp. 278282.