Emotion Based Music Recommendation System Using Wearable Physiological Sensors

Deger Ayata[ORCID], Yusuf Yaslan, and Mustafa E. Kamasak

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. This paper proposes an emotion based music recommendation framework 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 feed to any collaborative or content based recommendation engine as a supplementary data. Thus, existing recommendation engine performances can be increased using these data. Therefore, in this paper emotion recognition problem is considered as arousal and valence prediction from multi-channel physiological signals. Experimental results are obtained on 32 subjects' GSR and PPG signal data with/out feature fusion using decision tree, random forest, support vector machine and k-nearest neighbors algorithms. 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.

Index Terms—Emotion aware recommendation engine, emotion recognition, galvanic skin response, machine learning, physiological signals, photo plethysmography.

I. 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 [1], [2].

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 user's choices and produce the outcomes that co-relates to their needs and interests [3]. 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 [4], [5] 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 system's 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 [6]. The proposed wearable attached music recommendation framework utilizes not only the user's 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.

II. RELATED WORKS

Ekman *et al.* [7] 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. 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 / conductivity etc have been used [5], [8]–[10] 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 [11], [12]. However, using human emotion state
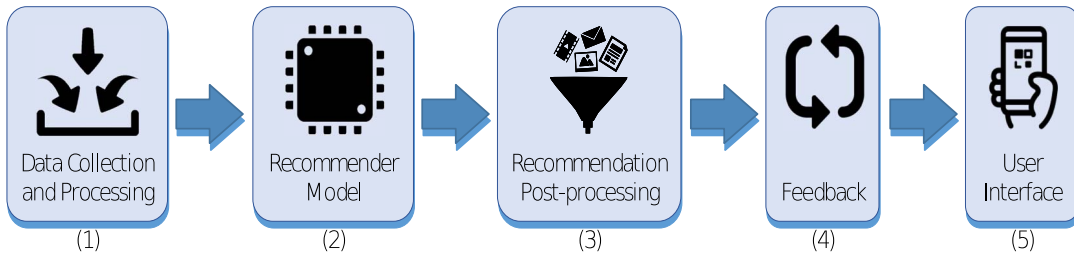
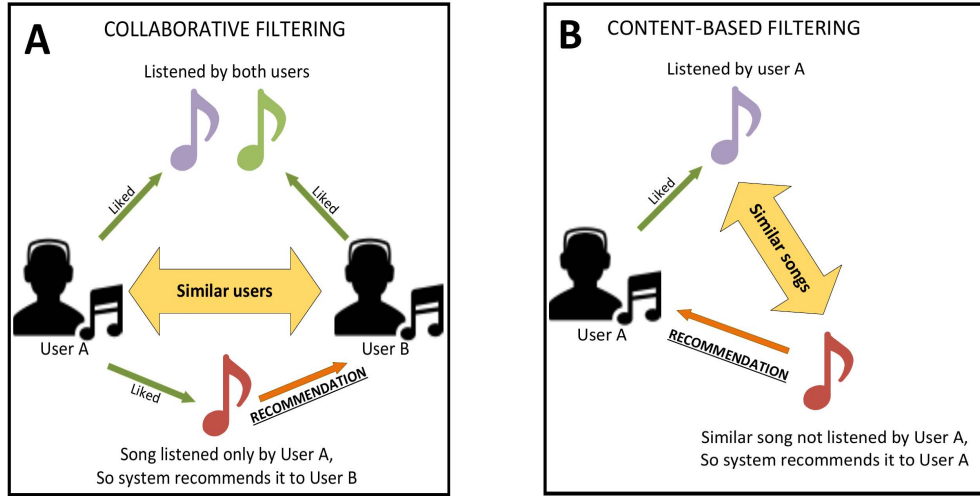Fig. 1. Recommender system components and data flow.

Fig. 2. Collaborative and Content-based filtering methods(A and B scenarios).

with recommendation engines may increase recommendation engines performance. Shin *et al.* [8] presented an automatic stress-relieving music recommendation system. System used wireless and portable finger-type PPG sensor. Nirjon *et al.* [9] proposed a context-aware, biosensor—based, music recommender system for mobile phones. Liu *et al.* [10] presented a music recommendation system which is aware of user heartbeat and preference. Yoon *et al.* [13] implemented personalized music recommendation system using selected features, context information and listening history. Rosa *et al.* [14] 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 user's 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 user's sentiment intensity. We propose a music recommendation system which considers user's emotional state in its recommendations. System's recommendations are mostly based on two factors: user's past preferences, and the possible effects of recommended songs' on the user emotion. The system detects user emotion and evaluates the emotional effect's feedback before and after a song is recommended. The framework uses GSR and PPG to continuously track user's emotional state changes. Before current work, we have proposed an emotion recognition system based on only GSR signals [6]. 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.

III. TRADITIONAL RECOMMENDATION ENGINES AND PROPOSED APPROACH

A. Traditional Recommendation

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 [11]. Figure 1 describes recommender system components, working order and data flow in recommendation engine. Data collection and processing unit (DCPU, Figure 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) component defines what users see and how they can interact with the recommender [15]. Traditional recommendation engines use a number of different technologies to generate recommendations. Collaborative filtering (Figure 2A) is an approach to making recommendations
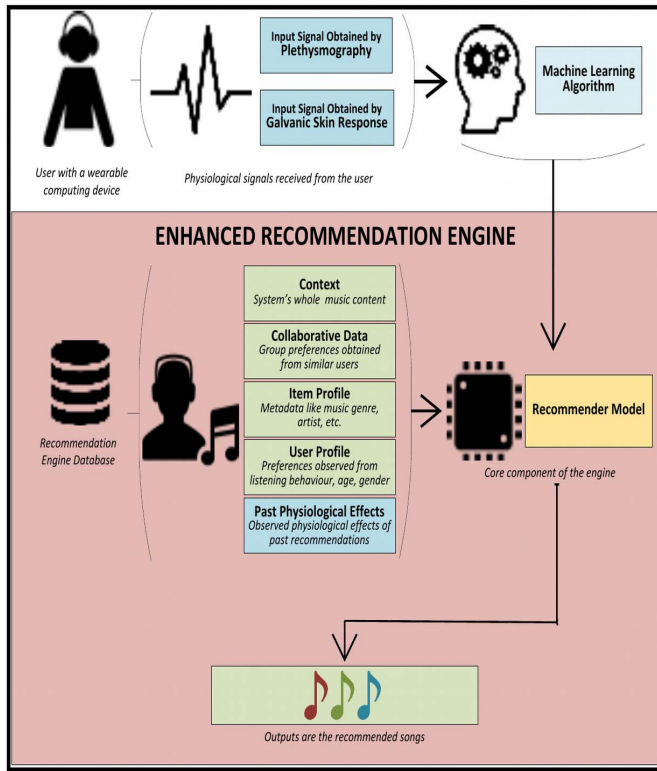
Fig. 3.   System architecture.

Algorithm 1: Flow of the Proposed Algorithm

Input: First Source Data Signal Photo plethysmography S_{PPG}

Input: Second Source Data Signal Galvanic Skin Response S_{GSR}

Output: Predicted Target emotion labels L_E based on Arousal and Valence values

Output: Recommended Song R_S

1 Get signal data from Photo plethsymograhpy S_{PPG} and Galvanic Skin Response sensors S_{GSR}

2 Sample and Extract features from S_{PPG} and S_{GSR} (described in Section IV)

3 Predict Target emotion labels L_E based on Arousal and Valence values in Machine Learning Pipeline (described in Section IV)

4 Feed L_E to enrich Recommendation Engine Decision Algorithm

5 Combine L_E with User Profile, Item Profile and feed Recommendation Engine

6 Get R_S and send it to player

---

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 methods (Figure 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.

B. Proposed Approach

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 user's emotional state through these signals. Emotional effects of the past recommendations on the user are stored in the system's database and used in future recommendations, as the same musical track's effects can be varied between different users. System architecture of this framework is given in Figure 3. In Figure 3, 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.

IV. MATERIALS AND METHODS

A. Emotion Recognition Using Physiological Signals

Biosensors 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 / conductance based, easily captured, low cost physiological signal technique. 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 skin's electrical conductance changes.

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 participant's 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' micro-vascular bed. 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 [12], [16].

B. Emotion Representation

Two models are common among proposed models for emotion representation by psychologists: the dimensional model
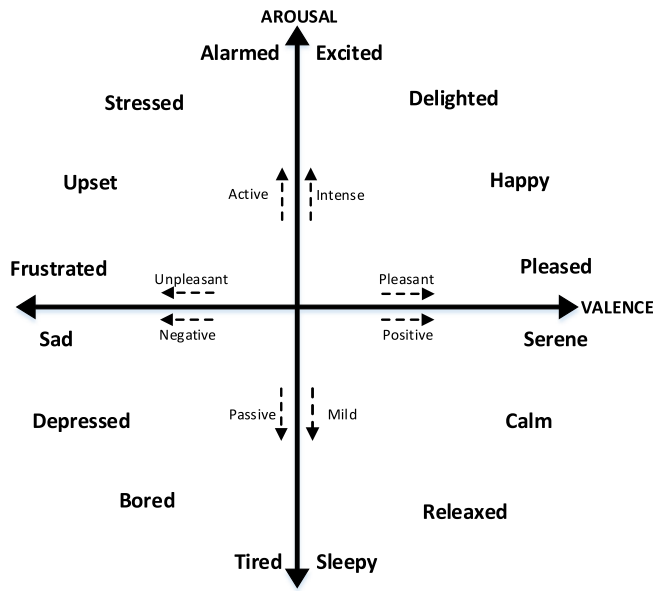
Fig. 4.   Valence - Arousal Model.

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

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 [17]–[21].

The emotion valence-arousal dimensional model (EVADM) is depicted in Figure 4. The pleasure (valence) - displeasure scale measures pleasantness degree of an human as an reaction to an external stimuli [18], [21] 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 a external stimuli.

C. 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 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. Sub-signal 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.

D. Data Fusion

Lahat *et al.* [22] 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.

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 Figure 5. 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.

TABLE II
BASIC FEATURES AND FORMULAS USED

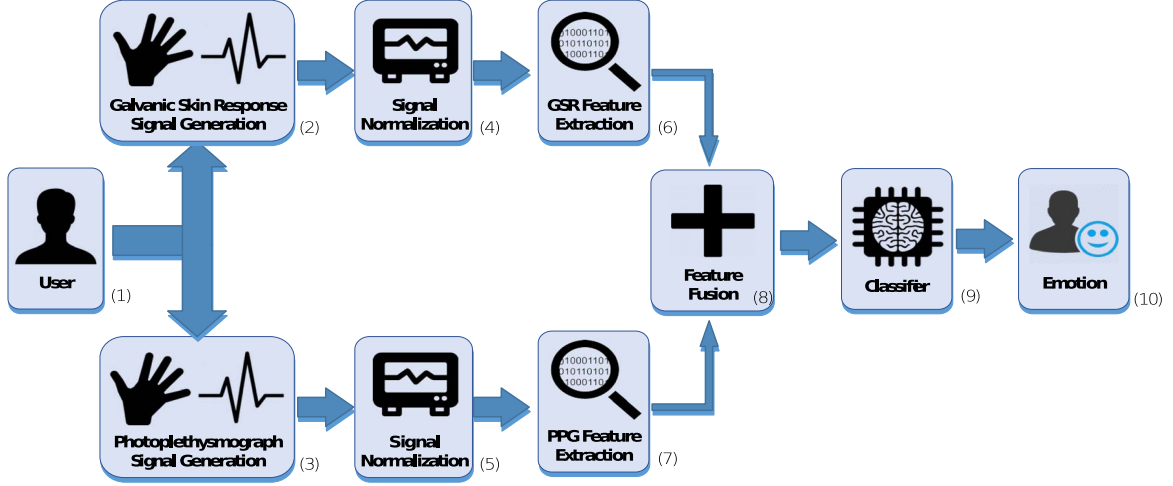| Attribute	Formula	Attribute	Formula
Min	$\min\{X_n\}$	Skewness	$\sum_{n=1}^{N}(X_n - AM)\frac{3}{(N-1)SD^3}$
Max	$\max\{X_n\}$	Kurtosis	$\sum_{n=1}^{N}(X_n - AM)\frac{4}{(N-1)SD^4}$
Arithmetic mean (μ)	$\frac{1}{N}\sum_{n=1}^{N} X_n$	Median	$\frac{(\frac{N}{2})^{\text{th}}+(\frac{N}{2}+1)^{\text{th}}}{2}$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

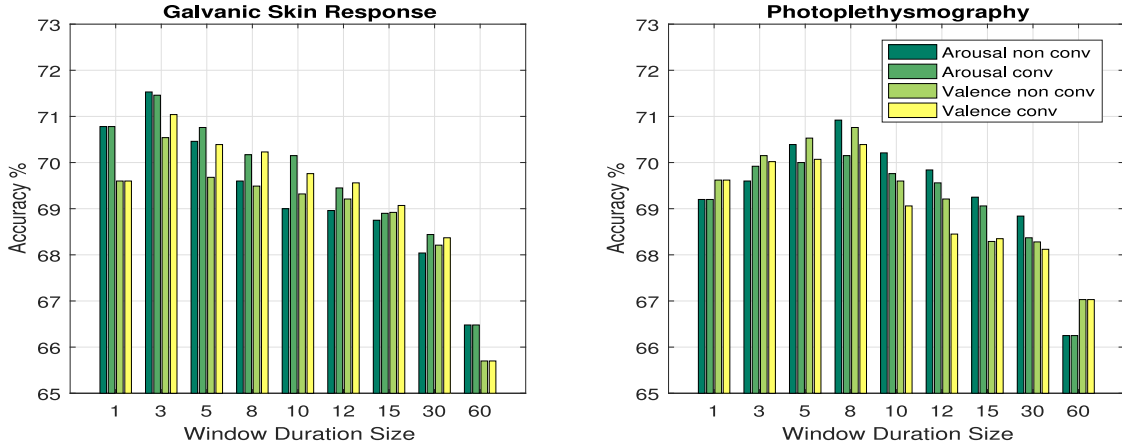Fig. 5.   Emotion recognition framework with GSR and PPG.

Fig. 6.   Window Duration and Convolution effect for GSR and PPG.

E. 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 [24]. 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.
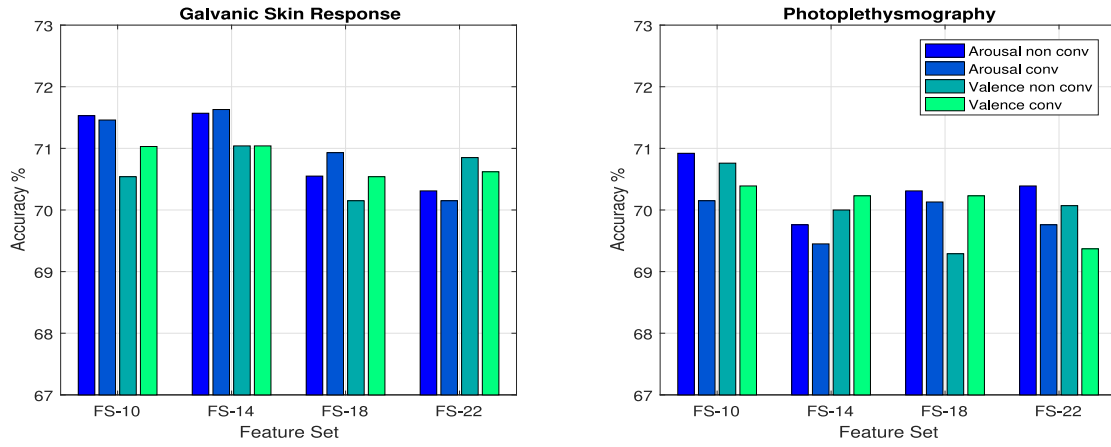
Fig. 7. Feature Set and Convolution effect for GSR and PPG.

V. EXPERIMENTS AND RESULTS

A. Dataset

Experiments 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.* [25]. The data was down sampled to 128Hz, EOG artefacts were removed, a band pass frequency filter from 4.0 - 45.0Hz was applied.

B. 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 $W \in$ 1, 3, 5, 8, 10, 12, 15, 30, 60 seconds. Afterwards features extracted from signals have been studied and various feature set size $F \in$ 10, 14, 18, 22 have been tested. We have also evaluated 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 (non-convolution) manner. Overlapped and one second slide duration performed better compared to non-overlapping window sliding. Figure 6 confirms that convolution is a better approach to increase accuracy rate (see Figure 6).

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 7.

C. 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 $C \in DecisionTree(J48)$, Random Forests (RF), k - Nearest Neighbor(kNN), Support Vector Machine (SVM) and feature set size $F \in$ 10, 14, 18, 22 configurations.

For the arousal, best accuracy rate was obtained with Feature Set-22 and Classifier RF 72.06%. SVM and FS-18 also gave very close results for arousal as shown in Figure 8.

For the valence, best accuracy rate was obtained with Feature Set-22 and Classifier RF with 71.05% (see Figure 8).

VI. 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 revealed that multi modality may help to increase accuracy rate slightly compared to single modality as depicted in Figure 9. For GSR only, we obtained 71.53% and 71.04% accuracy rate for arousal and valence prediction respectively. For PPG we have obtained 70.92% and 70.76% accuracy rate for arousal and valence prediction respectively. 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.
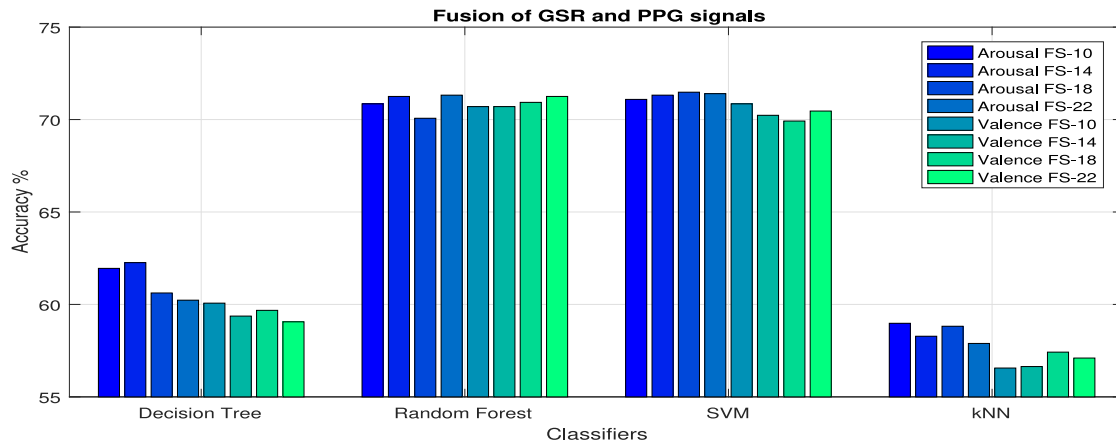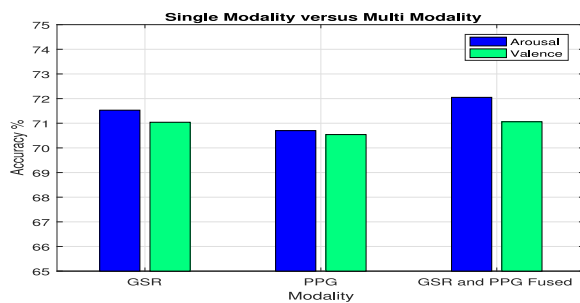
Fig. 8.   Fusion of GSR and PPG signals.

Fig. 9.   Comparison of Accuracy for Single and Multi Modality Approaches.

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-consumer's 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. 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.

VII. 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 photo-plehysmography 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] S. Jhajharia, S. Pal, and S. Verma, "Wearable computing and its application," *Int. J. Comput. Sci. Inf. Tech.*, vol. 5, no. 4, pp. 5700–5704, 2014.

[2] K. A. Popat and P. Sharma, "Wearable computer applications: A future perspective," *Int. J. Eng. Innov. Technol.*, vol. 3, no. 1, pp. 213–217, 2013.

[3] P. Melville and V. Sindhwani, "Recommender systems," in *Encyclopedia of Machine Learning*. New York, NY, USA: Springer, 2011, pp. 829–838.

[4] N. Sebe *et al.*, "Multimodal emotion recognition," in *Handbook of Pattern Recognition and Computer Vision*, vol. 4. Singapore: World Sci., 2005, pp. 387–419.

[5] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[6] D. Ayata, Y. Yaslan, and M. Kamasak, "Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods," *IU J. Elect. Elect. Eng.*, vol. 17, no. 1, pp. 3129–3136, 2017.

[7] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," *Amer. Assoc. Adv. Sci.*, vol. 221, no. 4616, pp. 1208–1210, 1983.

[8] I.-H. Shin *et al.*, "Automatic stress-relieving music recommendation system based on photoplethysmography-derived heart rate variability analysis," in *Proc. IEEE Int. Conf. Eng. Med. Bio. Soc.*, Chicago, IL, USA: IEEE, 2014, pp. 6402–6405.

[9] S. Nirjon *et al.*, "MusicalHeart: A hearty way of listening to music," in *Proc. ACM Conf. Embedded Netw. Sensor Syst.*, Toronto, ON, Canada: ACM, 2012, pp. 43–56.

[10] H. Liu, J. Hu, and M. Rauterberg, "Music playlist recommendation based on user heartbeat and music preference," in *Proc. Int. Conf. Comput. Technol. Develop.*, vol. 1. Kota Kinabalu, Malaysia: IEEE, 2009, pp. 545–549.

[11] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egypt. Informat. J.*, vol. 16, no. 3, pp. 261–273, 2015.

[12] A. Nakasone, H. Prendinger, and M. Ishizuka, "Emotion recognition from electromyography and skin conductance," in *Proc. Int. Workshop Biosignal Interpretation*, 2005, pp. 219–222.

[13] K. Yoon, J. Lee, and M.-U. Kim, "Music recommendation system using emotion triggering low-level features," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 612–618, May 2012.

[14] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," *IEEE Trans. Consum. Electron.*, vol. 61, no. 3, pp. 359–367, Aug. 2015.

[15] J. B. Schafer, "DynamicLens: A dynamic user-interface for a meta-recommendation system," in *Proc. Workshop Beyond Pers.*, 2005, pp. 72–76.

[16] 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. Aust. Comput.-Human Interaction Conf.*, Melbourne, VIC, Australia: ACM, 2012, pp. 420–423.

[17] J. A. Russell, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.

[18] A. Mehrabian, *Basic Dimensions for a General Psychological Theory Implications for Personality, Social, Environmental, and Developmental Studies*. Cambridge, MA, USA: Oelgeschlager, Gunn & Hain, 1980.

[19] P. J. Lang, M. Bradley, P. Margaret, and B. N. Cuthbert, "International affective picture system (IAPS): Instruction manual and affective ratings," Center Res. Psychophysiol., Univ. Florida, Gainesville, FL, USA, Rep. A-6, 2005.

[20] A. Mehrabian, "Communication without words," *Psychol. Today*, vol. 2, no. 4, pp. 53–56, 1968.

[21] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychol. Rev.*, vol. 110, no. 1, pp. 145–172, 2003.

[22] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.

[23] S. Planet and I. Iriondo, "Comparison between decision-level and feature-level fusion of acoustic and linguistic features for spontaneous emotion recognition," in *Proc. Iber. Conf. Inf. Syst. Technol.*, Madrid, Spain: IEEE, 2012, pp. 1–6.

[24] T. K. Ho, "Random decision forests," in *Proc. Int. Conf. Doc. Anal. Recognit.*, vol. 1. Montreal, QC, Canada: IEEE, 1995, pp. 278–282.

[25] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

Deger Ayata received the B.Sc. degree in electronics and telecommunication engineering, the first M.Sc. degree in computer and control engineering from Istanbul Technical University, and the second M.Sc. degree in electrical and electronics engineering and the M.B.A. degree from Bogazici University, Istanbul, Turkey. He is currently pursuing the Ph.D. degree with Computer Engineering Department, Istanbul Technical University. As a Technology, Research and Development Director, he managed projects including big data and analytics, data science, artificial intelligence, machine learning, IoT, and large scale software integration programs. His research interests are machine learning, pattern recognition, signal processing, computer vision, natural language processing, and artificial intelligence theory and applications.

Yusuf Yaslan received the B.Sc. degree in computer science engineering from Istanbul University, Turkey, in 2001 and the M.Sc. degree in telecommunication engineering and the Ph.D. degree in computer engineering from Istanbul Technical University (ITU), in 2004 and 2011, respectively, where he is an Assistant Professor with the Computer Engineering Department. In 2002, he joined the Multimedia Signal Processing and Pattern Recognition Laboratory, ITU. From 2001 to 2002, he was an Intern with FGAN-FOM Research Institute, Germany. He was a Visiting Researcher with Statistical Machine Learning and Bioinformatics Group, Aalto University, Finland, from 2007 to 2008. His research interests are watermarking, machine learning theory and applications, especially in semi supervised learning, social networks, music recognition/recommendation, and data mining applications on IoT data and smart cities.

Mustafa E. Kamasak received the B.S.E.E. and M.S.E.E. degrees from Bogazici University, Istanbul, Turkey, in 1997 and 1999, respectively, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA, in 2005. He is currently a Professor with the Department of Computer Engineering, Istanbul Technical University, Istanbul. His research interests include medical informatics, medical imaging, and image processing.