Classification of Human Emotions from Physiological signals using Machine Learning Algorithms

Recognition of Pain, Boredom, and Surprise Emotions

Eun-Hye Jang, Byoung-Jun Park, Sang-Hyeob Kim, Myoung-Ae Chung
BioHeath IT Convergence Research Department, IT Convergence Services Future Technology Research Department
Electronics and Telecommunications Research Institute
Daejeon, S. Korea
cleta4u@etri.re.kr; bj_park@etri.re.kr; shk1028@etri.re.kr; machung@etri.re.kr

Mi-Sook Park, Jin-Hun Sohn
Department of Psychology/Brain Research Institute
Chungnam National University
Daejeon, S. Korea
peaceful1011@nate.com ; jhsohn@cnu.ac.kr

***Abstract*—Emotion recognition is one of the key steps towards emotional intelligence in advanced human-machine interaction. Recently, emotion recognition using physiological signals has been performed by various machine learning algorithms as physiological signals are important for emotion recognition abilities of human-computer systems. The purpose of this study is to classify three different emotional states (boredom, pain, and surprise) from physiological signals using several machine learning algorithms and to identify the optimal algorithms being able to classify these emotions. 217 subjects participated in this experiment. The emotional stimuli designed to induce three emotions (boredom, pain, and surprise) were presented to subjects and physiological signals were measured for 1 minute as baseline and for 1-1.5 minutes during emotional states. The obtained signals were analyzed for 30 seconds from the baseline and the emotional state and 27 parameters were extracted from these signals. For classification of three different emotions, machine learning algorithms of Decision tree, k-NN (k-nearest neighbor algorithm), LDA (linear discriminant analysis), and SVM (support vector machine) were done by using the difference values of signal parameters subtracting baseline from the emotional state. Classification accuracy using LDA was 74.9% and the result of emotion recognition using Decision Tree showed that accuracy to recognize all emotions was 67.8%. In analysis of k-NN and SVM, classification accuracy was 62.0%. The result of emotion recognition shows that LDA is the best algorithm being able to classify pain, surprise, and boredom emotions. This led to better chance to recognize other emotions except human basic emotions and to assist more accurate and greater understanding on emotional interactions between man and machine based on physiological signals.**

Keywords-emotion; pain; surprise; boredom; physiological signals; machine learning algorithm

I. INTRODUCTION

Recently, in an attempt to categorize and understand human emotions, psychologists and engineers have tried to analyze various modalities such as facial expressions, voices, gestures, and physiological signals [1]. In particular, various physiological signals have been widely used to recognize human emotions for the following advantages. Although physiological signal may happen to artifact due to motion or other environmental factors, its signal acquisition by non-invasive sensors is relatively simple and it is possible to observe user's state in real time. Also, physiological responses can be acquired spontaneous emotional responses not by responses to social masking or factitious emotion expressions and are less sensitive in social and cultural difference [2]. Finally, various physiological signals offer more information for emotion recognition, because physiological responses are related to emotional state [3] and are considered a great potential for emotion recognition in computer systems.

Many emotion researches have studied physiological signals induced by basic emotions [4-12] and recently, emotion recognition based on physiological signals was performed by various algorithms. Studies on emotion classification from physiological responses using machine learning algorithms (e.g., Fisher projection, k-nearest neighbor algorithm, and support vector machines, etc.) have mainly focused on responses induced by basic emotions such as happiness, sadness, anger, fear, and disgust [13-17]. On the other hand, other emotions such as boredom, pain etc. have been least investigated and there are little results of emotion classification on these emotions. Although these emotions aren't basic emotion, they are emotion that human have often experienced in real life and it is needed to classify them from multi-channel physiological signals using machine learning algorithms.

The purposes of this study are to classify three different emotions (pain, boredom, and surprise) using multi-channel physiological signals (ECG, EDA, PPG, and SKT) and to identify the optimal algorithms being able to recognize them. We have operationally defined that surprise emotion is 'startle' response to a sudden unexpected stimulus such as a flash of light, a loud noise, or a quick movement near the face [17-18]. For emotion classification, there are used Decision Tree (which is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences), k-NN (k-nearest neighbor algorithm, which is a method for classifying objects based on closest training examples in the feature space), LDA (linear discriminant analysis, which is a method used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events), and SVM (support vector machine, supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis).

II. EXPERIMENTAL METHODS

A. *Subjects*

129 college students (60 males, 69 females, ages 22.0±2.2 years) and 88 high school students (37 males, 51 females, ages 16±1.3 years) participated in this experiment. They were normal persons who reported no history of medical illness due to heart disease, respiration, or central nervous system disorder. They were introduced to the experiment protocols and filled out a written consent before the beginning of experiment. Also, they were paid $30 USD per session to compensate for their participation.

B. *Emotional Stimuli*

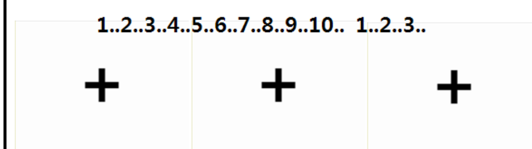
The emotional stimuli used in experiment, which are the 1-3 min long audio-visual stimuli and stimulus provoking pain, had been demonstrated their appropriateness and effectiveness by preliminary psychometric experiment. The appropriateness of emotional stimuli means a consistency between the intended emotion by experimenter and the participants' experienced emotion. The effectiveness is an intensity of emotions that participants rated on a 1 to 7 point Likert-type scale (e.g.., 1 being "least boring" and 7 being "most boring"). The apporiateness and effectiveness of these stimuli were as follows; appropriateness and effectiveness of pain were 97.3% appropriateness and 4.96 ± 1.34 effectiveness, in boredom were 86.0% and 5.23±1.36, and 94.1% appropriateness and 6.12±1.14 effectivess in surprise.

The example of each emotion stimulus is like Table I. The pain provoking stimulus is that it's the more pressure an experimenter put on it after wearing a blood pressure cuff on subjects' arm during 1 minute. The boring stimulus is the combination a presentation of "+" symbol on screen and a repetitive sound of number from 1 to 10 during 3 minutes. The surprise provoking stimulus is the sudden presentation of above images and hog-caller, sound of breaking glass, and thunder during concentration on task during 1 minute.

C. *Experimental Settings and Procedures*

The laboratory is a room with 5m x 2.5m size having a sound-proof (lower than 35dB) of the noise level where any outside artifact are completely blocked. A comfortable chair is placed in the middle of the laboratory and TV monitor set for presentation of film clips is placed in front of the chair. An intercommunication device is placed to the right side of chair for subjects to communicate with an experimenter. A CCTV is installed on the top of the monitor set to observe participant's behaviors and their behaviors are storage through the monitor and a video cassette recorder outside the laboratory.

TABLE I. THE EXAMPLE OF EMOTION STUMULI

Emotion	Stimulus
pain	
	Induction of pain using blood pressure cuff (1 min)
boredom	
	Repetitive sounds of number from 1 to 10 (3 min)
surprise	
	Sudden presentation of above images and hog-caller, sound of breaking glass, and thunder during concentration on task (1 min)

Prior to the experiment, subjects are introduced to detail experiment procedures and have an adaptation time to feel comfortable in the laboratory setting. Then they are attached electrodes on their wrist, finger, and ankle for measurement of physiological signals. Physiological signals are measured for 1 minute prior to the emotional stimuli (baseline) and for 1 to 3 minnutes during the presentation of stimuli (emotional state), then for 1 minute after presentation of the emotional stimuli as recovery term. Subjects rated the own emotion that they experienced during emotional state (Fig. 1).

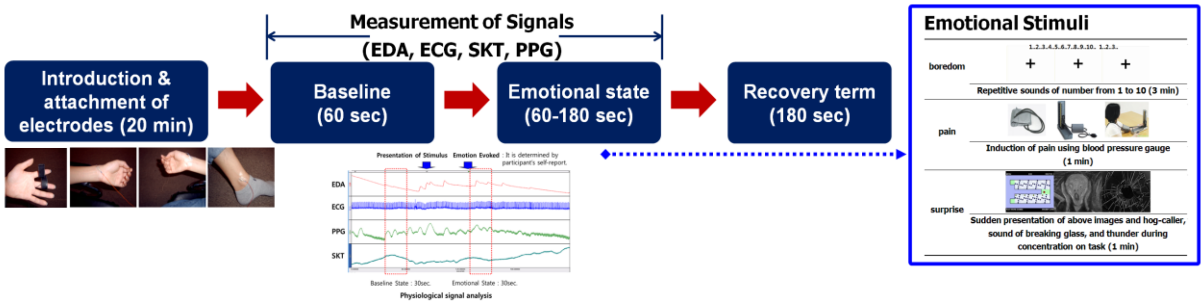
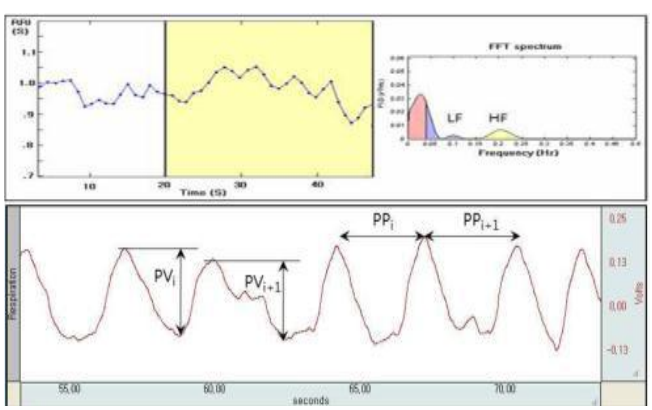


Figure 1. Experimental procedure.

D. Measuremets of Physisological Signals and Feature extraction

The dataset of physiological signals, electrocardiogram (ECG), electrodermal activity (EDA), skin temperature (SKT), and photoplethysmography (PPG) were collected by MP150 Biopac system Inc. (USA). For measurement of ECG, ECG electrodes were placed on both wrists and one left ankle with two kinds of electrodes, sputtered and AgCl ones. The electrode on left-ankle was used as a reference. EDA was measured with the use of 8 mm AgCl electrodes placed on the volar surface of the distal phalanges of the index and middle fingers of the non-dominant hand. The electrodes were filled with a 0.05 molar isotonic NaCl paste to provide a continuous connection between the electrodes and the skin. SKT electrode was attached on the first joint of the non-dominant ring finger and on the first joint of the non-dominant thumb for PPG. These signals were sampled with sampling rate 250Hz, and appropriate amplification and band-pass filtering were performed.

The signals are acquired for 1 minute long baseline state prior to presentation of emotional stimuli and 1-3 minutes long emotional states during presentation of the stimuli as emotional state. To extract features, the obtained signals are analyzed for 30 seconds from the baseline and the emotional state by AcqKnowledge (Ver. 3.8.1) software (USA) as shown in Fig. 2. 27 features are extracted and analyzed from the obtained physiological signals (Table II).



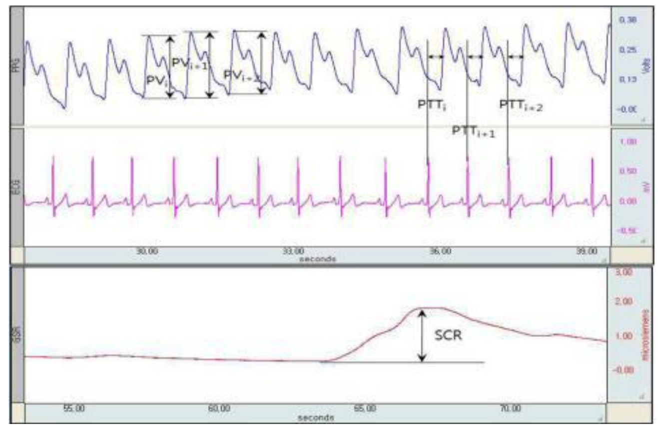


Figure 2. The example of acquired physiological signals.

TABLE II. THE EXTRACTED PHYSIOLOGICAL FEATURES

Signals	Features
EDA	b_SCL, b_SCR, e_SCL, e_SCR, d_SCL, d_SCR
SKT	b_meanSKT, e_meanSKT, d_meanSKT
PPG	b_BVP, b_PPT, e_BVP, e_PPT, d_BVP, d_PPT
ECG	b_HR, b_LF, b_HF, b_HRV, e_HR, e_LF, e_HF, e_HRV, d_HR, d_LF, d_HF, d_HRV

b_: baseline
e_: emotional state
d_: 'e_' – 'b_'

E. Machine Learning Algorithms for Emotion Recognition

For three different emotion classification, 4 machine learning algorithms, Dicision tree, k-NN, LDA, and SVM were applicated by using the extracted features. Decision tree is a hierarchy based classifier in which each branch node represents an option between a number of alternatives, and each leaf node represents a decision and a decision support tool that uses a tree-like graph or model of decisions and their possible consequences[19]. It can select from among a large number of variables those and their interactions that are most important in determining the outcome variable to be explained. Given the data represented at a node, either declare that node to be a leaf (and state what category to assign to it), or find another property to use to split the data into subsets. Decision trees have various advantages: it is possible to validate a model using statistical tests and performs well with large data in a short time. But, decision tree learners can create over-complex trees that do not generalise the data well.

K-NN is a method for classifying objects based on closest training examples in the feature space. It is a method for classifying objects based on closest training examples in the feature space and is a simple and efficient classifier, so it is proper to apply KNN to emotion recognition. The k-nearest neighbor classifier assigns an utterance to an emotional state according to the emotional state of the k utterances that are closest to uξ in the measurement space. It's a method for classifying patterns based on closest training datasets without probability arguments in the feature space. K-NN decision rule provides a simple nonparametric procedure for the assignment of a class label to the input pattern based on the class labels represented by the k-closest neighbors of the vector. However, the disadvantages of k-NN is that systematic methods for selecting the optimum number of the closest neigh bors and the most suitable distance measure are hard to find.

LDA which is one of the linear models is a method used in statistics, pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of objects or events. LDA finds the direction to project data on so that between-class variance in maximized and within-class variance in minimized, and then offers a linear transformation of predictor variables which provides a more accurate discrimination [20]. In LDA, the measurement space is transformed so that the separability between the emotional states is maximized. The separability between the emotional states can be expressed by several criteria.

SVM is non-linear model, which are used the well-known emotion algorithms and support vector classifier separates the emotional states with a maximal margin. The advantage of support vector classifier is that it can be extended to nonlinear boundaries by the kernel trick. SVM supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. SVM is designed for two class classification by finding the optimal hyperplane where the expected classification error of test samples is minimized and has utilized as a pattern classifier to overcome the difficulty in pattern classification due to the large amount of within-class variation of features and the overlap between classes, although the features are carefully extracted [20]. The goal in training SVM is to find the separating hyperplane with the largest margin. We expect that the larger the margin, the better generalization of the recognizer [21].

In the next section, we will discuss the compartive results of emotion classification by the four algorithms as the mentioned above. These algorithms are well-known general methods studied in lots of literatures. We have used the Classification Toolbox of MATLAB for Decision tree and Duda's Toolbox, see www.yom-tov.info/toolbox.html, for k-NN, LDA and SVM. We used feature normalization and the related parameters of algorithms used default values, which have offered with toolbox.

III. RESULTS

The purpose of this study is to compare the performance of each classifier and we used the recognition accuracy as the performance of a classifier for three emotions, i.e., pain, boredom, and surprise. The performance of each classifier was evaluated by 10 fold cross-validation for the overfitting problem and the results of this study are reported for those. For the recognition of three emotions, Table III contrasts the recognition accuracy (%) of the used algorithms. Our result showed that the optimal algorithm being able to recognize three emotions was LDA (74.9%).

The more detail results of emotion recognition accuracy by each algorithm are like from Table IV to VII. Decision tree provided accuracy of 67.8% when it recognized all emotions and accuracy of each emotion had range of 58.9% to 76.1%. Pain was recognized by Decision tree with 69.8%, boredom 76.1%, and surprise 58.9% as shown in Table IV. In analysis of k-NN, the accuracy of all emotions was 62.0% and accuracy of each emotion showed that accuracy of 61.5% was achieved in pain, 68.2% in boredom, and 56.8% in surprise. LDA had recognition accuracy of 74.9% in all emotions as shown in Table III. LDA showed recognition accuracy of 76.3%, 75.6%, and 72.9% according to orders of pain, boredom, and surprise. Finally, as can be seen in Table VII, the result of the SVM was 62.0% in all emotions and this algorithm successfully recognized pain (62.1%), boredom (67.0%), and surprise (57.3%).

TABLE III. RESULT OF EMOTION CLASSIFICATION BY MACHINE LEARNING ALGORITHMS

Algorithm	Accuracy (%)	Features (N)
Decision tree	67.8	27
k-NN	62.0	27
LDA	74.9	27
SVM	62.0	27

 ISBN: 978-1-61208-250-9

TABLE IV. RESULT OF EMOTION CLASSIFICATION BY DECISION TREE

	Pain	Boredom	Surprise
Pain	69.2	6.5	24.3
Boredom	6.8	76.1	17.0
Surprise	21.9	19.3	58.9

TABLE V. RESULT OF EMOTION CLASSIFICATION BY K-NN

	Pain	Boredom	Surprise
Pain	61.5	7.7	30.8
Boredom	8.0	68.2	23.9
Surprise	27.6	15.6	56.8

TABLE VI. RESULT OF EMOTION CLASSIFICATION BY LDA

	Pain	Boredom	Surprise
Pain	76.3	1.8	21.9
Boredom	5.7	75.6	18.8
Surprise	20.8	6.3	72.9

TABLE VII. RESULT OF EMOTION CLASSIFICATION BY SVM

	Pain	Boredom	Surprise
Pain	62.1	5.9	32.0
Boredom	8.0	67.0	25.0
Surprise	28.6	14.1	57.3

IV. CONCLUSION

We identified that three different emotions (pain, boredom, and surprise) were classified by machine learning algorithms from various physiological features. For this, twenty seven features were extracted by means of the statistical and the geometric approaches in time and frequency domain from physiological signals i.e., ECG, EDA, SKT, and PPG and these signals were induced by emotional stimuli.

Also, we recognized three emotions by 4 machine learning algorithms of Decision tree, k-NN, LDF, and SVM. Our result showed that LDA is the best algorithm being able to classify these emotions. The LDA algorithm offers many advantages in other pattern recognition tasks such as face recognition or speech recognition etc. LDA finds the vectors in the underlying space that best discriminate among classes. LDA method tries to maximize the between-class differences and minimize the within-class ones. LDA method is good at discriminating different classes because it is a surveillance method. But LDA always suffers from a small sample size problem. The problem will happen when the number of training samples is less than the total number of physiological features. Although LDA method has some problems, we think that our result is reliable and stable because it is based on sufficient sample size of 227 subjects' data and 27 features.

The result of this study could help emotion recognition studies lead to better chance to recognize various human emotions by using physiological signals. Also, this result can be useful in developing an emotion theory, or profiling emotion-specific physiological responses, as well as establishing the basis for emotion recognition system in human-computer interaction. Physiological signals offer a great potential for the recognition of emotions in computer systems. But, in order to fully exploit the advantages of physiological measures, standardization needs to be established on the emotional model, stimulus used for the identification of physiological patterns, physiological measures, parameters for analysis, and model for pattern recognition and classification [22].

Future studies are needed to obtain additional signals from other modalities such as facial expression, face temperature, or voice to improve classification rate. And more research is needed to obtain stability and reliability of this result compare with accuracy of emotion classification using other algorithms.

ACKNOWLEDGMENT

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (No.012K001330 & 2012K001339).

REFERENCES

[1] J.Wagner, J. Kim, and E. Andre, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," *IEEE International Conference on Multimedia and Expo*, pp. 940-943, 2005.

[2] N. Sebe, I. Cohen, and T.S. Huang, Multimodal emotion recognition, in *Handbook of Pattern Recognition and Computer Vision*, Amsterdam: Publications of the Universiteit van Amsterdam, 2005, pp. 1-23.

[3] P.D. Drummond, and S.-H. Quah, "The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians," *Psychophysiology*, pp. 190-196, 2001.

[4] O. Alaoui-Ismaili, O. Robin, H. Rada, A. Dittmar, A. and E. Vernet-Maury, "Basic emotions evoked by odorants: comparison between autonomic responses and self-evaluation," *Physiology and Behavior*, 62, pp. 713-720, 1997.

[5] A.R. Ax, "The physiological differentiation between fear and anger in humans," *Psychosomatic Medicine*, vol. 15, pp. 147-150, 1953.

[6] F.A. Boiten, "Autonomic response patterns during voluntary facial action," *Psychophysiology*, vol. 33, pp. 123-131, 1996.

[7] T.C. Kanade, and Y. Tian, "Comprehensive database for facial expression analysis," *Proceeding of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46-53, 2000.

[8] D. Palomba, M. Sarlo, A. Angrilli, and A. Mini, "Cardiac responses associated with affective processing of unpleasant film stimulus," *International Journal of Psychophysiology*, vol. 36, pp. 45-57, 2000.

[9] R.W. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Transactions of Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1175-1191, 2001.

[10] R. Sinha, and O.A. Parsons, "Multivariate response patterning of fear and anger," *Cognition and Emotion*, vol. 10, pp. 173-198, 1996.

[11] G. Stemmler, *Physiological processes during emotion*. In: Phillippot, P., Feldman, R. S. (Eds.), The regulation of emotion, Erlbaum, Mahwah, NJ, 2004, pp. 33-70.

[12] C.L. Stephens, I.C. Christie, and B.H. Friedman, "Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis," *Biological Psychology*, vol. 84, pp. 463-473, 2010.

[13] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. G. Taylor, "Emotion recognition in human computer interaction," *IEEE Signal Process. Mag*., vol. 18, pp. 32-80, January 2001.

[14] J.A. Healey, "*Wearable and Automotive Systems for Affect Recognition from Physiology*," PhD thesis, MIT, Cambridge, MA, May 2000.

[15] R.W. Picard, E.Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. On Pattern Analysis and Machine Intell*., vol. 23, pp.1175-1191, October 2001.

[16] A. Haag, S. Goronzy, P. Schaich, J. Williams, "Emotion recognition using bio-sensors: First step towards an automatic system," in *Affective Dialogue Systems Tutorial and Research Workshop*, Kloster Irsee, Germany, June 2004.

[17] F. Nasoz, K. Alvarez, C.L. Lisetti, and N. Finkelstein, "Emotion recognition from physiological signals using wireless sensors for presence technologies," *Cognition, Technology and Work*, vol. 6, pp. 4-14, 2004.

[18] T. Verhoef, C. Lisetti, A. Barreto, F. Ortega, T. Zant, and F. Cnossen, "Bio-sensing for emotional characterization without word labels," Human-Computer Interaction: Ambient, ubiquitous and intelligent interaction, *13th International Conference*, San Diego, CA, USA, July 19-24, pp. 693-702, 2009.

[19] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision trees: an overviewand their use in medicine," *Journal of Medical Systems*, vol. 26, pp. 445-463, October 2002

[20] P.D. Wasserman, *Advanced Methods in Neural Computing*, New York, Van Nostrand Reinhold, 1993, pp. 35–55.

[21] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," *2nd International Conference on Autonomous Robots and Agents*, Palmerston North, pp. 186-191, December, 2004.

[22] J. Arroyo-Palacios, and D.M. Romano, "Towards a standardization in the use of physiological signals for affective recognition systems," *Proceedings of Measuring Behavior 2008*, 2008.

 ISBN: 978-1-61208-250-9