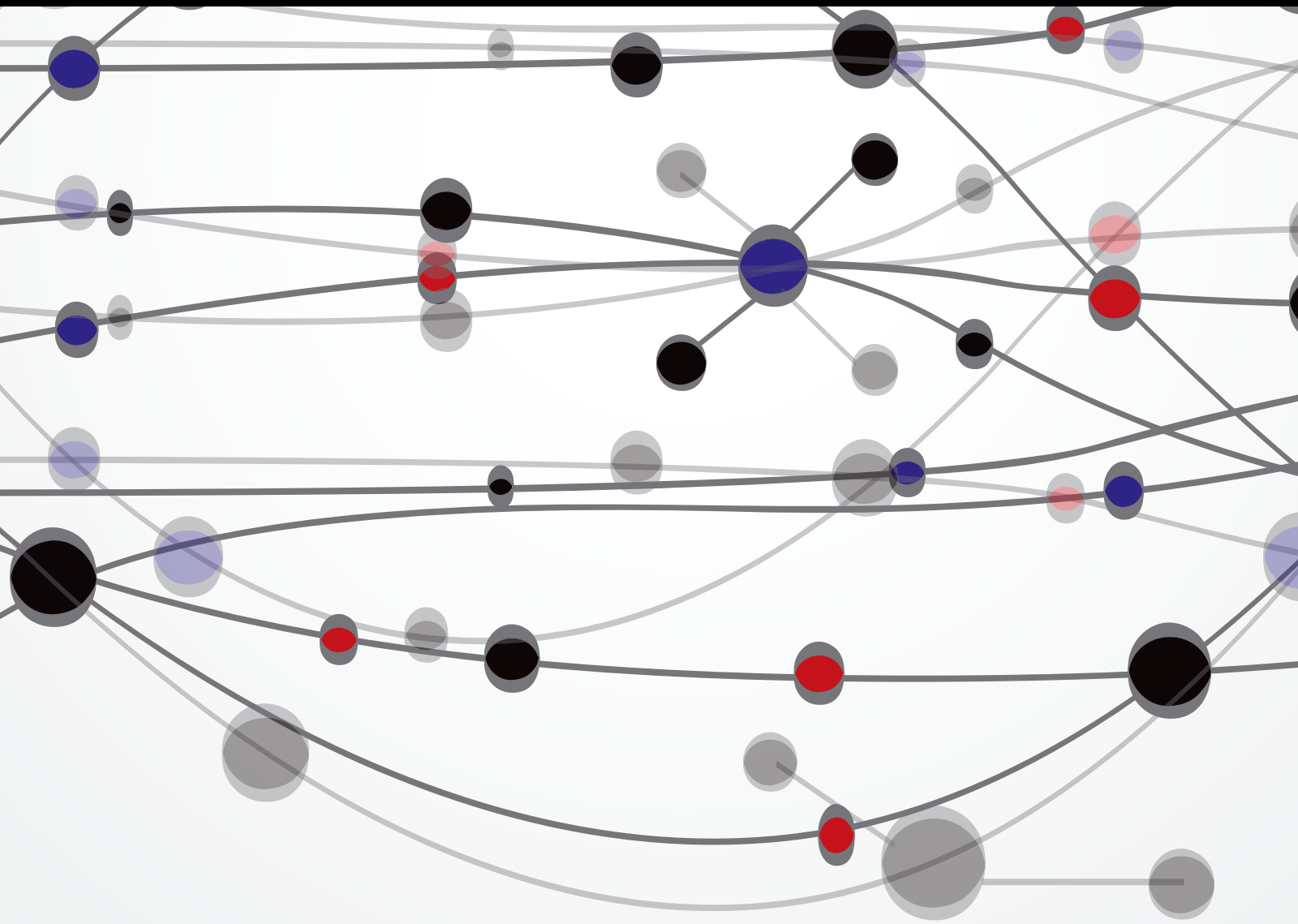
The Scientific World Journal

Detection, Measurement, and Enhancement of Happiness

Guest Editors: Jhing-Fa Wang, Chung-Hsien Wu, Shulan Hsieh, Shyhnan Liou, and Bo-Wei Chen



Detection, Measurement, and Enhancement of Happiness

The Scientific World Journal

Detection, Measurement, and Enhancement of Happiness

Guest Editors: Jhing-Fa Wang, Chung-Hsien Wu, Shulan Hsieh, Shyhnan Liou, and Bo-Wei Chen

Contents

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2014, Article ID 841206, 1 page
http://dx.doi.org/10.1155/2014/841206

Editorial

Detection, Measurement, and Enhancement of Happiness

Jhing-Fa Wang,[1] Chung-Hsien Wu,[2] Shulan Hsieh,[3] Shyhnan Liou,[4] and Bo-Wei Chen[1]

[1] *Department of Electrical Engineering, National Cheng Kung University, No. 1 Daxue Road, Tainan 70101, Taiwan*
[2] *Department of Computer Science and Information Engineering, National Cheng Kung University, No. 1 Daxue Road, Tainan 70101, Taiwan*
[3] *Department of Psychology and Institute of Allied Health Sciences, National Cheng Kung University, No. 1 Daxue Road, Tainan 70101, Taiwan*
[4] *Institute of Creative Industries Design, National Cheng Kung University, No. 1 Daxue Road, Tainan 70101, Taiwan*

Correspondence should be addressed to Bo-Wei Chen; dennisbwc@gmail.com

Received 26 February 2014; Accepted 26 February 2014; Published 20 May 2014

Research on happiness has gained much attention in recent years since the United Nations published a report on the worldwide happiness index. Such publication subsequently stimulates academia and national policies; related research therefore becomes a hot spot. The idea of happiness research is dated back to 1972 when the king of Bhutan proposed using Gross National Happiness to replace traditional economic indices. The pursuit of well-being happiness gradually initiates the formation of new theories, for example, Satisfaction with Life Index, Quality-of-life Index, Happy Planet Index, and positive psychology.

Although a great number of studies have been conducted in social science, there are still few papers accommodating the needs of technology. In view of this, the special issue particularly highlights happiness research on information technology. More specifically, this issue discusses how technology can be used for enhancement of positive emotions, not simply for detection and measurement.

Five papers selected from the submission present a recent update and the advances of technologies in detection, measurement, and enhancement of happiness. The paper by Y. Chin et al. focused on kernel design for Support Vector Machines (SVMs). The authors used music classification as a case study to evaluate the proposed SVM. Their research helped users select music of interest based on emotions in the audio. C. Chuang et al. developed a digital learning system based on somatosensors. Such a system was capable of enhancing the learning performance of users by analyzing physical expressions and interactions. The work by C. Lin et al. investigated feature extraction for recognition of emotional speech. The recognition result could be further evaluated by medical doctors or psychologists and subsequently made into personal profiles. N. Jatupaiboon et al. proposed an emotion classification system based on real-time electroencephalograms (EEGs). The brain waves of subjects were elicited by pictures and sound and subsequently analyzed by the proposed spectral characterization. The authors also constructed a feedback system that allowed users to practice controlling emotions. Finally, P. Cipresso et al. created a multidimensional valence-arousal model to present emotional changing paths. Their model demonstrated another research direction in psychophysiology and affective computing.

Through these papers, readers can obtain an overview of happiness research on information technology. Besides, readers can also understand several practical issues, including feature extraction, pattern recognition, and feedback analysis during processing audiovisual and biomedical signals. In the near future, we expect the happiness research on information technology will become a new subject, "Happiness Informatics," and have an impact on society.

Jhing-Fa Wang
Chung-Hsien Wu
Shulan Hsieh
Shyhnan Liou
Bo-Wei Chen

Research Article

Emotion Identification Using Extremely Low Frequency Components of Speech Feature Contours

Chang-Hong Lin, Wei-Kai Liao, Wen-Chi Hsieh, Wei-Jiun Liao, and Jia-Ching Wang

Department of Computer Science and Information Engineering, National Central University, Taiwan

Correspondence should be addressed to Jia-Ching Wang; jcw@csie.ncu.edu.tw

Received 2 September 2013; Accepted 19 November 2013; Published 20 May 2014

Academic Editors: B.-W. Chen, S. Liou, and J.-F. Wang

The investigations of emotional speech identification can be divided into two main parts, features and classifiers. In this paper, how to extract an effective speech feature set for the emotional speech identification is addressed. In our speech feature set, we use not only statistical analysis of frame-based acoustical features, but also the approximated speech feature contours, which are obtained by extracting extremely low frequency components to speech feature contours. Furthermore, principal component analysis (PCA) is applied to the approximated speech feature contours so that an efficient representation of approximated contours can be derived. The proposed speech feature set is fed into support vector machines (SVMs) to perform multiclass emotion identification. The experimental results demonstrate the performance of the proposed system with 82.26% identification rate.

1. Introduction

As technology advances, computers or machines with human emotion are no longer an unreachable dream. A system that understands human emotions can provide more personalized services and be extended to more applications. Professor Rosalind Picard stated that affective computing explores the topic about computing that relates to, arises from, or influences emotions. This concept creates a new computing system or idea of human-machine interface design which can realize, recognize, and utilize human emotions. The concept above coincides with "Technology derives from humanity." Today, in the twenty-first century, human's demand for high-tech product is no longer merely some specific functionality. Humanity is an essential element for successful product. The design focusing on affectivity will become the mainstream.

Speech emotion identification [1–8] can be divided into two parts, features and classifiers. In the feature part, many features have been considered having a significant influence on emotion identification. For example, both of [1, 2] use pitch, energy, and Mel-frequency cepstral coefficients (MFCCs) and so forth. Other features such as voice quality [8], spectrum features [3], and nonlinear Teager energy operator features [9] are also exploited to represent the characteristics of emotions. In [4, 10], the authors use linear regression to describe graphical trend of pitch and use polynomial coefficients as features to train and identify [8]. Legendre trend uses polynomial bases that are different from linear regression to describe speech information and obtain features [1, 8]. In this paper, besides the statistical analysis of the selected acoustical features, approximated speech feature contours with (PCA) [11, 12] are also utilized in the feature extraction. In the classifier part, we use a multiclass support vector machine (SVM) [7, 12, 13] to achieve emotion identification.

The rest of this paper is organized as follows. Section 2 introduces the overview of the system. The proposed speech feature set for emotion identification is illustrated in Section 3. Section 4 describes the emotion classifier used in this paper. Next, Section 5 gives the experimental results. Conclusions are finally drawn in Section 6.

2. System Overview

Figure 1 illustrates the proposed system block diagram. The proposed emotion identification system can be divided into

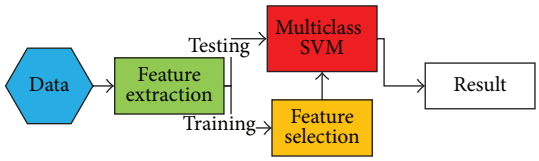


Figure 1: Block diagram of the proposed system.

two main parts, feature extraction and emotion classifier. In the feature extraction, we extract all the acoustical features from both of training and testing speeches. The acoustical features comprise silence/active, voiced/unvoiced, pitch, log frame energy (LFE), subband power, spectral centroid, spectral bandwidth, spectral flatness, MFCCs [3], and ratio of spectral flatness to spectral centroid (RSS) [7]. The statistical analysis features are generated from the aforementioned frame-based acoustical features. The chosen feature is important for classification task [14, 15]. A feature selection procedure chooses the effective statistical analysis features to form the selected feature set. The selected statistical analysis feature set as well as the selected approximated speech feature contours is combined to form the proposed feature set. In the emotion classifier, a multiclass SVM is used as the classifier to identify the emotion class of a speech utterance.

3. Emotion Feature Extraction

Extracting appropriate speech features is an important issue in emotion identification. Features can be used to describe the emotion characteristics of speech. Suitable feature set can effectively increase the performance of classifiers. For emotional speech identification, most systems partition speech waveforms into segments called frames, and a classifier is trained using features extracted from frames. These features are frame-based speech features, which usually have an excellent performance for highly nonstationary signals. To further enhance the frame-based features, statistical analysis such as ratio, mean, and standard deviation of frame-based features usually creates more reliable speech features. In addition to statistical analysis of frame-based features, this paper also considers the approximated speech feature contours.

3.1. Frame-Based Features with Statistical Analysis. The frame-based acoustical features extracted in this paper include silence/active, voiced/unvoiced, pitch, LFE, subband power, spectral centroid, spectral bandwidth, spectral flatness, MFCCs [3], and RSS [7]. Among these frame-based acoustical features, several simpler ones are briefly explained below. Silence/active denotes if a frame is a silence frame or not, while voiced/unvoiced represents if a frame is a voiced frame. Pitch expresses the repeating duration in a voiced speech frame. LFE makes a logarithm operation on a frame energy. In the following, the more complicated frame-based acoustical features are explained.

The subband powers are extracted from specified subband intervals. In this paper, we adapt four intervals which are $[0, 0.125f_0]$, $[0.125f_0, 0.25f_0]$, $[0.25f_0, 0.5f_0]$, and $[0.5f_0, f_0]$, where f_0 is half of the sampling rate. Spectral centroid [16, 17] denotes the weighted average of the frequencies of a speech power spectrum generated from a speech frame. Spectral bandwidth [16, 17] indicates whether the shape of the speech power spectrum concentrates in the neighborhood of the spectral centroid or else spreads out over the speech power spectrum. The spectral flatness is obtained by computing the ratio of the geometric mean and the arithmetic mean of the speech power spectrum coefficients. MFCCs are nonparametric representations of audio signal, which models the human auditory perception system. The RSS proposed by Kim et al. for emotional speech recognition [7] is the ratio of spectral flatness to spectral centroid. The RSS of ith frame is calculated by

$$\mathrm{RSS}(i) = \frac{1000 \times \mathrm{SF}(i)}{\mathrm{SC}(i)}, \tag{1}$$

where $\mathrm{SF}(i)$ and $\mathrm{SC}(i)$ denote spectral flatness and spectral centroid of ith frame.

To convert these frame-based features into more effective features, statistical analysis is applied to each of them. This paper adopts three statistical analysis types, ratio, mean, and standard deviation. The ratio analysis is applied to silence/active and voiced/unvoiced to calculate the silence ratio and voiced ratio, respectively. As for the remaining features, both mean and standard deviation analysis are performed.

3.2. Approximated Speech Feature Contours. In the previous subsection, statistical analysis of frame-based acoustical features is performed. This subsection provides another point of view to extract emotion related information, that is, the temporal shape of a feature contour. As the detailed shape information may be merely aroused from the various phone pronunciations rather than the various emotions, an approximated speech feature contour is proposed. Figure 2 depicts the flowchart to obtain the approximated speech feature contour.

Let x_{lk} denote an original feature contour, where l is the speech utterance index, $k = 1, 2, \ldots, T_l$ is the frame index, and T_l is the frame number of lth speech utterance. As the durations of different utterances differ, a resampling scheme is performed to normalize the feature contour durations so that the feature contour comparisons are fair.

Besides, speech feature contours usually contain too much local fluctuations that are unrelated to emotion expression. To lessen this impact, this paper presents the approximated speech feature contours. First, the well-known discrete Fourier transform [18] is applied to the resampled speech and transformed it to frequency domain by

$$y_{lj} = \sum_{k=0}^{F\_\mathrm{length}-1} e^{-(2\pi i/n)jk} x'_{lk}, \quad j = 0, 1, \ldots, F\_\mathrm{length} - 1, \tag{2}$$

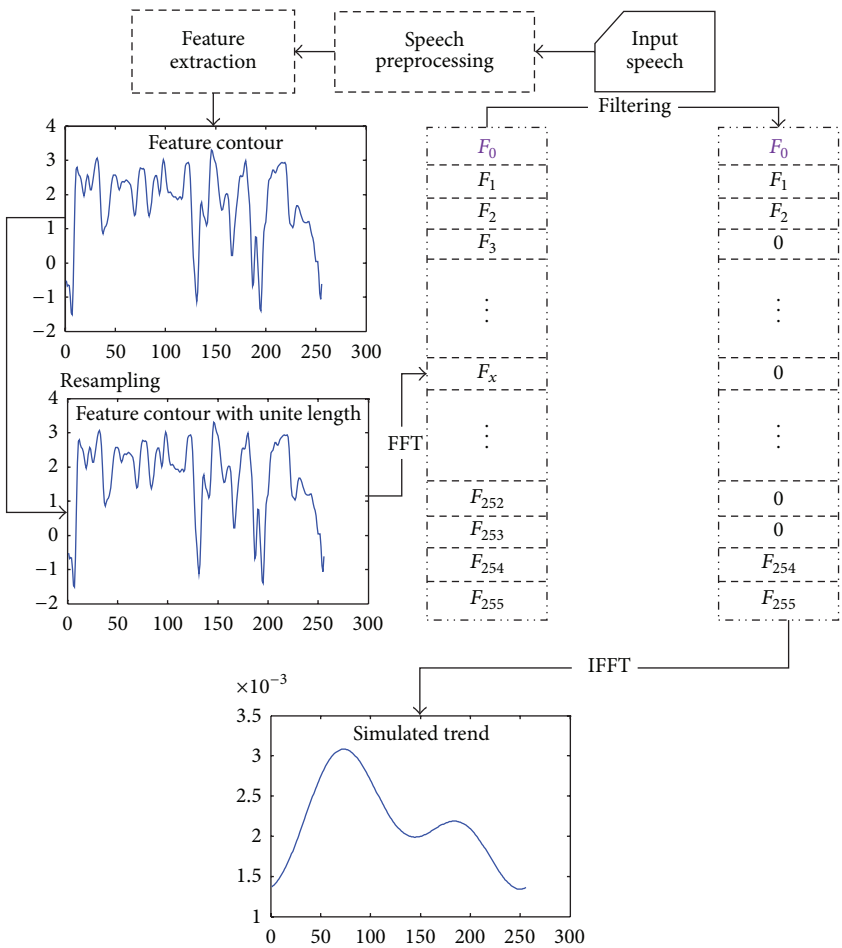


Figure 2: Flowchart to obtain the approximated speech feature contour.

where x' represent the resampled feature contour and its duration has been normalized to $F_$length. Moreover, x'_{lk} refers to the kth feature in the feature contour of the lth speech utterance, and y_{lj} is the jth Fourier coefficient of the lth speech utterance.

The presented approximated speech feature contours rely on extracting extremely low frequency components to speech feature contours. To extract these frequency components, the symmetric property of discrete Fourier spectrum is also required to be taken into account. Assume m lowest-frequency Fourier coefficients (frequency bins) are to be extracted; the frequency component extraction is achieved by converting y_{lj} in (2) into (3):

$$y'_{lj} = \begin{cases} 0, & m < j < n - m, \\ y_{lj}, & \text{others}. \end{cases} \tag{3}$$

Finally, the extremely low frequency components y'_{lj} in (3) are transformed to time-domain by inverse Fourier transform. The resynthesized time-domain signal is the proposed approximated speech feature contour.

In this paper, parameter m is chosen as 1, 2, and 3. Taking log energy feature, for example, Figures 3, 4, and 5 exemplify the approximated log energy contours of an angry, a bored, and a sad speech utterances, respectively.

3.3. Approximated Speech Feature Contours with Principal Component Analysis. In this paper, principal component analysis is adopted to represent the approximated speech feature contours in an efficient way. Principal component analysis is a well-known technique in multivariate analysis and pattern recognition [11]. In this study, PCA is used to reduce the high feature dimension of an approximated speech feature contour.

To generate the PCA bases of approximated speech feature contours, a training speech database is required so that eigenvectors of the approximated speech feature contours can be found. The goal of PCA is to search a linear combination of the original bases that maximizes the total variance of training approximated speech feature contours. By selecting the top κ principle bases or eigenvectors, PCA projection matrix is capable of representing approximated

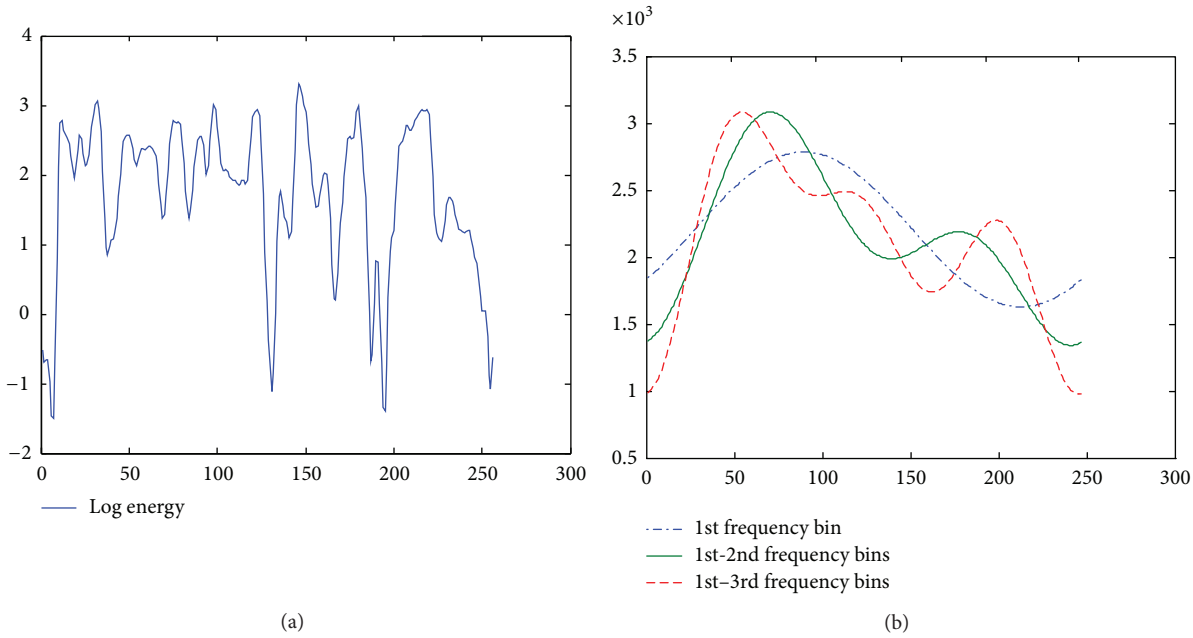


Figure 3: Examples of approximated LFE contours from an angry speech utterance.

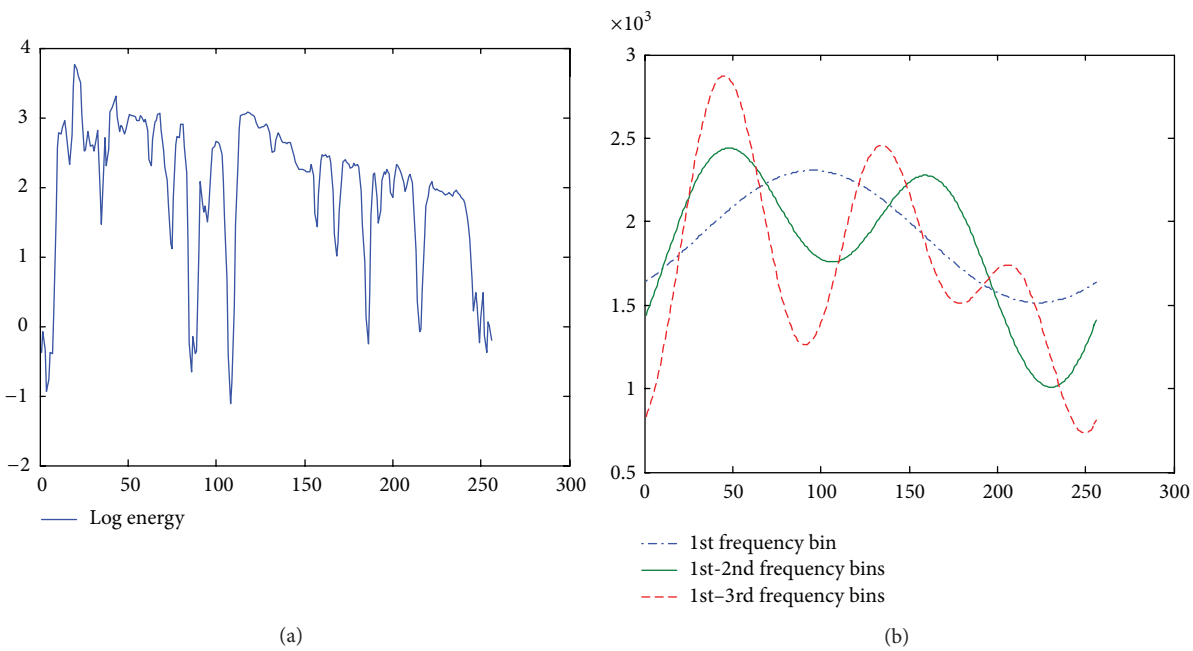


Figure 4: Examples of approximated LFE contours from a bored speech utterance.

speech feature contour accurately with lower dimensions of projection coefficient vector.

To perform emotion identification, the PCA projection coefficients are utilized as the speech features. The PCA projection coefficients are computed by projecting the approximated speech feature contour onto the PCA bases. With the obtained PCA bases $\overline{v}_i$, $i = 1, 2, \ldots, \kappa$, the PCA projection matrix is formed as $\mathbf{V} = (\overline{v}_1, \overline{v}_2, \ldots, \overline{v}_\kappa)$. Let the approximated speech feature contour be denoted by $\overline{a} = (a_1, a_2, \ldots, a_N)^T$, where N is the normalized contour length. Let $\overline{c}_{\mathrm{PCA}}$ refer to the PCA projection coefficient vector. The $\overline{c}_{\mathrm{PCA}}$ is computed by

$$\overline{c}_{\mathrm{PCA}} = \mathbf{V}^T \overline{a}. \tag{4}$$

Take approximated LFE contour, for example; each m value (Section 3.2) associates with a set of PCA bases derived from the training approximated LFE contours. Figures 6, 7, and 8 give the PCA bases of approximated LFE contours with

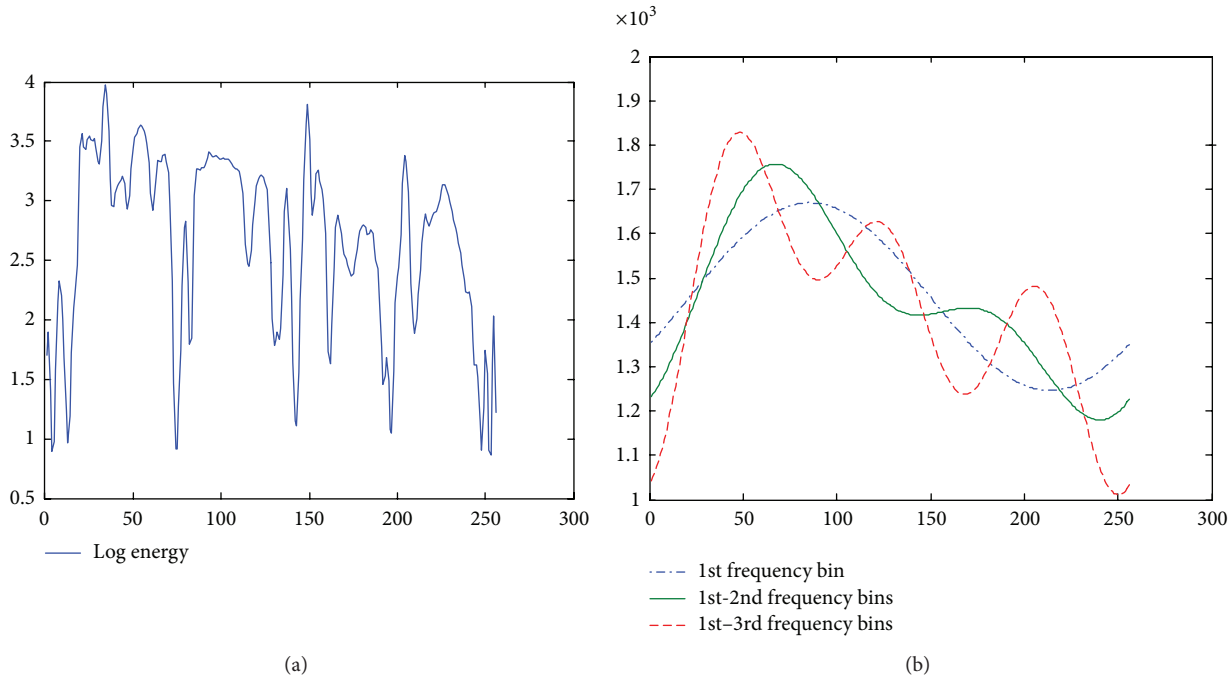


Figure 5: Examples of approximated LFE energy contours from a sad speech utterance.

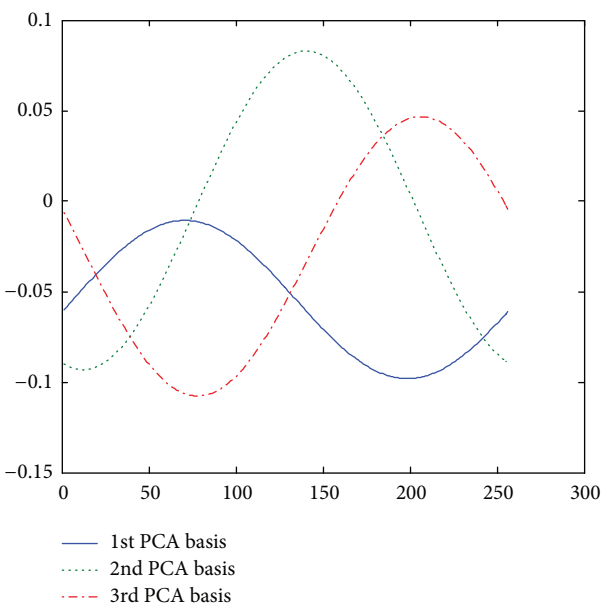


Figure 6: PCA bases of approximated LFE with m value being 1.

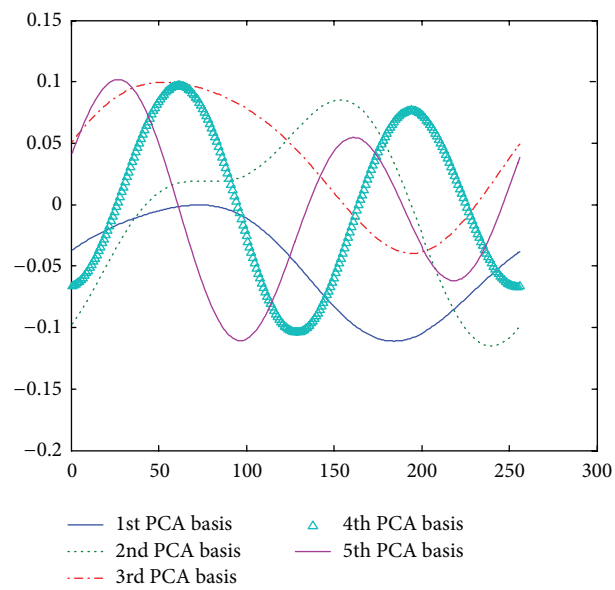


Figure 7: PCA bases of approximated LFE contours with m value being 2.

m value being 1, 2, and 3, respectively. It is noted that only significant PCA bases are shown in these figures.

4. Emotion Identification Using SVM

The SVM based on statistical machine learning is a powerful classifier [13, 19]. Using several crucial support vectors, the SVM has not only a clear structure but also a good classification performance. Considering data from two different classes, an SVM attempts to solve an optimization problem that finds a hyperplane that separates the data with maximum margin. Suppose the optimal separating hyperplane $(\overline{w} \cdot \overline{x}) + b = 0$, with $\overline{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$, maximizes the margin $2/\|\overline{w}\|^2$. A data point $\overline{x}$ is then labeled $y \in \{1, -1\}$ based on the decision function. To introduce kernel concepts,

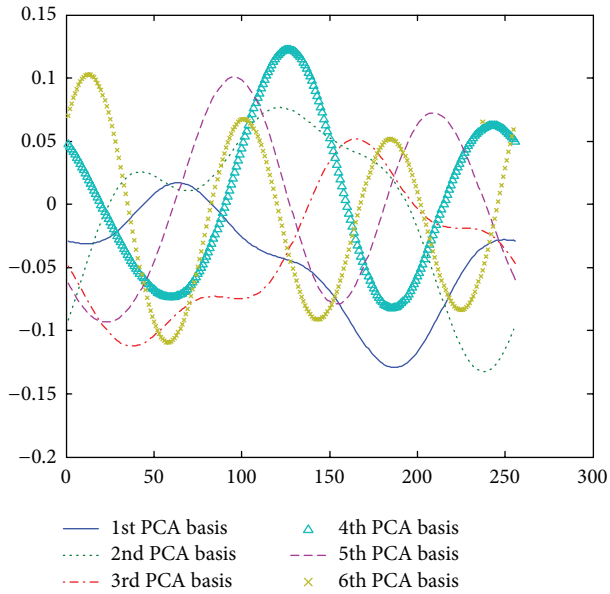


Figure 8: PCA bases of approximated LFE contours with m value being 3.

the separating hyperplane function in terms of the inner product of $\overline{x}$ can be rewritten as

$$f(\overline{x}) = \operatorname{sign}((\overline{w} \cdot \overline{x}) + b) = \operatorname{sign}\left(\sum_{i=1}^{m} \alpha_i \overline{x}_i \cdot \overline{x} + b\right) = \operatorname{sign}\left(\sum_{i=1}^{m} \alpha_i k(\overline{x}, \overline{x}_i) + b\right), \quad (5)$$

where α is a Lagrange multiplier, i is the number of vectors, and $k(\overline{x}, \overline{x}_i)$ is a kernel function. Using Mercer's theory, we can introduce a mapping function $\varphi(\overline{x})$, such that $k(\overline{x}_j, \overline{x}_i) = \phi(\overline{x}_j)\phi(\overline{x}_i)$. This provides the ability of handling nonlinear data. Typical kernel functions include linear kernel, polynomial, and radial basis kernel.

The 2-class SVM is extended to multiclass emotion identification by one versus one technique. Totally, there are C_2^E 2-class SVMs built, where E is the emotion class number.

5. Experimental Results

The database we used in this paper is German emotional speech database (GES) [20]. This database consists of seven emotion classes that are anger, joy, sadness, boredom, disgust, fear, and neutral and records utterances of five males and five females. Each speaker recorded ten speech utterances for each emotion. The number of speech files is about 800. Because we use only those speech files which are voted by over 80% voters, finally, the number of valid speech files is 535. The longest period of files is 8 seconds. The sampling rate of each file is 16 kHz with a resolution of 16 bits per sample. The frame size is 256 samples, with a 50% overlap in the two adjacent frames. There are 127 angry, 81 bored, 46 disgust, 69 fear, 71 joy, 79 neutral, and 62 sad files. Finally, 50% of the dataset was used for training and 50% for testing. We extracted features from both of the training and testing set.

Table 1: Performance evaluation using approximated LFE contours.

Feature	Dimension	Identification rate (%)
$P_{m=1}$	3	44.9
$P_{m=2}$	5	43.77
$P_{m=3}$	6	44.52

Table 2: Adopted statistical analysis features after feature selection.

Statistical analysis feature set	Dim.
Silence ratio	1
Voiced ratio	1
Mean and standard deviation of pitch	2
Mean and standard deviation of log frame energy	2
Mean and standard deviation of subband powers	8
Mean and standard deviation of spectral centroid	2
Mean and standard deviation of bandwidth	2
Mean and standard deviation of MFCCs	26

First, performance of various approximated speech feature contours was evaluated. The experiment indicates approximated LFE contour has the best performance of them. Considering the approximated log energy contours, the m value of Section 3.2 is set as 1, 2, and 3. For each m value, only significant PCA bases are used. We choose 3, 5, and 6 significant PCA bases for m value equivalent to 1, 2, and 3, respectively. The emotion identification results using PCA projection coefficients are given in Table 1. We abbreviate the PCA projection coefficients as P_m, which are $P_{m=1}$, $P_{m=2}$, and $P_{m=3}$.

In Section 3.1, we introduce many statistical analysis features. In this paper, a feature selection procedure was conducted to choose an effective feature set for emotion identification. The chosen statistical analysis feature set after feature selection procedure is given in Table 2.

In the second experiment, the performance evaluation related to the adopted statistical analysis feature set as well as its combination with approximated LFE contours is summarized in Table 3. In this table, we abbreviate the adopted statistical analysis feature set as Γ. Moreover, the combination of the adopted statistical analysis feature set and approximated log energy contours are represented by Γ, P_m. With the P_m, the experimental result reveals that the identification rate of sadness is decreased slightly by 3.2%, but the identification rates of disgust, joy, and neutral are enhanced by 8.7%, 5.7%, and 5.2%, respectively. The Γ, P_3 feature set achieves the best total identification rate 82.3%, which is 2.26% higher than merely using statistical analysis feature set Γ. The confusion matrix corresponding to the optimal feature set Γ, P_3 is given in Table 4.

6. Conclusion

This work proposes a method to generate approximated speech feature contours using forward and inverse Fourier

Table 3: Performance evaluation using different feature sets.

	Ang.	Bor.	Dis.	Fear	Hap.	Neu.	Sad.	Total
Γ	93.7	80	69.6	76.5	62.9	76.9	87.1	80.0
Γ, P_1	93.7	80	69.6	76.5	65.7	79.5	87.1	80.8
Γ, P_2	93.7	80	73.9	79.4	65.7	82.1	83.9	81.5
Γ, P_3	93.7	80	78.3	79.4	68.6	82.1	83.9	82.3

Table 4: Confusion matrix of Γ, P_3 feature set. The left column denotes actual emotions, and the top row represents predicted emotions.

	Ang.	Bor.	Dis.	Fear	Hap.	Neu.	Sad.
Ang.	59	0	0	1	3	0	0
Bor.	0	32	0	0	0	7	1
Dis.	2	0	18	1	1	0	1
Fear	4	0	0	27	1	2	0
Hap.	8	0	0	3	24	0	0
Neu.	0	7	0	0	0	32	0
Sad.	0	2	0	0	1	2	26

transform. PCA projection coefficients provide an efficient feature representation of the approximated speech feature contours. For PCA projection coefficients of approximated log frame energy contour, 44.39% average identification rate can be achieved. After integrating PCA projection coefficients with the selected statistical analysis feature set, the average identification rate coefficients are increased from 80% to 82.26%. This result demonstrates the effectiveness of the proposed PCA projection coefficients generated from approximated speech feature contours. In the future, the effectiveness of other different Fourier coefficients can be exploited. Moreover, wrapper selection and linear discriminant analysis may further increase the performance.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] C. F. Wu, *Bimodal emotion recognition from speech and facial expression [M.S. thesis]*, Department of Computer Science and Information Engineering, National Cheng Kung University, 2002.

[2] D. Wu, T. D. Parsons, E. Mower, and S. Narayanan, "Speech emotion estimation in 3D space," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '10)*, pp. 737–742, Suntec, Singapore, July 2010.

[3] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis et al., "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.

[4] C. Busso, S. Lee, and S. Narayanan, "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 4, pp. 582–596, 2009.

[5] T. L. Nwe, S. W. Foo, and L. C. de Silva, "Speech emotion recognition using hidden Markov models," *Speech Communication*, vol. 41, no. 4, pp. 603–623, 2003.

[6] I. Luengo, E. Navas, and I. Hernaez, "Feature analysis and evaluation for automatic emotion identification in speech," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 490–501, 2010.

[7] E. H. Kim, K. H. Hyun, S. H. Kim, and Y. K. Kwak, "Improved emotion recognition with a novel speaker-independent feature," *IEEE/ASME Transactions on Mechatronics*, vol. 14, no. 3, pp. 317–325, 2009.

[8] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[9] A. Georgogiannis and V. Digalakis, "Speech emotion recognition using nonlinear Teager energy based features in noisy environments," in *Proceedings of the Signal Processing Conference (EUSIPCO '12)*, Bucharest, Romania, August 2012.

[10] E. Grabe, G. Kochanski, and J. Coleman, "Connecting intonation labels to mathematical descriptions of fundamental frequency," *Language and Speech*, vol. 50, no. 3, pp. 281–310, 2007.

[11] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2001.

[13] J. C. Wang, C. H. Lin, E. Siahaan, B. W. Chen, and H. L. Chuang, "Mixed sound event verification on wireless sensor network for home automation," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 803–812, 2013.

[14] T. L. Nwe, S. W. Foo, and L. C. de Silva, "Speech emotion recognition using hidden Markov models," *Speech Communication*, vol. 41, no. 4, pp. 603–623, 2003.

[15] H. M. Teager, "Some observations on oral air flow during phonation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 5, pp. 599–601, 1980.

[16] ISO-IEC/JTC1 SC29 WG11 Moving Pictures Expert Group, "Information technology—multimedia content description interface—part 4: audio," Comittee Draft 15938-4, ISO/IEC, 2000.

[17] M. Casey, "MPEG-7 sound-recognition tools," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 737–747, 2001.

[18] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 1999.

[19] V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, NY, USA, 1998.

[20] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of German emotional speech," in *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech '05)*, pp. 1517–1520, Lisbon, Portugal, September 2005.

Research Article

The Pursuit of Happiness Measurement: A Psychometric Model Based on Psychophysiological Correlates

Cipresso Pietro,[1] Serino Silvia,[1] and Riva Giuseppe[1,2]

[1] *Applied Technology for Neuro-Psychology Laboratory (IRCCS Istituto Auxologico Italiano), Via Pellizza da Volpedo 41, 20149 Milan, Italy*
[2] *Department of Psychology, Catholic University of Milan, Largo Gemelli 1, 20123 Milan, Italy*

Correspondence should be addressed to Cipresso Pietro; p.cipresso@auxologico.it

Received 31 August 2013; Accepted 17 October 2013; Published 30 April 2014

Academic Editors: B.-W. Chen, S. Liou, and C.-H. Wu

Everyone is interested in the pursuit of happiness, but the real problem for the researchers is how to measure it. Our aim was to deeply investigate happiness measurement through biomedical signals, using psychophysiological methods to objectify the happiness experiences measurements. The classic valence-arousal model of affective states to study happiness has been extensively used in psychophysiology. However, really few studies considered a real combination of these two dimensions and no study further investigated multidimensional models. More, most studies focused mainly on self-report to measure happiness and a deeper psychophysiological investigation on the dimensions of such an experience is still missing. A multidimensional model of happiness is presented and both the dimensions and the measures extracted within each dimension are comprehensively explained. This multidimensional model aims at being a milestone for future systematic study on psychophysiology of happiness and affective states.

It seems everyone has a view on happiness. Joan Collins, the Dalai Lama and over 100 others have released new titles on the subject since the beginning of 2001
Richard Tooth
"The Psychology of Happiness (2nd Edition)"
Michael Argyle, Routledge

1. Introduction

"Life, Liberty and the pursuit of Happiness" is a sentence in the United States Declaration of Independence [1]. The sentence is considered an example of "unalienable rights" to be considered for all human beings.

Everyone is interested in the pursuit of happiness, but the real problem for the researchers is how to measure it. An interesting distinction is between Subjective Well Being (SWB), measures of happiness based on self-reports and surveys, and Objective Well Being, measures of observable variables, for example, based on life expectancy and other variables that we believe important for a good life. Among several methods between these two extremes, our aim is to deeply investigate happiness measurement through biomedical signals, using psychophysiological methods to objectifying the subjective experiences measurements.

Psychophysiology research has come to age to allow sophisticated and objective measurement of perceived experiences. However, there is still room for improvement in the research methods and in the consequent modeling of the involved processes.

The goal of our study was to model subjective experiences by measuring different dimensions of the affective states and the related psychological and physiological spheres.

According to the classic valence-arousal model [2, 3] for identifying affective states in subjects during an experimental session, we can consider the two dimensions of "activation," namely, physiological arousal and emotional valence.

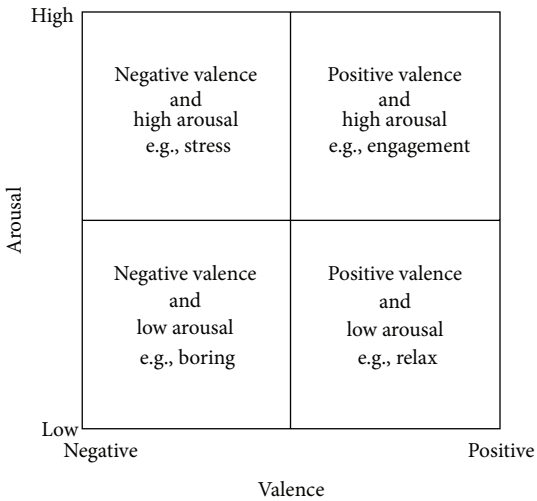


Figure 1: The classic valence-arousal model [2, 3] with the two dimensions of "activation": physiological arousal and emotional valence.

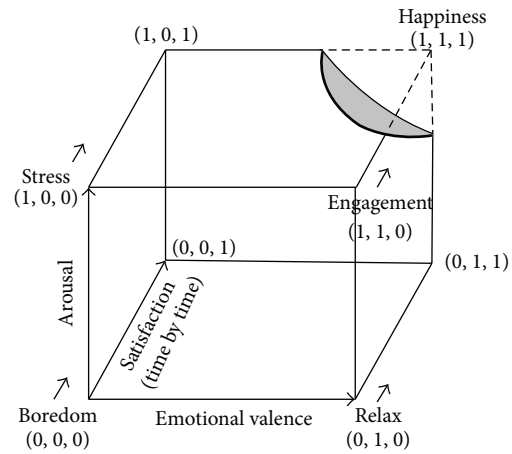


Figure 2: Arousal-Valence-Satisfaction space, identifying on frontal plan the valence-arousal model. The happiness is an extension of the engagement state with a higher level of life satisfaction.

Figure 1 offers an intuitive identification of affective states based on these two dimensions [3].

This approach has been extensively used in psychophysiological research as an objective way to measure affective states during a mediated experience [4–13]. More, recently an extensive research has been done also to discern different emotions by the means of cardiovascular measures [14, 15], and this hugely helps the analysis of affective states, by confirming the results that can be obtained, with specific patterns of the cardiovascular indexes [16–18].

Researchers measuring SWB have been able to deconstruct happiness into separate but related dimensions of positive effect, satisfaction, and negative effect [19–21]. Aroual-valence model can be used in this sense, but, however, we need to add another dimension: the life satisfaction.

Considering this new dimension, we add to the pursuit of proximal goals and immediate pleasure (hedonic enjoyment) also the long-term commitment to pursue "self-realization" (eudaimonia) [22, 23].

Life satisfaction can be considered as the opposite of depression. According to a classic study of Headey et al., "life satisfaction, is quite strongly (negatively) correlated with a distress dimension, depression; life satisfaction and depression are near opposites" [24]. Also recent studies keep the same relationship in clinical and experimental studies [25–36].

Psychophysiology of depression has been studied through Heart Rate Variability measurements historically [37–41], continuing also recently [42–53].

So psychophysiology of life satisfaction dimension can be easily computed referring to several studies and researches from the last twenty years.

In Figure 2, we represented the Arousal-Valence-Satisfaction space, identifying on frontal plan the valence-arousal model and, consequently, the relative affective states and the happiness as an extension of the engagement state due to a higher level of life satisfaction.

Thus, in the model, the happiness is identified as the situation in which subjects have a high physiological arousal, a positive emotional valence, and a high level of life satisfaction. Its vantage is to combine the three dimensions, making a specific experience measurable in a more effective way.

To summarize, we used psychophysiological measures to evaluate life satisfaction, emotional valence, and physiological arousal. In this perspective, engagement and happiness are strictly related to the link between short and long run: the more the subjects will be engaged and satisfied, the more they experience happiness, characterized by positive valence, high arousal, and high life satisfaction. Thus, we aimed to objectively model specific pattern of users' affective state in the Arousal-Valence-Satisfaction plane.

2. Model Hypotheses

The model purpose is to work on typical ground truths in a multidimensional space of objectively measurable variables and to explore if a possible subjective experience can be identified as a happiness experience. Thus, for an effective assessment of subjects' experience, we have to identify stable ground truths in the tridimensional space that we considered.

Several studies, recently, established stable ground truths. Some of the most important databases at this purpose are the IAPS (for the images) [54], the IADS (for the audios) [55], the Affective Norms for English Words (ANEW) [56], the Affective Norms for English Text (ANET) [57], and the Age-Dependent Evaluations of German Adjectives (AGE) [58]. In these databases, several stimuli are classified on the basis of physiological arousal and emotional valence and are used also to investigate other dimensions, like, for example, in the recent study of Leite and Colleagues [59]. These databases have been also investigated in hundreds of psychophysiological studies and also with patients [60], which made evidence of objectivity and effectiveness of the used stimuli.

Table 1: Our pseudohypotheses (~ is for similar, > is for greater than, and < is for lower than).

Hypothesis	Dimension	Comparing happiness with	
		Relax	Stress
Hp1	High physiological arousal	Happiness > relax	Happiness ~ stress
Hp2	Positive emotional valence	Happiness ~ relax	Happiness > stress
Hp3	High life satisfaction	Happiness > relax	Happiness > stress

These and other databases have a great role in psychophysiology, neuroscience, and many other related fields; however, an aspect that often is not considered is that to experiment beyond the basic research, in particular, into the applied field, hugely complexify the situations.

The pursuit of happiness measurement may be a complex issue, and also, in the bidimensional models used at the moment, no one conceptualized correct research methods to analyze the outcome of the involved arousal and valence dimensions yet. This lack leaded to tons of studies where the analysis of statistical differences are considered just "good enough" to publish bidimensional model based on arousal-valence plan, that however considered only single variables without going deeper on the combination or the relations among them.

One statistically correct study, that got beyond the statistical differences, is of Von Leupoldt and Colleagues [61] where an analysis of polynomial contrasts has been conducted to analyze the trends. Another relevant study has been really well conducted by Grühn and Scheibe [10], where the relations between arousal and valence are taken into consideration.

Of course, many studies considered sophisticated statistical technique to analyze the data results; however, no one considered yet combinations and analyses beyond the statistical differences, that come to be essential in multidimensional studies.

The pursuit of the happiness measurement regards a plenty of fields about the human sciences. Typical examples are the studies on ergonomics but also all the studies in the field of positive psychology, where the idea is to investigate the optimal experience and the flow state [62, 63]. Also a new emergent paradigm, the positive technology [64–66], seems going in the same direction.

Since this model aimed at being of a wide interest for several researchers, our approach will be toward the simplification of the complexity naturally embedded in a multidimensional model.

Also the statistical analyses are explained in details in a way to fit well also nonmathematical users, more descriptive than equation-based, however scientifically rigorous.

Based on this approach, we formulated three simple pseudohypotheses (following hypotheses).

Each hypothesis is based on one dimension of the multidimensional space in particular, we considered physiological arousal, emotional valence, and life satisfaction.

As ground truth, we defined two basic affective states, namely, "Relax" and "Stress," to be elicited in a way that can represent the ground truth of relax state and stress state, respectively, in the tridimensional space considered. The way to elicit these two states strongly depends on the study that the researcher is carrying out. Relax can also be induced using panorama slides show with a soft music and a cognitive stress is easily induced by standard cognitive tasks, such as Stroop task or arithmetic task [17, 18, 67, 68].

Thus, in a possible experiment, it will be necessary to foresee at least three epochs: (1) the phenomenon that is investigated to be a happiness experience or not (following, to simplify, the happiness), (2) a relax epoch, and (3) a stress epoch.

This operation is to compare along each axis the happiness with a standard affective state elicited in the subject.

Each hypothesis needs also to be verified for the significant quadratic trend using the within-subjects contrasts (further specifications are given following, in a specific section).

The first hypothesis is on the dimension of physiological arousal. In particular, we hypothesize that a happiness experience leads to be more "activated," that is, with an arousal activation similar to the stress states and enough different from the relax one.

The second hypothesis is on emotional valence, for which we expect, by definition, that a happiness experience is able to generate positive emotions and thus we hypothesize that emotional valence during a happiness experience is similar to a relax state and quite different from the stress state, that generates negative emotions.

The third hypothesis is on life satisfaction, for which we expect to have a high level of satisfying experience repeated time by time to make a happiness experience attracting continuously. The process is dissimilar to the one activated during a stress state, where the alertness toward the complex task leads users to move far from satisfaction. Thus we hypothesize that satisfaction during a happiness experience is different from both relax and stress states.

Thus, happiness experience differs from stress for the emotional valence and the satisfaction, being similar in the physiological arousal. This could be a great weakness of the model, since a few errors in measuring a variable could lead to opposite conclusions, considering a stressful experience as a happiness one. To avoid these misleading consequences is our strong suggestion to avoid considering the physiological arousal to measure happiness: this would bring to great errors.

A synthesis of the hypotheses is reported in Table 1. Of course, these hypotheses make it difficult to find a happiness experience; however, this is due to the fact that it is a complex phenomenon and not to the experimental variables, that are only used to objectively measure the subjects' states.

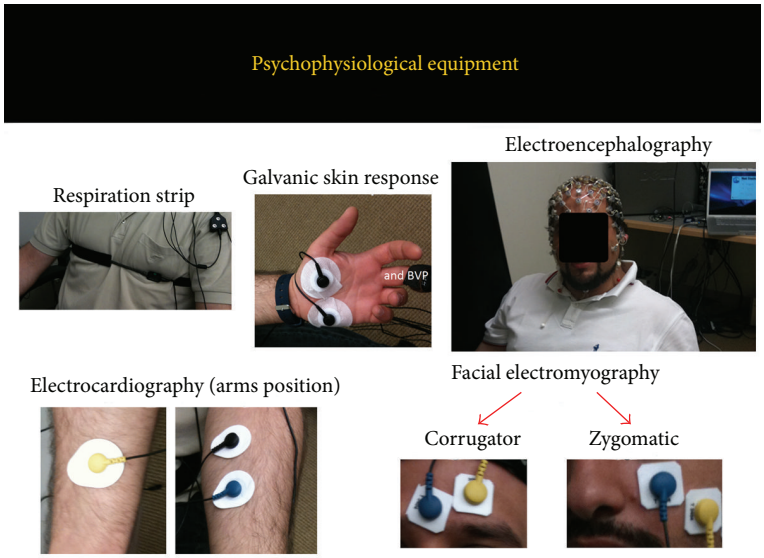


Figure 3: Psychophysiological equipment.

3. Psychophysiological Assessment

The multidimensional model aimed at measuring in an objective way the subjective experience. At this purpose we describe which biosensors and biomedical signals came to age to be considered consolidated enough to allow an objective measurement. More, it is to be taken into account that psychophysiological analysis is not easy and requires specific mathematical competences and not only sophisticated instruments; thus, following, we will give a short insight on the correct signal processing procedures necessary to extract the indexes (measures) that eventually can be used for the statistical data analysis.

3.1. Biosensors and Biomedical Signals. A number of biosensors and biomedical signals can be used; most biosensors are nowadays also wearable and their obtrusiveness is more and more reduced. Eye tracker acts at distance (about one meter); the other biosensors are electrodes-based, reading the electrophysiological signals by contact (see an example in Figure 3). The setting phase is simple but needs to be made by an expert researcher or physician to detect the exact locations or the signals extracted risk to be compromised.

Following a (nonexhaustive) list of typical biosensors/biomedical signals: electroencephalogram (EEG), galvanic skin response (GSR), electrocardiogram (ECG), blood volume pulse (BVP), respiration signal (RSP), eye tracker (ET), and facial electromyography (fEMG). In the next session a deeper insight on the biosensors use, and the sense of the extracted measures based on the multidimensional model will be given.

3.2. Signal Processing and Extracted Measures. Cardiovascular and respiratory activity is monitored to evaluate both voluntary and autonomic effect of respiration on heart rate, analyzing R-R interval extracted from electrocardiogram (ECG) and respiration (RSP) from chest strip sensor and their interaction. It is also possible to extract IBI (interbeat-interval) from blood volume pulse (BVP), that is an acceptable (even if worse) alternative to ECG's R-R. According to the guidelines of Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, typical heart rate variability (HRV) spectral indexes can be extracted to evaluate the autonomic nervous system response [16, 17, 69]. Spectral analysis can be performed using Fourier spectral methods. The rhythms can be classified as very low frequency (VLF, i.e., less than 0.04 Hz), low frequency (LF, from 0.04 to 0.15 Hz), and high frequency (HF, from 0.15 to 0.5 Hz) oscillations. This procedure allows us to calculate the LF/HF ratio, also known as the sympathovagal balance index. Cardiovascular and respiratory activity interaction can also be taken into account through Respiratory Sinus Arrhythmia (RSA) index [17, 69]. As temporal domain measures of heart rate variability are generally calculated NN50 index, that is, the number of interval differences of successive NN intervals greater than 50 milliseconds. This index describes the short-term NN variability. Just to simplify, NN intervals can be seen as a sort of beat-to-beat representation of heart rate; according to Camm and Colleagues [69], "In a continuous ECG record, each QRS complex is detected, and the so-called normal-to-normal (NN) intervals (that is, all intervals between adjacent QRS complexes resulting from sinus node depolarization) or the instantaneous heart rate is determined."

Skin conductance mean (SC) can be extracted from a GSR biosensor. It is critical to remove possible movement artifacts before computing the index (since on the hand, it can be affected by consistent involuntary grasping). SC is an interesting measure, since the sweat glands are regulated by the sympathetic nervous system without a direct "contamination" of parasympathetic nervous system (that for example exists for HR). Thus SC is an excellent candidate to measure pure physiological arousal [70, 71].

The raw electromyography (EMG raw) is a collection of positive and negative electrical signals; their frequency and amplitude give us information on the contraction or rest state of the muscle. Amplitude is measured in μV (microvolts). As the subject contracts the muscle, the number and amplitude of the lines increase; as the muscle relaxes, amplitude decreases [72–74]. It is generally considered the Root Mean Square (RMS) for rectifying the raw signal and converting it to an amplitude envelope [18, 75]. In particular cases we can also be interested in frequency, related to muscle fatigue [73]. There are a number of measures that can be extracted from this signal that depend on the muscle corresponding to the electrodes locations. For the model, there are three facial locations that give relevant information about emotional valence. In particular, RMS of EMG signal was recorded in correspondence of facial zygomatic major muscle (following EMG Zygomatic), that increases when positive emotions arise [72, 75]. On the other hand, the RMS of EMG signal recorded in correspondence was with facial corrugator supercilii muscle (following EMG Corrugator), that increases when negative emotions arise [72, 75]. Eventually, the RMS of EMG signal was recorded in correspondence of facial orbicularis oculi muscle underneath the eye with miniature electrodes muscle (following Startle Reflex), that is inversely proportional to the pleasantness of the stimuli [75].

Respiration signal can be elaborated to compute the respiration depth (RSP depth), the point of maximum inspiration minus the point of maximum expiration to be determined from the respiratory tracing. Smaller values indicate more shallow respiration and higher activation [18, 76]. It is also possible to calculate respiration rate (also measured in breaths per minutes) from peak-to-peak computing.

EEG signals need to be extensively worked to remove ocular artifacts and blinks, if possible basing on electrooculography (EOG) signals using automatic algorithm and subsequent visual inspection. Then the corrected matrixes can be computed to calculate means of the Beta EEG (e.g., 13–30 Hz) bands, of the Alpha EEG (e.g., 7–13 Hz) bands, and of the Slow Alpha EEG (e.g., 7–10 Hz) bands, one per each channel recorded, through spectral analyses [77–79]. Frontal EEG activation asymmetry has been generally used, giving evidences that greater left frontal activity seems to be higher related to positive emotional valence, whereas greater right frontal activity seems to be more involved in negative emotional valence [80]. Alpha index seems to be the most adapt to study the frontal EEG activation asymmetry [80]. Alpha Asymmetry index can be calculated in many different ways to take into account one hemispheric prevalence on the other one and correcting the sign accordingly. In calculating this index, it is crucial to consider that higher cortical activation is revealed by lower Alpha waves, and thus this needs to be considered in the computation and formula derivation. This Alpha Asymmetry is also a recognized index of depression [81–91] and hence can be used to measure life satisfaction and emotional valence.

Beta EEG bands (following Beta indexes) are often used to identify physiological arousal [79].

Using eye-tracker data, we can calculate the measure of cognitive and visual information processing, although they are limited in what they reveal about higher-order processes [92]. Eye movement data consist of moment-to-moment measures of the eyes' displacements along the vertical and horizontal axes (in mm) within the spatial working area of the monitor screen. The pupil size and gazes are acquired, based on the corneal reflection on the frontal surface of participants' eyes (caused by an infrared light source). After the experiment, the signals can be extracted and processed taking into account the blinks. The mean of pupil size (following pupil size) is considered as an important indicator of emotional arousal, that is, an arousal due to emotional stimuli, that is, one of the few indexes that take into account the physiological arousal as emotional consequences.

Every channel needs to be synchronously acquired at 2048 Hz and exported at last at 256 Hz sampling rate (256 records per second, one every 3.90625 millisecond). Some signals may be required to be extracted to a higher sampling rate (for example a minimum of 1024 is suggested for EMG signals).

To make interpretation relevant to actual users' affective state and to avoid contaminations, light and temperature sensors should be used to monitor the conditions of the room and, if possible, two three-axis accelerometers should be integrated into the biosensors and used to monitor subjects' stability and remove possible artifacts.

3.3. Synchronization and Epochs' Definition. Even if the use of eye tracker in the model is totally justified from pupil and gazes analyses are allowed by this tool, it is also crucial to underline its usefulness for synchronizing the psychophysiological signals within the experimental epochs. In fact, such synchronization become important when it is critical the timing for the presentation of the stimuli. In these cases, it become crucial to synchronize psychophysiological signals with eye-tracker data and to synchronize all these data with the sequences of stimuli presented to the subjects.

Usually, one way to overcome this problem is represented by the use of a webcam to record the stimuli screen or through a video screen capture program. However, these methods are not precise. In fact, even if the stimuli are synchronized with a computer clock, they require the visualization of a video to establish the periods, but this affects the time acquisition due to normal video latency. It is better, using the eye-tracker data extraction, to obtaine for each participant a matrix of gaze and pupil data corresponding to stimuli presentation, in particular, to collect a number of rows for each second (depending on the sampling rate used), thereby making it possible to establish the exact periods previously indicated.

A second step is represented by the synchronization of the stimuli with psychophysiological signals. In this case, it can use algorithms to synchronize eye-tracker systems with a psychophysiological device by using a photodiode, which can also be configured through a physical channel on the equipment used, just capturing the light (i.e., by identifying black and white). Practically, thanks to this photodiode actually applied on the screen and an algorithm [93]. Moreover, based on gazes and pupil signals acquired, it is possible to identify eye blinks, which enable us to align the matrixes containing the eye-blink data from gazes and pupil signals,

with the matrixes containing the psychophysiological signals. Thanks to these procedures, it is possible to synchronize all signals and to correctly identify the experimental epoch, with an error of ±0.01 second [94].

To synchronize the presentation of objects with electrophysiological recordings, an interesting tool to keep in consideration is the box for interaction with objects (BIO) [95].

3.4. Personal Data Archiving. All participants data need to be memorized in encrypted and password protected files, possibly following the criteria to protect personal health information [96] and using PsychoPass or improved methods [97, 98] to generate and share passwords information among pairs.

4. Data Analysis

The model produces three successive observations of the same variable (measure considered for the analysis) on each subject. Repeated measures are defined as measurements sequentially conducted in time (temporal factor) or location (spatial factor) on the same subject. Repeated measurements are commonly employed to estimate measure parameters, investigate the factor effect on the process, and model and monitor the production and its process [99].

It is highly suggested to follow the recommendation of Bakker and Wicherts in reporting statistical results [100].

4.1. Repeated Measure Analysis of Variance. Repeated measure analysis of variance (rmANOVA, also known as ANOVARM) design requires three basic assumptions: (1) normal distribution of measures, (2) independent samples (if a between variable is taken into account, for example: depressed subjects versus nondepressed subjects), (3) homoscedasticity (equal variances of measures).

Additional assumptions are needed for rmANOVA as a result of the presence of correlations between measurements taken on the same subject at different levels (time, space, order, etc.).

In particular, an additional assumption to make the F-test of the repeated-measures valid is the sphericity (assumption of compound symmetry, that is, circularity of variance-covariance matrix) requiring homogeneity of the covariances among repeated measures.

A significant value for Mauchly's test of sphericity at P level .05 indicates that the assumption of homogeneity of covariance has been violated for some measures.

Commonly, two correction methods are used for adjusting to the F-test in terms of degree of freedom: the Huynh-Feldt and the Greenhouse-Geisser test [101, 102]. In these cases, corrected P values need to be reported accordingly.

Girden [103] recommended that if epsilon (Greenhouse-Geisser estimate) is larger than 0.75, then the correction according to Huynh and Feldt should be used. On the other hand, if epsilon is smaller than 0.75, then the more conservative correction according to Greenhouse-Geisser is preferred.

4.2. Pairwise Comparisons. A common error on data analysis is the use of paired-samples t-test to compare couples of repeated measures. In the model, it is necessary to have a precise idea of each dimension, by comparing "relax versus happiness" and "happiness versus stress," through the use of pairwise comparison adjusting the alpha level to avoid an inflated type I error rate making multiple statistical comparisons (using, for example, Bonferroni correction). Most statistical softwares used for behavioral sciences have an embedded tool to correct these values.

It is also possible to use contrasts, based on F test, for comparison. In particular, simple contrasts (baseline versus each other level) or repeated contrasts (comparison of adjacent levels) can be used.

4.3. Polynomial Contrasts. Polynomial a priori contrasts can be computed by testing the hypothesized quadratic trends for main effects in physiological arousal (hypothesis 1) measures, with higher values for the happiness and stress epochs in comparison with the relax epoch. Polynomial a priori contrasts can also be performed by testing the hypothesized quadratic trends for main effects in emotional valence (hypothesis 2) and life satisfaction (hypothesis 3) measures, with positive values for the relax and happiness epochs in comparison with the stress epoch.

A monotonic trend with measures increasing (hypothesis 1) or decreasing (Hypotheses 2 and 3) from the relax to happiness to stress epochs is also expected.

4.4. Sample Size, Sensitivity, and Post Hoc Power Analysis. Hypothesis test tells us the probability of a result of that magnitude occurring, if the null hypothesis is correct (i.e., there is no effect in the population). It does not tell us the probability of that result, if the null hypothesis is false (i.e., there actually is an effect in the population).

Specifically, we consider the effect size, the sample size, and the criterion required for significance (α, where α is probability of type I error). These three factors, together with power ($1 - \beta$, where β is probability of type II error), form a closed system; once any three are established, then the fourth is completely determined [104].

A sample size calculation for the model experimental design, based on rmANOVAs, leaded an estimation of minimum 28 subjects to be used for possible experiments, in order to achieve a minimum power of 0.8, considering a medium effect size of 0.25 and a significance level of 0.05, and sphericity assumption satisfied (see Table 2) [104]. As can also be seen from Figure 4, a lower effect size leads to a necessary increase of sample size to achieve the same minimum power.

Once the results are computed, a power analysis can be used to anticipate the likelihood that the study yielded significant effects. In particular, the goal of a post hoc power analysis is to compute achieved power, given the effective other three factors, which can be read or deducted by data (output of statistical data analysis). Since many statistical softwares give $p\eta^2$ (partial eta-square) values instead of cohen's f effect size, it is important to compute $f = \text{sqrt}[\eta^2/(1 - \eta^2)]$, where sqrt is for square root calculation.

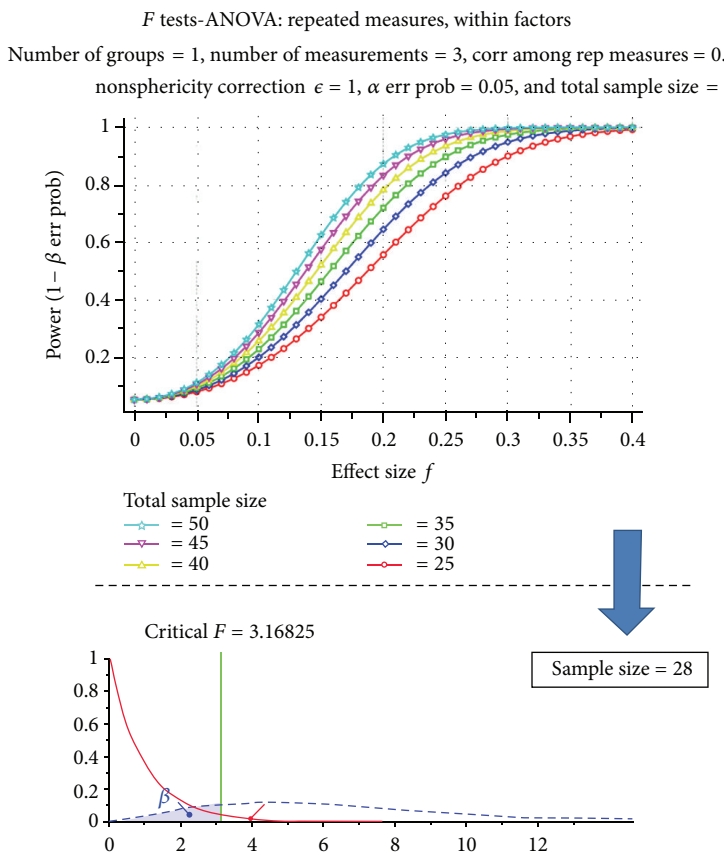


Figure 4: Lower effect sizes lead to a necessary increase of sample size to achieve the same minimum power. Estimated sample size from a literature analysis is of at least 28 participants.

Table 2: A priori analysis to compute the required sample size in a repeated measures ANOVA.

Input		Output	
Effect size f	0.25	Noncentrality parameter λ	10.5
α err prob	0.05	Critical F	3.168246
Power ($1-\beta$ err prob)	0.8	Numerator df	2
Number of groups	1	Denominator df	54
Number of measurements	3	Total sample size	28
Corr among rep measures	0.5	Actual power	0.8124546
Nonsphericity correction ε	1		

According to post hoc power analysis, some significance level could be high informative even if slightly higher than .05 (it depends on achieved $p\eta^2$ for that measure).

5. Hypotheses Testing

As explained and reported in Table 1, three hypotheses are given for the model. Each hypothesis refers to one dimension of the model. A detail on expected measures within each hypothesis/dimension is presented, and comparisons with the ground truths are discussed. Consideration on statistical significance and polynomial contrasts is also taken into account. Table 3 summarizes the measures within each hypothesis.

5.1. Hypothesis 1: Physiological Arousal. HR, SC, Beta indexes, and pupil size should be lower in relax epoch and higher in stress epoch; respiration depth for physiological arousal is in the smaller-is-higher form; consequently it is expected to be higher in relax epoch and lower in stress epoch, because relax produces lower physiological arousal and stress produces higher physiological arousal. Values of indexes for physiological arousal during happiness should be more similar to the ones in stress epoch. Practically, in both stress and happiness, the subject is in a situation of elevated physiological activation. Thinking of positive and engaging situation, such as gaming or other highly involving situations, it is easy to understand this state of higher activation in happiness.

Table 3: Hypotheses per each measure (− is for lower and + is for higher).

Dimension	Measure	Biomedical signal	Session type			Pairwise comparisons (Bonferroni correction)	
			Relax	Happiness	Stress	Happiness versus Relax	Happiness versus Stress
Higher arousal	HR	ECG	−	+	+	Sig.	−
	SC	GSR	−	+	+	Sig.	−
	Beta indexes	EEG	−	+	+	Sig.	−
	Pupil size	ET	−	+	+	Sig.	−
	Respiration depth	RSP	+	−	−	Sig.	−
Positive valence	EMG zygomatic	EMG	+	+	−	−	Sig.
	EMG Corrugator	EMC	−	−	+	−	Sig.
	Startle reflex	EMG	−	−	+	−	Sig.
	EEG Alpha Asymmetry	EEG	+	+	−	−	Sig.
Higher satisfaction	LF/HF	ECG	−	−	+	Sig.	Sig.
	NN50	ECG	+	+	−	Sig.	Sig.
	RMSSD	ECG	+	+	−	Sig.	Sig.
	HF power	ECG	+	+	−	Sig.	Sig.
	LF power	ECG	−	−	+	Sig.	Sig.

Repeated measures ANOVAs can be used with epoch type (relax, happiness, and stress) as the within-subject variable for all the indexes used to measure physiological arousal. A main statistical significant effect of epoch type is expected for all the measures. Pairwise comparisons using the Bonferroni (or others) corrections can be used to reveal if there are statistically significant differences between relax epoch and happiness for all the indexes and no statistically significant differences between happiness and stress epoch for all the indexes (i.e., these two epochs are supposed to be so similar to produce no differences in arousal activation).

Polynomial a priori contrasts resulting by testing the hypothesized quadratic trends for main effects, with lower value of respiration depth and higher values for all the other indexes, can be used to measure physiological arousal for the happiness and stress epoch in comparison with the Relax epoch.

5.2. Hypothesis 2: Emotional Valence. EMG Zygomatic and Alpha Asymmetry measures should be higher in relax epoch and lower in stress epoch; EMG Corrugator and Startle Reflex measures for emotional valence are in the smaller-is-higher form; consequently, they are expected to be lower in relax epoch and higher in stress epoch, because relax elicits positive emotional valence and stress elicits negative emotional valence.

Values of indexes for emotional valence during happiness should be more similar to the ones in relax epoch. Practically, in both relax and happiness experience, the subjects are in a situation of positive emotional valence, that is, pleasantness. Being in relaxing situations or being in happiness experience leads to have positive emotional valence.

Repeated measures ANOVAs can be used with epoch type (relax, happiness, and stress) as the within-subject variable for all the indexes was used to measure emotional valence. A main statistically significant effect of epoch type is expected for all the measures. Pairwise comparisons using the Bonferroni (or others) corrections can be used to reveal if there are statistically significant differences between happiness and stress epoch for all the indexes and no statistically significant differences between Relax epoch and Happiness for all the indexes (i.e., these two epochs are supposed to be so similar to produce no differences in emotional valence).

Polynomial a priori contrasts resulting by testing the hypothesized quadratic trends for main effects, with higher values of EMG zygomatic and Alpha Asymmetry indexes and lower values for EMG Corrugator and Startle Reflex indexes can be used to measure emotional valence for the relax epoch and happiness in comparison with the stress epoch.

5.3. Hypothesis 3: Life Satisfaction. LF/HF and LF power measures are supposed to be lower in relax epoch and higher in stress epoch; NN50, RMSSD, and HF power measures for depression are in the smaller-is-higher form; consequently, they are expected to be higher in relax epochs and lower in stress epoch, because relax produces lower anxiety and stress produces higher anxiety. Since depression is negatively related to life satisfaction, the above indexes will be exactly the opposite to measure higher happiness.

Happiness is characterized by a low level of depression, resulting having a high level of life satisfaction, that is, different from the relax and stress states.

Repeated measures ANOVAs can be used with epoch type (relax, happiness, and stress) as the within-subject variable for all the indexes was used to measure depression. A main statistically significant effect of epoch type is expected for all the measures. Pairwise comparisons using the Bonferroni (or others) corrections can be used to reveal if there are

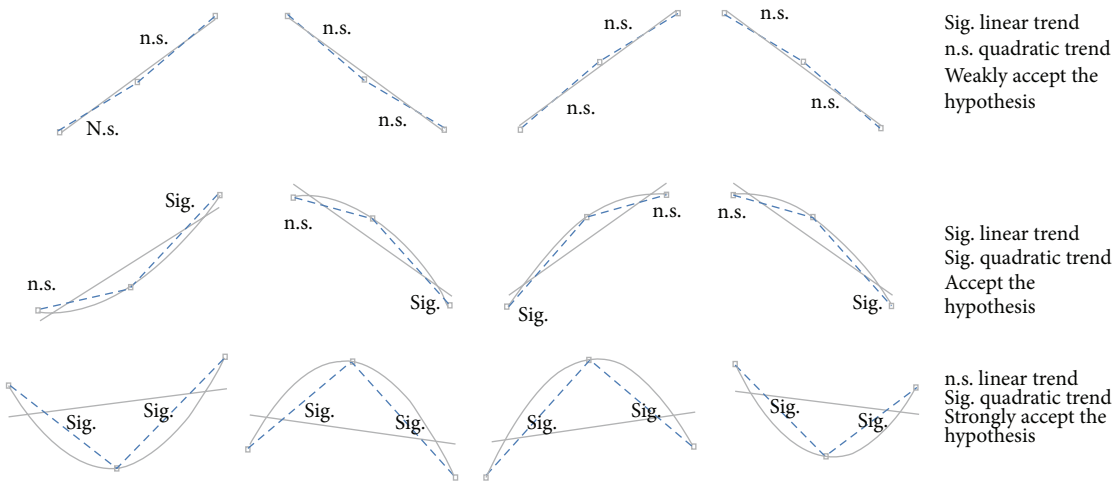


Figure 5: Schema of the listed situations and graphical representations of the linear-quadratic trend coexistence. Here, we illustrate linear or quadratic relationships among the three states regardless of Cartesian representation.

statistically significant differences between happiness and both relax and stress epochs for all the indexes.

Polynomial a priori contrasts resulting by testing the hypothesized quadratic trends for main effects, with higher values of NN50, RMSSD, and HF power indexes and lower values of LF/HF and LF power, can be used to measure lower life satisfaction for the relax and stress epochs compared to happiness.

5.4. Linear-Quadratic Trend Coexistence. A linear monotonic significant trend with measures changing from the relax to happiness to Stress epochs may also result for all the indexes considered in physiological arousal and emotional valence but not in life satisfaction (where only a quadratic form is expected, being the two extremes—relax and stress—at the same low level and the happiness to a higher level, by definition).

There are three possible situations of statistically significant polynomial contrasts:

(1) only linear trend is statistically significant;

(2) both linear and quadratic trends are statistically significant;

(3) only quadratic trend is statistically significant.

Figure 5 reports a schema of the listed situations and graphical representations of the linear-quadratic trend coexistence.

The first situation, with only statistically significant linear trend, in the model denotes a situation, where the happiness is in the middle between relax and stress epochs, but is not clear if it is closer to one or the other. Generally such a situation comparing the happiness with the other two states (relax and stress) leads to statistically significant differences in both the directions and becomes even more difficult to make a decision. In these cases, it becomes relevant to have more than one measure for the considered dimension, in order to strengthen the possible acceptance or rejection of the hypothesis.

In the second situation, both linear and quadratic trends are statistically significant. This denotes that there are two consecutive states with similar values, in our model "relax and happiness" or "happiness and stress." In this situation comparing the happiness with the other two states (relax and stress), it is probable to have statistically significant differences in only one direction, which would make it more easier to understand the happiness state. In this case, linear and quadratic trends and statistical significances provide more information about the closeness of the happiness to each conditions ("closer to relax" or "closer to stress"), strengthening the decision.

In the third situation listed before, an elevated increase in quadratic trend may lead to the loss of the monotonic trend and no statistical significant linear trend. In the model when this happens leads to strength the hypothesis. For example, let us consider the following scenario situation. Emotional valence measured through the EMG zygomatic shows (1) a significant quadratic trend and no significantly linear trend; (2) from simple (or repeated) contrasts (F-test based) or from the pairwise comparison (t-test based, with correction) result, statistically significant differences between happiness and stress and nonstatistically significant differences between relax and happiness, that is correct being EMG Zygomatic and index of emotional valence; (3) from descriptive results that the values for relax and happiness are higher than the values in stress, if not it means that EMG zygomatic is behaving like an EMG Corrugator and a deeper intelligence on signal processing or channels naming would be very suggested.

In the scenario situation just described the EMG Zygomatic is so high during happiness experience to generate a strong quadratic trend. This let us suppose, practically, that emotional valence has been more positive during the happiness experience, that actually strengthen our hypothesis.

Figure 6 shows the dynamics of a happiness state, *ceteris paribus*. While the happiness moves toward a new state, the quadratic trend increases (and its significance level decreases)

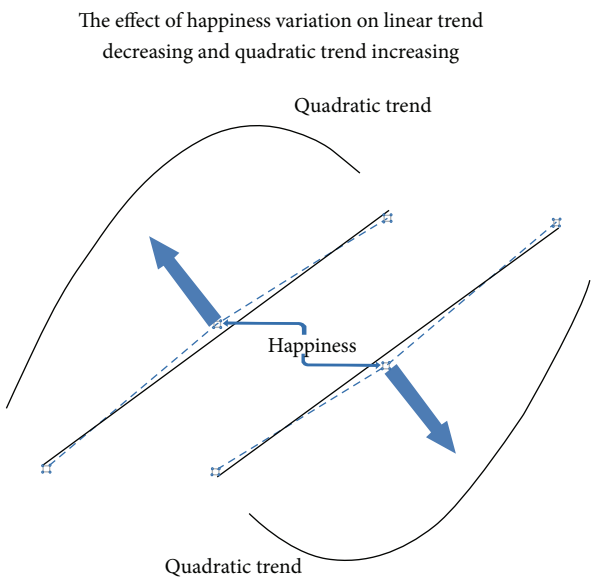


Figure 6: Dynamics of a happiness state, *ceteris paribus*. Here we illustrate linear or quadratic relationships among the three states regardless of Cartesian representation.

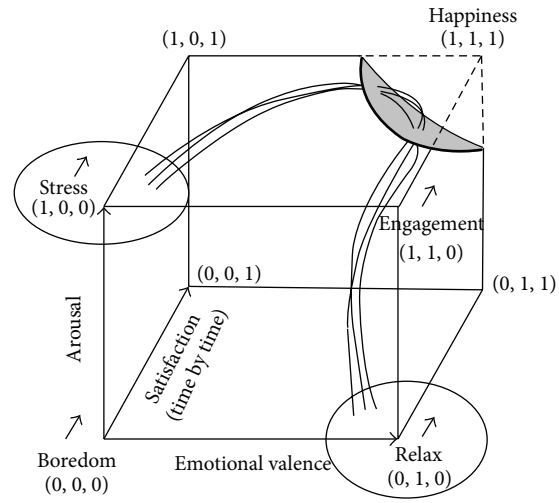


Figure 7: A representation of possible paths from relax to happiness experience to stress states.

and the linear trend decreases (and its significance level increases).

5.5. Other Elements against or Supporting the Hypotheses. In the multidimensional model, it is necessary to evaluate the strength or weakness of each hypothesis, working on the several dimensions step by step, one dimension per time.

Within each hypothesis, the possible weakness of a measure needs to be considered in a wider context. In the case of unexpected values of a measures beyond the typical considerations that arise from the intrinsic imperfection of statistics (remembering the tails of a normal distribution), it is crucial to consider the effect that a dimension may have on the measure of another dimension. Typical example is the HR index that, even if it is recognized as a physiological arousal measure, may show unexpected values due to the effect of the baroreceptor reflex that causes heart rate to decrease also producing a variation in sympathovagal balance and in particular in a part of the LF, known as Mayer waves, at 0.1 Hz [105]. This means that there is a complex interaction between physiological arousal and depression that may produce adjustment to a part of HR index. This becomes a great limitation if complex interaction between variables are not checked by the means of correct indexes in all the dimension. In the example of HR, a lower than expected HR needs to be inspected using the several cardiovascular measures available, even if calculated for another dimension (depression and life satisfaction, in this case).

6. Discussion

We presented a multidimensional model to measure happiness by the means of psychophysiological correlates.

Dimension considered was physiological arousal, emotional valence, and life satisfaction. Psychophysiological measures can be an extremely useful source of knowledge in each domain considered. A correct and complete statistical analysis based on rmANOVAs and polynomial contrasts can be used to create a map of the affective states and detect in a rigorous empirical way if happiness arose. Figure 7 is a representation of possible paths from relax to happiness experience to stress states.

Future works should focus on a deeper classification of the dimensions and their relationships; in fact, it is crucial to investigate measures in a dimension and also in a wider view, considering the effect of a dimension on another dimension.

The multidimensional model is based on a statistical approach. Practically three sessions of physiological data are collected: during a relax epoch, during a stress epoch, and during the condition to verify, supposed to be during a happiness experience.

The disadvantage of such an approach is that it is necessary to have well-ested ground truths to which the experimental condition can be compared. However we saw that many databases with well-classified affective states are available and use our same classifications, among the others (IAPS, IADS, etc.).

Another limit of this approach is that the results, to be robust, need a large sample, of minimum 28 participants, but better if they are more than 35 since the effect size could be less than 0.25 once actually calculated. This is a practically a huge problem; in fact doing an experiment with psychophysiological signals is really demanding, requiring big efforts and attention during the recording phase and huge works and abilities *a posteriori*. In fact, the signal processing phase may also take long time, for at least two reasons: (1) the huge amount of data, just to give an idea 5 minutes of recording sampled at 256 Hz (it is the minimum, but some signals require also 1024 or more), means a matrix with a minimum of 76,800 rows and one column per each channel recorded, that then need to be preprocessed, filtered, and computed

for a certain number of indexes extracted; (2) most signals need to be visually inspected for a corrected signal processing procedure; for example, from ECG signals it is necessary to extract R-R waves to compute cardiovascular indexes that we described; however, a good automatic detection algorithm may detect correctly 95% of R-peaks, and this means that the researcher needs to look at all the signals and correct the problems accordingly. Due to these considerations and to the general complexity of psychophysiological experiments, to collect data from 28 participants or more is not simple at all and often requires a well-consolidated team devoted to it.

On the other hand the advantage of such a complex approach is that when the experiments are conducted in a rigorous way, the results describe the subject reactions objectively allowing the researchers to achieve meaningful conclusions, of course in the limit of the study.

More, the increased computational capacity and the huge advancement in the field of artificial intelligence made another approach available that also gained a good credibility, that is, the affective computing. According to Rosalind Picard, who coined the term, "affective computing is computing that relates to, arises from, or deliberately influences emotion or other affective phenomena." [106].

But how affective computing is related to our pursuit of happiness? Which advantages may offer? To answer these questions we need to determine to which extent the affective computing approach is different from the statistical one.

Let us imagine that we want to build an mp3 reader that automatically plays Mozart when you are stressed, recognizing if you become happy. Is that possible? This is a typical research question that arises from affective computing approach. The main difference from a statistical approach is that the target in this case is the real-time monitoring of a single subject's affective states (e.g., in Figure 8 is represented an SC signal processed in real time and the computed Fit function, based on a sum of sinusoids model ($f(x) = a1 * \sin(b1 * x + c1) + a2 * \sin(b2 * x + c2) + a3 * \sin(b3 * x + c3) + a4 * \sin(b4 * x + c4) + a5 * \sin(b5 * x + c5) + a6 * \sin(b6 * x + c6) + a7 * \sin(b7 * x + c7) + a8 * \sin(b8 * x + c8)$.)). In this case the signal analysis can be an automatic and continuous process that collects, seconds-by-seconds, physiological signals, classifies them, and gives an immediate output of the pattern recognition to the subject; if not correct the system may be instructed to autocorrect and improve its recognition algorithms, by the means of data mining techniques.

Of course affective computing is a fascinating field; however, it contains intrinsic limitations, mainly due to lacking in classification. It is in fact complex to recognize an affective state after months of signal processing and data analysis, let imagine in real time. However the big advantage of affective computing is that it fostered the development of a plenty of classification methods with a vivid international discussion at really high scientific levels. The continuous development of new artificial intelligence techniques further enriches this scenario.

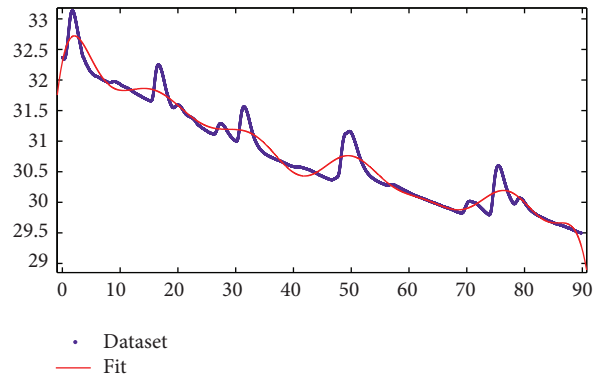


Figure 8: An SC signal processed in real time with a Fit function, based on a sum of sinusoids ($f(x) = a1 * \sin(b1 * x + c1) + a2 * \sin(b2 * x + c2) + a3 * \sin(b3 * x + c3) + a4 * \sin(b4 * x + c4) + a5 * \sin(b5 * x + c5) + a6 * \sin(b6 * x + c6) + a7 * \sin(b7 * x + c7) + a8 * \sin(b8 * x + c8)$.). Horizontal axis represents time (in seconds) and vertical axis represents the SC level (in microsiemens (μS)).

At this very point, thus, the real question is if it can be possible to integrate the two approaches for a better measurement of the experience and a better detection of happiness.

A possible approach could follow these steps: (1) an experiment with a statistical approach, to create a model of the experience based on random sampling; (2) data reduction and extraction, that is, on the basis of the results obtained with the experiments to find the most informative measures for that specific happiness experience; and (3) experience tracking and automatic detection of happiness experience, based on the specific measures extracted. Once the measures are extracted this can constitute a new sample for statistical analysis for the perfection of the process.

We created in this way a closed loop of analysis aimed at fostering a better detection and measurement of happiness experiences.

However, also this approach presents some limitations; in fact while the first phase is a standard experiment in a laboratory setting, the last phase requires a naturalistic setting to be really effective, at least for many kinds of tracked experiences.

Naturalistic setting means considering "real mobile" setting, that is, *in vivo* experiments out of the lab, using a computerized ecological momentary assessment [107–109], with the combined integration of wearable biosensors.

The last challenge that we need to consider for the multidimensional model presented is the possible use of other measures collected by the means of less obtrusive biosensors, however being objective and reliable. Unobtrusiveness in data collection will probably represent the next most exciting challenge in psychophysiology.

Conflict of Interests

The authors declare that there is no conflict of interests.

References

[1] Congress, U.S.C., "A declaration by the representatives of the United States of America," in *Proceedings of the General Congress Assembled*, Philadelphia, Pa, USA, July 1776.

[2] J. A. Russell, "Affective space is bipolar," *Journal of Personality and Social Psychology*, vol. 37, no. 3, pp. 345–356, 1979.

[3] P. J. Lang, "The emotion probe: studies of motivation and attention," *The American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.

[4] J. D. Morris, "Observations: SAM: the self-assessment manikin—an efficient cross-cultural measurement of emotional response," *Journal of Advertising Research*, vol. 35, no. 6, pp. 63–68, 1995.

[5] R. B. Rubin, A. M. Rubin, E. Graham, E. M. Perse, and D. Seibold, "Self-Assessment Manikin," in *Communication Research Measures II: A Sourcebook*, pp. 336–341, 2009.

[6] R. W. Backs, S. P. Da Silva, and K. Han, "A comparison of younger and older adults' self-assessment Manikin ratings of affective pictures," *Experimental Aging Research*, vol. 31, no. 4, pp. 421–440, 2005.

[7] P. Cipresso, S. Serino, D. Villani et al., "Is your phone so smart to affect your state? An exploratory study based on psychophysiological measures," *Neurocomputing*, vol. 84, pp. 23–30, 2012.

[8] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, 2000.

[9] A. Keil, M. M. Bradley, O. Hauk, B. Rockstroh, T. Elbert, and P. J. Lang, "Large-scale neural correlates of affective picture processing," *Psychophysiology*, vol. 39, no. 5, pp. 641–649, 2002.

[10] D. Grühn and S. Scheibe, "Age-related differences in valence and arousal ratings of pictures from the international affective picture system (LAPS): do ratings become more extreme with age?" *Behavior Research Methods*, vol. 40, no. 2, pp. 512–521, 2008.

[11] B. Kuhr, J. Jacobi, C. Krause et al., "Arousal, valence, dominance...and desire? Evidence from an Erp study concerning the necessity of a new motivational dimension to describe affective states," *Psychophysiology*, vol. 48, p. S88, 2011.

[12] B. Rozenkrants and J. Polich, "Affective ERP processing in a visual oddball task: arousal, valence, and gender," *Clinical Neurophysiology*, vol. 119, no. 10, pp. 2260–2265, 2008.

[13] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Real-time EEG-based happiness detection system," *The Scientific World Journal*, vol. 2013, Article ID 618649, 12 pages, 2013.

[14] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio, "Basic emotions are associated with distinct patterns of cardiorespiratory activity," *International Journal of Psychophysiology*, vol. 61, no. 1, pp. 5–18, 2006.

[15] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, *Handbook of Psychophysiology*, Cambridge University Press, Cambridge, UK, 3rd edition, 2007.

[16] M. Mauri, P. Cipresso, A. Balgera, M. Villamira, and G. Riva, "Why is Facebook so successful? Psychophysiological measures describe a core flow state while using Facebook," *Cyberpsychology, Behavior, and Social Networking*, vol. 14, no. 12, pp. 723–731, 2011.

[17] V. Magagnin, M. Mauri, P. Cipresso et al., "Heart rate variability and respiratory sinus arrhythmia assessment of affective states by bivariate autoregressive spectral analysis," *Computers in Cardiology*, vol. 37, no. 5737930, pp. 145–148, 2010.

[18] M. Mauri, V. Magagnin, P. Cipresso et al., "Psychophysiological signals associated with affective states," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '10)*, pp. 3563–3566, Buenos Aires, Argentina, September 2010.

[19] M. Argyle, *The Psychology of Happiness*, Routledge, London, UK, 2nd edition, 2001.

[20] S. A. David, I. Boniwell, and A. C. Ayers, *The Oxford Handbook of Happiness*, Oxford Library of Psychology, Oxford University Press, Oxford, UK, 1st edition, 2013.

[21] S. Achor, *The Happiness Advantage: The Seven Principles of Positive Psychology That Fuel Success and Performance at Work*, Broadway Books, New York, NY, USA, 1st edition, 2010.

[22] A. S. Waterman, "Two conceptions of happiness: contrasts of personal expressiveness (Eudaimonia) and hedonic enjoyment," *Journal of Personality and Social Psychology*, vol. 64, no. 4, pp. 678–691, 1993.

[23] C. O. Walker, T. D. Winn, and R. M. Lutjens, "Examining relationships between academic and social achievement goals and routes to happiness," *Education Research International*, vol. 2012, Article ID 643438, 7 pages, 2012.

[24] B. Headey, J. Kelley, and A. Wearing, "Dimensions of mental health: life satisfaction, positive affect, anxiety and depression," *Social Indicators Research*, vol. 29, no. 1, pp. 63–82, 1993.

[25] P. Steca, A. Greco, D. Monzani et al., "How does illness severity influence depression, health satisfaction and life satisfaction in patients with cardiovascular disease? The mediating role of illness perception and self-efficacy beliefs," *Psychology and Health*, vol. 28, no. 7, pp. 765–783, 2013.

[26] R. B. Nes, N. O. Czajkowski, E. Røysamb, R. E. Orstavik, K. Tambs, and T. Reichborn-Kjennerud, "Major depression and life satisfaction: a population-based twin study," *Journal of Affective Disorders*, vol. 144, no. 1-2, pp. 51–58, 2013.

[27] P. B. Gnilka, J. S. Ashby, and C. M. Noble, "Adaptive and maladaptive perfectionism as mediators of adult attachment styles and depression, hopelessness, and life satisfaction," *Journal of Counseling and Development*, vol. 91, no. 1, pp. 78–86, 2013.

[28] M. Eskin, A. Akyol, E. Y. Çelik, and B. K. Gültekin, "Social problem-solving, perceived stress, depression and life-satisfaction in patients suffering from tension type and migraine headaches," *Scandinavian Journal of Psychology*, vol. 54, no. 4, pp. 337–343, 2013.

[29] J. Yorgason, H. Choi, K. Gustafson, W. Godfrey, and A. Bond, "Daily associations between family and community support with daily levels of depression, anxiety, and life satisfaction," *Gerontologist*, vol. 52, pp. 496–496, 2012.

[30] A. Greco, D. Monzani, L. Pancani, E. R. Cappelletti, M. D'Addario, and P. Steca, "Relationships of illness severity with depression, health- and life-satisfaction in patients with cardiovascular diseases," *Psychology and Health*, vol. 27, pp. 52–52, 2012.

[31] K. Garver, K. Bogda, R. Westrup et al., "Effects of resistance training on mood, life satisfaction, and depression," *Gerontologist*, vol. 52, pp. 587–587, 2012.

[32] C. Fagerström, M. Lindwall, A. I. Berg, and M. Rennemark, "Factorial validity and invariance of the life satisfaction Index in older people across groups and time: addressing the heterogeneity of age, functional ability, and depression," *Archives of Gerontology and Geriatrics*, vol. 55, no. 2, pp. 349–356, 2012.

[33] P. C. Britton, P. C. Ouimette, and R. M. Bossarte, "The effect of depression on the association between military service and life

satisfaction," *Quality of Life Research*, vol. 21, no. 10, pp. 1857–1862, 2012.

[34] O. N. Bamishigbin, C. Carver, R. Spillers et al., "Effects of optimism, benefit-finding, and spirituality on depression and life satisfaction in cancer caregivers," *Annals of Behavioral Medicine*, vol. 43, p. S10, 2012.

[35] K. Yamasaki, M. Sasaki, K. Uchida, and L. Katsuma, "P02-109—effects of positive and negative affect and emotional supression on short-term life satisfaction and depression: considering activation of affect," *European Psychiatry*, vol. 26, supplement 1, no. 705, 2011.

[36] B. Grinde, "An evolutionary perspective on the importance of community relations for quality of life," *TheScientificWorldJOURNAL*, vol. 9, pp. 588–605, 2009.

[37] R. M. Carney, M. W. Rich, A. TeVelde, J. Saini, K. Clark, and K. E. Freedland, "The relationship between heart rate, heart rate variability and depression in patients with coronary artery disease," *Journal of Psychosomatic Research*, vol. 32, no. 2, pp. 159–164, 1988.

[38] V. K. Yeragani, R. Pohl, R. Balon et al., "Heart rate variability in patients with major depression," *Psychiatry Research*, vol. 37, no. 1, pp. 35–46, 1991.

[39] K. Yeragani, R. Berger, R. Pohl et al., "Heart-rate-variability in patients with panic disorder and depression," *Biological Psychiatry*, vol. 33, no. 6, p. A146, 1993.

[40] J. Runions, M. V. Kamath, and E. L. Fallen, "Heart-rate-variability and depression—a linkage," *Circulation*, vol. 90, no. 4, pp. 591–591, 1994.

[41] V. K. Yeragani, R. Balon, R. Pohl, and C. Ramesh, "Depression and heart rate variability," *Biological Psychiatry*, vol. 38, no. 11, pp. 768–770, 1995.

[42] C. Choi, K. Kim, C. Kim, S. H. Kim, and W. Choi, "Reactivity of heart rate variability after exposure to colored lights in healthy adults with symptoms of anxiety and depression," *International Journal of Psychophysiology*, vol. 79, no. 2, pp. 83–88, 2011.

[43] B. A. Song, S. Yoo, H. Kang et al., "Post-traumatic stress disorder, depression, and heart-rate variability among North Korean defectors," *Psychiatry Investigation*, vol. 8, no. 4, pp. 297–304, 2011.

[44] I. Tonhajzerova, I. Ondrejka, Z. Turianikova et al., "P01-359—heart rate variability in adolescent major depression," *European Psychiatry*, vol. 26, supplement 1, no. 361, 2011.

[45] A. Celik, A. Ozturk, K. Ozbek, K. Ceyhan, H. Kadi, and F. Koc, "The value of heart rate variability and turbulence for discriminating the true coronary artery disease from false positive results in patients with ST segment depression without angina during exercise stress testing," *Circulation*, vol. 125, no. 19, p. E789, 2012.

[46] A. H. Kemp, "Reply to: are antidepressants good for the soul but bad for the matter? Using noninvasive brain stimulation to detangle depression/antidepressants effects on heart rate variability and cardiovascular risk," *Biological Psychiatry*, vol. 71, no. 7, pp. e29–e30, 2012.

[47] A. H. Kemp, D. S. Quintana, K. L. Felmingham, S. Matthews, and H. F. Jelinek, "Depression, comorbid anxiety disorders, and heart rate variability in physically healthy, unmedicated patients: implications for cardiovascular risk," *PLoS ONE*, vol. 7, no. 2, Article ID e30777, 2012.

[48] D. S. Mennin, D. M. Fresco, and A. Aldao, "Phasic heart rate variability changes predict clinical outcomes of emotion regulation therapy for generalized anxiety disorder and comorbid depression," *Psychophysiology*, vol. 49, p. S84, 2012.

[49] P. S. Munk, K. Isaksen, K. Brønnick, M. W. Kurz, N. Butt, and A. I. Larsen, "Symptoms of anxiety and depression after percutaneous coronary intervention are associated with decreased heart rate variability, impaired endothelial function and increased inflammation," *International Journal of Cardiology*, vol. 158, no. 1, pp. 173–176, 2012.

[50] S. M. Benvenuti, E. Patron, G. Favretto, R. Gasparotto, and D. Palomba, "Depression and reduced heart rate variability after cardiac surgery: the mediating role of emotion regulation," *Psychophysiology*, vol. 50, p. S108, 2013.

[51] C. B. Harte, G. I. Liverant, D. M. Sloan et al., "Association between smoking and heart rate variability among individuals with depression," *Annals of Behavioral Medicine*, vol. 46, no. 1, pp. 73–80, 2013.

[52] H. T. Huang and K. S. Wan, "Heart rate variability in junior high school students with depression and anxiety in Taiwan," *Acta Neuropsychiatrica*, vol. 25, no. 3, pp. 175–178, 2013.

[53] S. Suh, R. J. Ellis, J. J. Sollers 3rd, J. F. Thayer, H. C. Yang, and C. F. Emery, "The effect of anxiety on heart rate variability, depression, and sleep in chronic obstructive pulmonary disease," *Journal of Psychosomatic Research*, vol. 74, no. 5, pp. 407–413, 2013.

[54] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): affective ratings of pictures and instruction manual," Tech. Rep. A-8, University of Florida, Gainesville, Fla, USA, 2008.

[55] M. M. Bradley and P. J. Lang, "International affective digitized sounds (IADS): stimuli, instruction manual and affective ratings," Tech. Rep. B-2, The Center for Research in Psychophysiology, University of Florida, Gainesville, Fla, USA, 1999.

[56] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): stimuli, instruction manual and affective ratings," Tech. Rep. C-1, The Center for Research in Psychophysiology, University of Florida, Gainesville, Fla, USA, 1999.

[57] M. M. Bradley and P. J. Lang, "Affective norms for English text (ANET): affective ratings of text and instruction manual," Tech. Rep. D-1, University of Florida, Gainesville, Fla, USA, 2007.

[58] D. Grühn and J. Smith, "Characteristics for 200 words rated by young and older adults: age-dependent evaluations of German adjectives (AGE)," *Behavior Research Methods*, vol. 40, no. 4, pp. 1088–1097, 2008.

[59] J. Leite, S. Carvalho, S. Galdo-Alvarez, J. Alves, A. Sampaio, and Ó. F. Gonçalves, "Affective picture modulation: valence, arousal, attention allocation and motivational significance," *International Journal of Psychophysiology*, vol. 83, no. 3, pp. 375–381, 2012.

[60] A. von Leupoldt, K. Taube, M. Henkhus, B. Dahme, and H. Magnussen, "The impact of affective states on the perception of dyspnea in patients with chronic obstructive pulmonary disease," *Biological Psychology*, vol. 84, no. 1, pp. 129–134, 2010.

[61] A. Von Leupoldt, B. Hartle-Bremerich, and B. Dahme, "Application of video glasses for sustained affective picture presentations: a comparison with video projector presentations," *Behavior Research Methods*, vol. 37, no. 4, pp. 602–607, 2005.

[62] M. Csikszentmihalyi, *Beyond Boredom and Anxiety*, Jossey-Bass Behavioral Science Series, Jossey-Bass Publishers, San Francisco, Calif, USA, 1st edition, 1975.

[63] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, Harper & Row, New York, NY, USA, 1st edition, 1990.

[64] G. Riva, R. M. Baños, C. Botella, B. K. Wiederhold, and A. Gaggioli, "Positive technology: using interactive technologies to

promote positive functioning," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 2, pp. 69–77, 2012.

[65] B. K. Wiederhold and G. Riva, "Positive technology supports shift to preventive, integrative health," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 2, pp. 67–68, 2012.

[66] S. Serino, P. Cipresso, A. Gaggioli, and G. Riva, "The potential of pervasive sensors and computing for positive technology: the interreality paradigm," in *Pervasive and Mobile Sensing and Computing for Healthcare*, vol. 2 of *Smart Sensors, Measurement and Instrumentation*, pp. 207–232, 2013.

[67] D. Villani, A. Grassi, C. Cognetta, D. Toniolo, P. Cipresso, and G. Riva, "Self-help stress management training through mobile phones: an experience with oncology nurses," *Psychological Services*, vol. 10, no. 3, pp. 315–322, 2013.

[68] P. Cipresso, A. Gaggioli, S. Serino et al., "EEG alpha asymmetry in virtual environments for the assessment of stress-related disorders," *Studies in Health Technology and Informatics*, vol. 173, pp. 102–104, 2012.

[69] A. J. Camm, M. Malik, J. T. Bigger et al., "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[70] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, no. 8, pp. 1017–1034, 2012.

[71] D. C. Fowles, M. J. Christie, and R. Edelberg, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 18, no. 3, pp. 232–239, 1981.

[72] J. T. Larsen, C. J. Norris, and J. T. Cacioppo, "Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii," *Psychophysiology*, vol. 40, no. 5, pp. 776–785, 2003.

[73] I. J. T. Veldhuizen and A. W. K. Gaillard, "Can differences in fatigue be reflected by differences in tonic EMG corrugator supercilii muscle activity?" *Psychophysiology*, vol. 38, p. S97, 2001.

[74] C. W. Goodmurphy and W. K. Ovalle, "Morphological study of two human facial muscles: orbicularis oculi and corrugator supercilii," *Clinical Anatomy*, vol. 12, no. 1, pp. 1–11, 1999.

[75] T. D. Blumenthal, B. N. Cuthbert, D. L. Filion, S. Hackley, O. V. Lipp, and A. van Boxtel, "Committee report: guidelines for human startle eyeblink electromyographic studies," *Psychophysiology*, vol. 42, no. 1, pp. 1–15, 2005.

[76] U. Kunzmann, C. S. Kupperbusch, and R. W. Levenson, "Behavioral inhibition and amplification during emotional arousal: a comparison of two age groups," *Psychology and Aging*, vol. 20, no. 1, pp. 144–158, 2005.

[77] A. I. Bagić, R. C. Knowlton, D. F. Rose, and J. S. Ebersole, "American clinical magnetoencephalography society clinical practice guideline 3: MEG-EEG reporting," *Journal of Clinical Neurophysiology*, vol. 28, no. 4, pp. 362–363, 2011.

[78] A. I. Bagić, R. C. Knowlton, D. F. Rose, and J. S. Ebersole, "American clinical magnetoencephalography society clinical practice guideline 1: recording and analysis of spontaneous cerebral activity," *Journal of Clinical Neurophysiology*, vol. 28, no. 4, pp. 348–354, 2011.

[79] V. V. Nikulin and T. Brismar, "Long-range temporal correlations in alpha and beta oscillations: effect of arousal level and test-retest reliability," *Clinical Neurophysiology*, vol. 115, no. 8, pp. 1896–1908, 2004.

[80] S. Debener, A. Beauducel, K. Fiehler, S. Rabe, and B. Brocke, "Frontal EEG alpha asymmetry and affective style: are individual differences related to fundamental dimensions of emotion?" *Psychophysiology*, vol. 38, p. S35, 2001.

[81] E. Gordon, D. M. Palmer, and N. Cooper, "EEG alpha asymmetry in schizophrenia, depression, PTSD, panic disorder, ADHD and conduct disorder," *Clinical EEG and Neuroscience*, vol. 41, no. 4, pp. 178–183, 2010.

[82] D. Mathersul, L. M. Williams, P. J. Hopkinson, and A. H. Kemp, "Investigating models of affect: relationships among EEG alpha asymmetry, depression, and anxiety," *Emotion*, vol. 8, no. 4, pp. 560–572, 2008.

[83] R. de Raedt, E. Franck, K. Fannes, and E. Verstraeten, "Is the relationship between frontal EEG alpha asymmetry and depression mediated by implicit or explicit self-esteem?" *Biological Psychology*, vol. 77, no. 1, pp. 89–92, 2008.

[84] R. Nusslock, A. J. Shackman, B. W. McMenamin, L. L. Greischar, M. Kovacs, and R. J. Davidson, "Anxiety moderates relations between frontal EEG alpha asymmetry and depression," *Psychophysiology*, vol. 45, p. S77, 2008.

[85] R. Nusslock, A. J. Shackman, L. L. Greischar, B. W. McMenamin, M. Kovacs, and R. J. Davidson, "Frontal EEG alpha asymmetry in depression: the role of clinical state and emotion regulation," *Psychophysiology*, vol. 44, p. S7, 2007.

[86] M. Tops, A. A. Wijers, A. S. J. van Staveren et al., "Acute cortisol administration modulates EEG alpha asymmetry in volunteers: relevance to depression," *Biological Psychology*, vol. 69, no. 2, pp. 181–193, 2005.

[87] A. J. Niemiec and B. J. Lithgow, "Alpha-band characteristics in EEG spectrum indicate reliability of frontal brain asymmetry measures in diagnosis of depression," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS '05)*, pp. 7517–7520, Shanghai, China, September 2005.

[88] M. A. Diego, T. Field, and M. Hernandez-Reif, "CES-D depression scores are correlated with frontal EEG alpha asymmetry," *Depression and Anxiety*, vol. 13, no. 1, pp. 32–37, 2001.

[89] S. Debener, A. Beauducel, D. Nessler, B. Brocke, H. Heilemann, and J. Kayser, "Is resting anterior EEG alpha asymmetry a trait marker for depression? Findings for healthy adults and clinically depressed patients," *Neuropsychobiology*, vol. 41, no. 1, pp. 31–37, 2000.

[90] H. L. Urry, S. K. Hitt, and J. J. B. Allen, "Internal consistency and test-retest stability of resting EEG alpha asymmetry in major depression," *Psychophysiology*, vol. 36, p. S116, 1999.

[91] I. H. Gotlib, C. Ranganath, and J. P. Rosenfeld, "Frontal EEG alpha asymmetry, depression, and cognitive functioning," *Cognition and Emotion*, vol. 12, no. 3, pp. 449–478, 1998.

[92] A. D. Lykins, M. Meana, and G. Kambe, "Detection of differential viewing patterns to erotic and non-erotic stimuli using eye-tracking methodology," *Archives of Sexual Behavior*, vol. 35, no. 5, pp. 569–575, 2006.

[93] P. Cipresso, M. Mauri, A. Balgera, E. Romanò, and M. Villamira, "Synchronization of a biofeedback system with an eye tracker through an audiovisual stimulus marker," *Applied Psychophysiology and Biofeedback*, vol. 35, no. 4, pp. 332–332, 2010.

[94] P. Cipresso, "Synchronizing physiological signals acquired from biofeedback equipment and eye-tracker systems," *Applied Psychophysiology and Biofeedback*, vol. 36, no. 1, pp. 58–58, 2011.

[95] J. M. Oliveira, E. Volchan, C. D. Vargas, S. Gleiser, and I. A. David, "Box for interaction with objects (BIO): a new device to

synchronize the presentation of objects with electrophysiological recordings," *Behavior Research Methods*, vol. 44, no. 4, pp. 1115–1120, 2012.

[96] K. E. Emam, K. Moreau, and E. Jonker, "How strong are passwords used to protect personal health information in clinical trials?" *Journal of Medical Internet Research*, vol. 13, no. 1, article e18, 2011.

[97] P. Cipresso, A. Gaggioli, S. Serino, S. Cipresso, and G. Riva, "How to create memorizable and strong passwords," *Journal of Medical Internet Research*, vol. 14, no. 1, article e10, 2012.

[98] B. Brumen, M. Heričko, I. Rozman, and M. Hölbl, "Security Analysis and Improvements to the PsychoPass Method," *Journal of Medical Internet Research*, vol. 15, no. 8, article e161, 2013.

[99] K. Lee and D. F. Gilmore, "Statistical experimental design for bioprocess modeling and optimization analysis: repeated-measures method for dynamic biotechnology process," *Applied Biochemistry and Biotechnology*, vol. 135, no. 2, pp. 101–115, 2006.

[100] M. Bakker and J. M. Wicherts, "The (mis)reporting of statistical results in psychology journals," *Behavior Research Methods*, vol. 43, no. 3, pp. 666–678, 2011.

[101] R. O. Kuehl, *Design of Experiments: Statistical Principles of Research Design and Analysis*, Duxbury/Thomson Learning, Pacific Grove, Calif, USA, 2nd edition, 2000.

[102] G. W. Oehlert, *A First Course in Design and Analysis of Experiments*, W. H. Freeman, New York, NY, USA, 2000.

[103] E. R. Girden, *ANOVA: Repeated Measures*, Sage University Papers Quantitative Applications in the Social Sciences, Sage, Newbury Park, Calif, USA, 1992.

[104] F. Faul, E. Erdfelder, A. Lang, and A. Buchner, "G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior Research Methods*, vol. 39, no. 2, pp. 175–191, 2007.

[105] P. Sleight, M. T. La Rovere, A. Mortara et al., "Physiology and pathophysiology of heart rate and blood pressure variability in humans: is power spectral analysis largely an index of baroreflex gain? (Clinical Science (1995) 88 (103–109))," *Clinical Science*, vol. 88, no. 6, p. 733, 1995.

[106] R. W. Picard, *Affective Computing*, MIT Press, Cambridge, Mass, USA, 1997.

[107] S. Shiffman and A. A. Stone, "Ecological momentary assessment: a new tool for behavioral medicine research," *Technology and Methods in Behavioral Medicine*, pp. 117–131, 1998.

[108] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Annual Review of Clinical Psychology*, vol. 4, pp. 1–32, 2008.

[109] S. Dockray, N. Grant, A. A. Stone, D. Kahneman, J. Wardle, and A. Steptoe, "A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method," *Social Indicators Research*, vol. 99, no. 2, pp. 269–283, 2010.

Research Article

Music Emotion Detection Using Hierarchical Sparse Kernel Machines

Yu-Hao Chin, Chang-Hong Lin, Ernestasia Siahaan, and Jia-Ching Wang

Department of Computer Science and Information Engineering, National Central University, Taoyuan 32001, Taiwan

Correspondence should be addressed to Jia-Ching Wang; jcw@csie.ncu.edu.tw

Received 30 August 2013; Accepted 17 October 2013; Published 3 March 2014

Academic Editors: B.-W. Chen, S. Liou, and C.-H. Wu

For music emotion detection, this paper presents a music emotion verification system based on hierarchical sparse kernel machines. With the proposed system, we intend to verify if a music clip possesses happiness emotion or not. There are two levels in the hierarchical sparse kernel machines. In the first level, a set of acoustical features are extracted, and principle component analysis (PCA) is implemented to reduce the dimension. The acoustical features are utilized to generate the first-level decision vector, which is a vector with each element being a significant value of an emotion. The significant values of eight main emotional classes are utilized in this paper. To calculate the significant value of an emotion, we construct its 2-class SVM with calm emotion as the global (non-target) side of the SVM. The probability distributions of the adopted acoustical features are calculated and the probability product kernel is applied in the first-level SVMs to obtain first-level decision vector feature. In the second level of the hierarchical system, we merely construct a 2-class relevance vector machine (RVM) with happiness as the target side and other emotions as the background side of the RVM. The first-level decision vector is used as the feature with conventional radial basis function kernel. The happiness verification threshold is built on the probability value. In the experimental results, the detection error tradeoff (DET) curve shows that the proposed system has a good performance on verifying if a music clip reveals happiness emotion.

1. Introduction

Listening to music plays an important role in human's daily life and people usually gain much benefit from listening to music. Besides the leisure purpose, music listening has other application areas such as education, inspiration production, therapy, and marketing [1]. Sometimes people try to be in particular emotion state by listening to music. However, in such situation, people need to choose the music which can make human have particular feelings. They should listen to each song at least once to know the music emotion of each song, and the whole process takes much time. If people can use computer to detect the emotion content in music, the problem can be solved. Besides this application, music emotion detection technology can be applied to other area as well, such as music research, music recommendation, and music retrieval. For the limitless potential of music emotion detection technology, many researchers focus on detecting emotion in music.

Many researches on music emotion detection have been proposed in music emotion detection [2]. Existing research methods could be divided into two main categories: dimension approach and categorical approach. Dimension approach defines an emotion plane and views the emotion plane as a continuous emotion state space. Each position of the plane means an emotion state [3]. The acoustical features can be mapped to a point in the emotion plane [4]. Categorical approach works by categorized emotions into a number of emotion classes. Each emotion class represents an area in the emotion plane [3]. Different from dimension approach, each emotion class is defined clearly. In the training phase, acoustical features are directly used to train classifiers to recognize the corresponding emotion classes [5]. In this paper, the proposed method belongs to the second type.

In previous music emotion detection studies, many machine learning algorithms are applied. In [5], features were mapped into emotion categories on the emotion plane, and two support vector regressors were trained to predict the

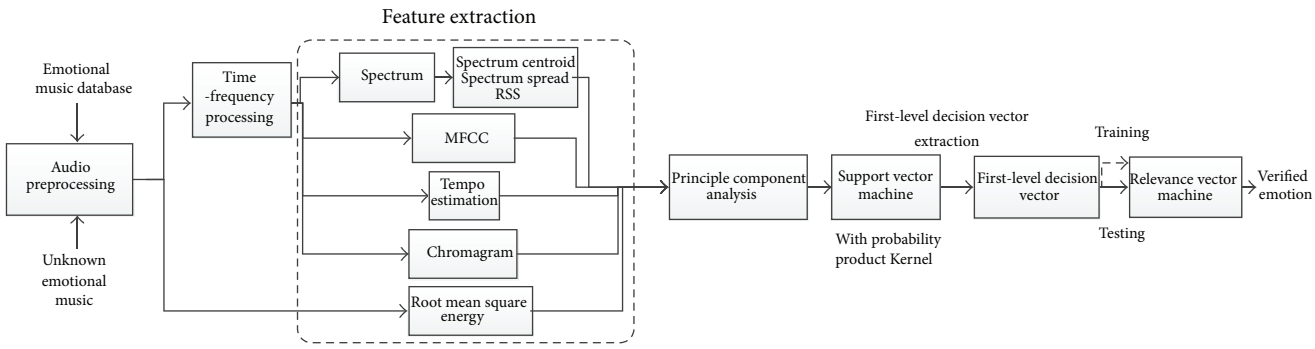


Figure 1: Block diagram of the proposed system.

arousal and valence value. In [6], hierarchical framework was adopted to detect emotion from acoustic music data. The method has the advantage of emphasizing proper feature in different detection work. In [7], support vector machine was applied to detect emotion content in music. In [8], kernel-based class separability is used to weight features. After feature selection, principal component analysis and linear discriminant analysis were applied, and k-nearest neighborhood (KNN) classifier was then implemented. In this paper, a music emotion detection system is proposed. The system establishes a hierarchical sparse kernel machine. In the first level, eight 2-class SVM models are trained, with eight emotion classes as the target sides, respectively. It is noted that emotion perception is usually not based on a single acoustical feature but a combination of acoustical features [4, 9]. This paper adopts an acoustical feature set comprising root mean square energy (RMS energy), tempo, chromagram, MFCCs, spectrum centroid, spectrum spread, and ratio of a spectral flatness measure to a spectral center (RSS). Each of them is normalized. In the second level of hierarchical sparse kernel machines, a 2-class relevance vector machine (RVM) model with happiness as the target side and other emotion as the background side is trained. Besides, first-level decision vector is used as the feature in this level.

The rest of this paper is organized as follows. The system overview is described in Section 2. The features and first-level decision vector extraction are described in Section 3. Principle component analysis is described in Section 4. The introduction of SVM and RVM is described in Section 5. Section 6 shows our experimental results. The conclusion is given in Section 7.

2. System Overview

The block diagram of the proposed system is presented in Figure 1. The system mainly comprises two level sparse kernel machines. For the first-level SVMs, we use a set of acoustical features which includes RMS energy, tempo, chromagram, MFCCs, spectrum centroid, spectrum spread, and RSS. In Table 1, the used acoustical features are classified into four main types, that is, dynamic, rhythm, timbre, and tonality.

Table 1: The proposed acoustical feature set.

Feature class	Feature name (dimension of feature)
Dynamic	RMS energy (1)
Rhythm	Tempo (1)
Timbre	MFCCs (13), spectrum centroid (1), spectrum spread (1), RSS (1)
Tonality	Chromagram (12)

Because each feature's scale is different, normalization of the whole feature set is performed [10]. After normalization, eight SVM models are trained to transform acoustical features into emotion profile features. Each of the eight SVM model is trained and tested using probability product kernel. We use the first-level decision vectors generated from the angry, happy, sad, relaxed, pleased, bored, nervous, and peaceful emotion classes. For an emotion, to calculate the corresponding value in the emotion profile features, we construct its 2-class SVM with calm emotion as the background side of the RVM. For the RVM, conventional radial basis function kernel is used, and the first-level decision vector extracted in the first level is utilized as the feature. To verify happiness emotion, a 2-class RVM with happiness as the target side and other emotion as the background side is constructed. For a tested music clip, the obtained probabilities value from this 2-class RVM is used to judge if this music clip belongs to happiness emotion or not.

3. Extraction of Acoustical Feature and First Level Decision Value Vector Feature

In the 2-level hierarchical sparse kernel machines, the first-level SVMs use acoustical features, while the second-level RVM adopts first-level decision vector. For acoustical features, the proposed system extracts RMS energy, tempo, chromagram, MFCCs, spectrum centroid, spectrum spread, and RSS. The extraction of these acoustical features as well as first-level decision vectors are described in the following.

3.1. Extraction of Acoustical Feature

3.1.1. RMS Energy. RMS energy is also called root mean square energy. It computes the global energy of input signal x [11]. The operation is defined as follows:

$$X_{\mathrm{RMS}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}, \tag{1}$$

where n means signal's length in hundredth of a second by default.

3.1.2. Tempo. Many tempo estimation methods have been proposed. The estimation of tempo is based on detecting periodicities in a range of BPMs [12]. Firstly, significant onset events are detected in the frequency domain [11]. Then find the events that best represents the tempo of the song, which means to choose the maximum periodicity score for each frame separately.

3.1.3. Chromagram. Chroma which is also called harmonic pitch class profile has a strong relationship with the structure of music [13]. Chromagram is a joint distribution of signal strength over the variables of time and chroma. Chroma is a frame-based representation of audio and is similar to short time Fourier transform. In music clips, frequency components belonging to the same pitch class are extracted by chromagram and transformed to a 12-dimensional representation, including C, C#, D, D#, E, F, F#, G, G#, A, A#, and B. The chromagram can present the distribution of energy along the pitches or pitch classes [11, 14]. In [14], chromagram is defined as the remapping of time-frequency distribution. The chromagram is extracted by

$$v(t,k) = \sum_{n\in S_k} \frac{X_t(n)}{Q_k} \quad k \in \{0,1,2,\ldots,11\}, \tag{2}$$

where $X_t(n)$ means the logarithmic magnitude of discrete Fourier transform of the tth frame, and Q_k is the number of elements in a subset of the discrete frequency space for each pitch class [15].

In Figure 2, the chromagram from a piece of music is exemplified.

3.1.4. Mel-Frequency Cepstral Coefficients (MFCCs). After signal is digitized, a large amount of information is not needed and cost plenty of storage space. Power spectrum is often adopted to encode the signal to solve the problem [16]. It is noted that MFCCs performs similar to human auditory perception system. The feature is adopted in various research topics, including speaker recognition, speech recognition, and music emotion recognition. For example, Cooper and Foote extracted MFCCs from music signal, and they found that MFCCs are similar to music timbre expression [17]. In [18], MFCCs were also proven to be having good performance in music recommendation.

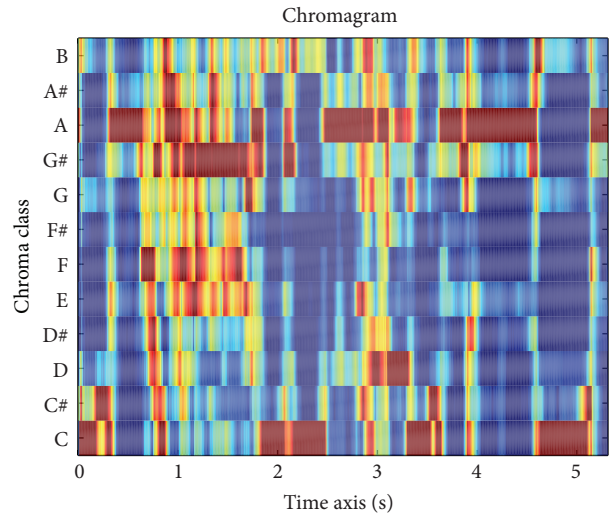


Figure 2: Example of chromagram from a piece of music.

MFCCs extraction is based on spectrum. The spectrum can be extracted by using discrete Fourier transform:

$$x_w(f) = \sum_{n=0}^{N} x_w(n) \exp\left\{-\frac{2\pi f n}{N}\right\}. \tag{3}$$

After power spectrum is extracted, subband energies can be extracted by using Mel filter banks and then evaluate logarithm value of the energies as follows:

$$S_i = \log \sum_{f=F_l}^{F_h} L(i,f)\left|X_w(f)\right|^2, \tag{4}$$

where F_h is the discrete frequency index corresponding to the high cutoff frequency, F_l is the discrete frequency index corresponding to low cutoff frequency, and $L(i,f)$ is the amplitude of the fth discrete frequency index of the ith Mel window. The number of the Mel windows often ranges from 20 to 24. Finally, MFCCs is obtained by performing discrete cosine transform (DCT) [19]. In Figure 3, the average MFCCs values from a piece of music are exemplified.

3.1.5. Spectrum Centroid. Spectrum centroid is an economical description of the shape of the power spectrum [20–22]. Additionally, it is correlated with a major perceptual dimension of timbre, that is, sharpness. Figure 4 gives an example of a spectrum and its spectrum centroid obtained from a frame in a piece of music. The spectrum centroid value is 2638 Hz in this example.

3.1.6. Spectrum Spread. Spectrum spread is an economical descriptor of the shape of the power spectrum that indicates whether it is concentrated in the vicinity of its centroid or else spread out over the spectrum [20–22]. It allows differentiating between tone-like and noise-like sounds. In Figure 5, an example of spectrum spread from a piece of music is provided.

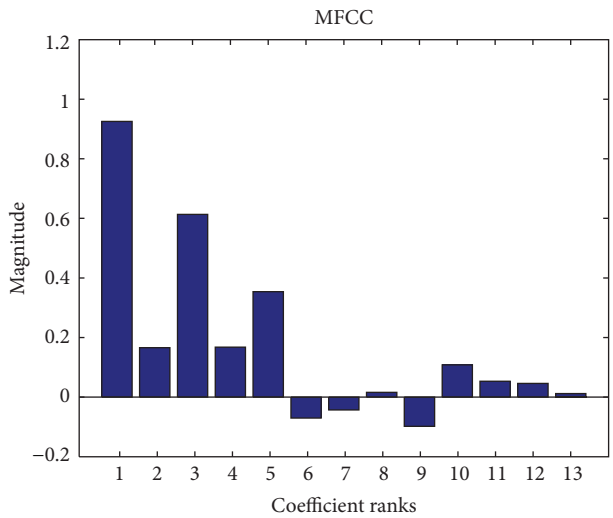


Figure 3: Example of average MFCCs values from a piece of music.

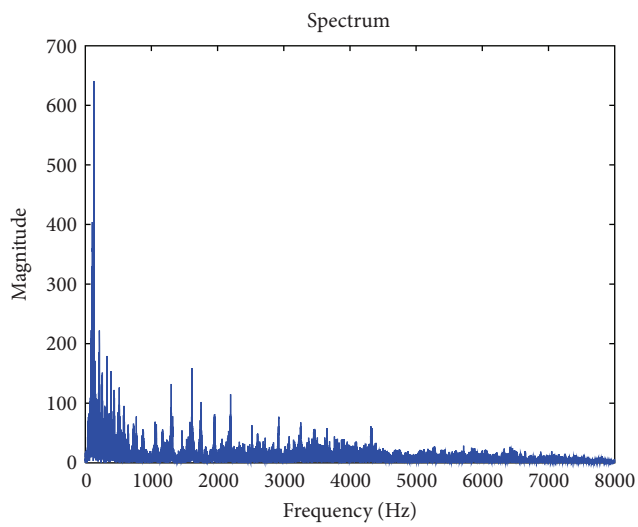


Figure 4: Example of a spectrum and its spectrum centroid from a frame in a piece of music.

3.1.7. Ratio of a Spectral Flatness Measure to a Spectral Center (RSS). RSS was proposed by Vapnik for speaker-independent emotional speech recognition [23]. RSS is the ratio of spectrum flatness to spectrum centroid and is calculated by

$$\mathrm{RSS} = \frac{1000 \times \mathrm{SF}}{\mathrm{SC}}, \tag{5}$$

where SF denotes spectrum flatness and SC represents spectrum centroid.

3.2. Extraction of First-Level Decision Vector. The acoustical feature set is utilized to generate the first-level decision vector with each element being a significant value of an emotion. This approach is able to interpret the emotional content by providing multiple probabilistic class labels, rather than a single hard label [24]. For example, happiness emotion not only contains happiness content, but also other properties that are similar to the content of peace. The similarity to peaceful may cause a music clip to be recognized as an incorrect emotion class. In this example, the advantage of first-level decision vector representation is its ability to convey both the evidences of happiness and peaceful emotions. This paper uses the significant values of eight emotions (angry, happy, sad, relaxed, pleased, bored, nervous, and peaceful) to construct an emotion profile feature vector. To calculate the significant value of an emotion, we construct its 2-class SVM with calm emotion as the background side of the SVM.

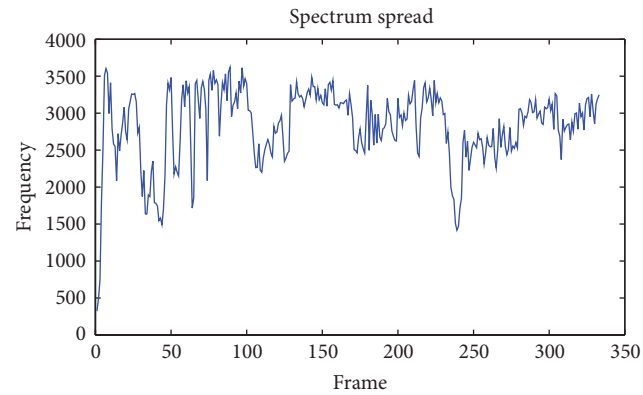


Figure 5: Example of spectrum spread from a piece of music.

4. Principle Component Analysis

PCA is an important mathematic technology in feature extraction approach. In this paper, PCA is implemented to reduce the dimensions of the extracted features. The first step of PCA is to calculate the d-dimension mean vector $\mathbf{u}$ and $d \times d$ covariance matrix $\mathbf{\Sigma}$ of the samples [25]. After that, the eigenvectors and eigenvalues are computed. Finally, the largest k eigenvectors are selected to form a $d \times k$ matrix M whose columns consist of the k eigenvectors. In fact, the other dimensions are noise. The PCA transformed data can be in the form

$$\mathbf{x}' = \mathbf{M}^T (\mathbf{x} - \mathbf{u}). \tag{6}$$

5. Emotion Classifier

The emotion classifier used in the proposed system adopts a 2-level hierarchical structure of sparse kernel machines. The first-level SVMs use probability product kernel, while the second-level RVM adopts traditional radial basis function kernel with first-level decision vector feature.

5.1. Support Vector Machine. The SVM theory is an effective statistical technique and has drawn much attention on audio classification tasks [7]. An SVM is a binary classifier that creates an optimal hyperplane to classify input samples. This optimal hyperplane linearly divides the two classes with the largest margin [23]. Denote $T = \{(\mathbf{x}_i, y_i), i = 1, 2, \ldots, N\}$ as a training set for SVM; each pair $(\mathbf{x}_i, y_i)$ means training

sample $\mathbf{x}_i$ belongs to a class y_i, where $y_i \in \{+1, -1\}$. The fundamental concept is to choose a hyperplane which can classify $\mathbf{T}$ accurately while maximizing the distance between the two classes. This means to find a pair $(\mathbf{w}, b)$ such that

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 0, \quad i = 1, \ldots, N, \tag{7}$$

where $\mathbf{w} \in R^N$ is normalized by itself and $b \in R$.

The pair $(\mathbf{w}, b)$ defines a separating hyperplane of equation

$$\mathbf{w} \cdot \mathbf{x} + b = 0. \tag{8}$$

If there exists a hyperplane satisfying (7), the set T is said to be linearly separable and we can change $\mathbf{w}$ and b so that

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 1, \quad i = 1, \ldots, N. \tag{9}$$

According to (9), we can derive an objective function under constraint

$$\begin{aligned} &\min \quad \|\mathbf{w}\|^2 \\ &\text{subject to} \quad y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 1, \quad i = 1, \ldots, N. \end{aligned} \tag{10}$$

Since $\|\mathbf{w}\|^2$ is convex, we can solve (9) by applying the classical method of Lagrange multipliers:

$$\min \|\mathbf{w}\|^2 + \mu_i [y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1], \quad i = 1, \ldots, N. \tag{11}$$

We denote $\mathbf{U} = (\mu_1, \mu_2, \ldots, \mu_N)$ as the N nonnegative Lagrange multipliers associated with (10). After solving (11), the optimal hyperplane has the following expansion:

$$\overline{\mathbf{w}} = \sum_{i=1}^{N} \mu_i y_i x_i. \tag{12}$$

$\overline{b}$ can be determined from $\mathbf{U}$ and from the Kühn-Tucker conditions. Consider

$$\mu_i\left(y_i\left(\overline{\mathbf{w}} \cdot x_i + \overline{b}\right) - 1\right) = 0, \quad i = 1, 2, \ldots, N. \tag{13}$$

Accordingly (11), the expected hyperplane is a linear combination of training samples. The corresponding training samples $(\mathbf{x}_i, y_i)$ with nonzero Lagrange multipliers are called support vectors. Finally, the decision value from a new data point $\mathbf{x}$ can be written as

$$\operatorname{dec}(x) = \sum_{i=1}^{N} \mu_i y_i \mathbf{x}_i \cdot \mathbf{x} + \overline{b}. \tag{14}$$

Functions that satisfy Mercer's theorem can be used as kernels. In this paper, probability product kernel is adopted.

5.2. Probability Product Support Vector Machine. A function can be considered as kernel function if the function satisfies Mercer's theorem. Using Mercer's theory, we can introduce a mapping function $\phi(\mathbf{x})$, such that $k(\mathbf{x}_j, \mathbf{x}_i) = \phi(\mathbf{x}_j)\phi(\mathbf{x}_i)$. This provides the ability of handling nonlinear data, by mapping the original input space $\mathbf{R}^d$ into some other space.

In this paper, the probability product kernel is utilized. The probability product kernel is a method of measuring similarity between distributions, and it has the property of simple and intuitively compelling conception [26]. Probability product kernel computes a generalized inner product between two probability distributions in the Hilbert space. A positive definite kernel $k : O \times O \rightarrow \mathbb{R}$ on input space O and examples $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_m \in O$ are defined. Firstly, the input data x is mapped to a probability distribution $p(x \mid O)$, which fits separate probabilistic models $p_1(x), p_2(x), \ldots, p_m(x)$ to $\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_m$. After that, a novel kernel $k^{\text{prob}}(p_i, p_j)$ between probability distributions on O is defined. At last, a kernel between examples is needed to be defined, and the kernel is equal to k^{prob} between the corresponding distributions. Consider

$$k(\mathbf{o}_i, \mathbf{o}_j) = k^{\text{prob}}(p_i, p_j). \tag{15}$$

Finally, this kernel is applied to SVM and proceeded as usual. The probability product kernel between distributions p_i and p_j is defined as

$$k(p_i, p_j) = \int_x p_i^\rho(\mathbf{x})\, p_j^\rho(\mathbf{x})\, dx = \left\langle p_i^\rho, p_j^\rho \right\rangle_{L_2}, \tag{16}$$

where p_i and p_j are probability distributions on a space O. Assume that $p_i^\rho, p_j^\rho \in L_2(O)$. L_2 is a Hilbert space and ρ is a positive constant. Probability product kernel allows us to introduce prior knowledge of data. In this paper, we assume a d-dimensional Gaussian distribution of our data.

5.3. First-Level Decision Vector Extraction. First-level decision vector presents perception probability of each of the eight emotion-specific decisions, which is extracted from input data by collecting decision values from each model. The decision value of SVM represents the degree of similarity between model and testing data. The advantage of similarity measure can be used to find out which model fits the data most accurately [24]. Using the first-level decision vector, the most probably perceived emotion in music can be detected.

5.4. Relevance Vector Machine. RVM is a development of SVM. Different from SVM, RVM tries to find a considerable number of weights which has highest sparsity [27]. The model defines a conditional distribution for target class $y = \{0, 1\}$, given an input set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ [28]. Assume that a training data can be a linear combination of weighted nonlinear basis functions $\phi_i(\mathbf{x})$, which is transformed by a logistic sigmoid function as follows:

$$f(\mathbf{x}; \mathbf{w}) = \mathbf{w}^T \phi(\mathbf{x}), \tag{17}$$

where $\mathbf{w} = (w_1, w_2, \ldots, w_G)$, $\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \ldots, \phi_G(\mathbf{x}))^T$ denotes the weights. In order to make weight sparse, the Bayesian probabilistic framework is implemented to find the distribution over the weights instead of using pointwise estimation; therefore, a separate hyperparameter a for each

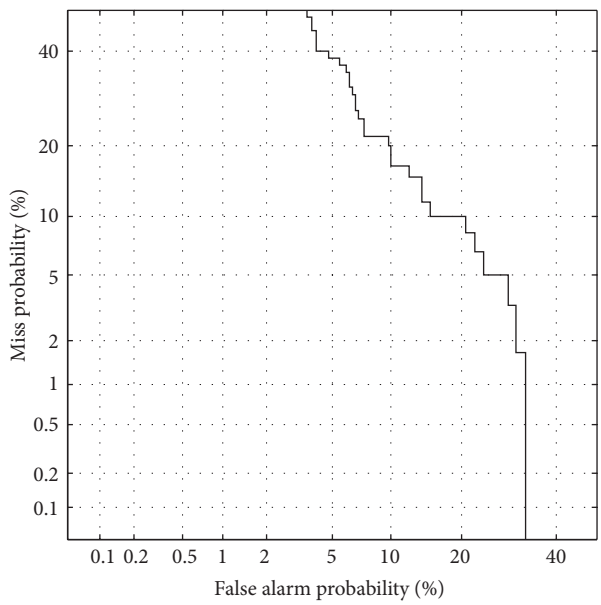


Figure 6: DET curve of the proposed system.

of the weight parameters w is introduced. According to Bayes rule, the posterior probability of w is

$$p(\mathbf{w} \mid y, \mathbf{a}) = \frac{p(y \mid \mathbf{w}, \mathbf{a})\, p(\mathbf{w} \mid \mathbf{a})}{p(y \mid \mathbf{a})}, \tag{18}$$

where $p(y \mid \mathbf{w}, \mathbf{a})$ is likelihood, $p(\mathbf{w} \mid \mathbf{a})$ is prior conditioned on weights $\mathbf{a} = [a_1, \ldots, a_n]^T$, and $p(y \mid \mathbf{a})$ denotes the evidence. For the reason that y is a binary variable, the likelihood function can be given by

$$p(y \mid \mathbf{w}, \mathbf{a}) = \prod_{i=1}^{n} [\sigma(f(\mathbf{x}_i; \mathbf{w}))]^{y_i} [1 - \sigma(f(\mathbf{x}_i; \mathbf{w}))]^{1-y_i}, \tag{19}$$

where $\sigma(f) = 1/(1+e^{-f})$ is the logistic sigmoid link function. According to (18), it can be found that a significant proportion of hyperparameters tend to be infinity, and the corresponding posterior distributions of weight parameters are concentrated at zero. Therefore, the basis functions that multiplied by these parameters will not be taken for reference when training the model. As a result, the model will be sparse.

6. Experimental Results

In the experiments, we collected one hundred songs from two websites to construct a music emotion database. These websites are All Music Guide [29] and Last.fm [30]. As mentioned before, music may contain multiple emotions. If we know which emotion class a song most likely belongs to, we may know the main emotion of the song. Songs in Last.fm are tagged by many people on the Internet. We choose the emotion which is tagged by most people to be the ground truth of data.

The database consists of nine classes of emotions, including happy, angry, sad, bored, nervous, relaxed, pleased, calm, and peaceful. Calm is taken as a model's opposite site when training models. Each emotion class contains twenty songs. Each song is thirty seconds long and is divided into five-second clips. Half of the songs are used as training data, and the others are used as testing data. In this paper, 240 music clips are tested. All of songs are western music and are encoded in 16 KHz WAV format. The used acoustical feature set are listed in Table 1. The whole feature set dimension is 30. The used SVM is based on LIBSVM library [31], and the used RVM is based on PTR toolbox [32]. The system performance is evaluated in terms of DET curve. Figure 6 depicts DET curve of the proposed happiness verification system. The proposed system can achieve 13.33% equal error rate (EER). From our results, we see that the system performs well on happiness emotion verification in music.

7. Conclusion

Detecting emotion in music has become the concern of many researchers in recent years. In this paper, we proposed a first-level decision-vector-based music happiness emotion detection system. The proposed system adopts a hierarchical structure of sparse kernel machines. First, eight SVM models are trained based on acoustical features with probability product kernel. Then eight decision values can be extracted to construct the first-level decision vector feature. After that, these eight decision values are considered as new feature to train and test a 2-class RVM with happiness as the target side. The probability value of the RVM is used to verify happiness content in music. Experimental results reveal that the proposed system can achieve 13.33% equal error rate (EER).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] R. E. Milliman, "Using background music to affect the behavior of supermarket shoppers," *Journal of Marketing*, vol. 46, no. 3, pp. 86–91, 1982.

[2] C.-H. Yeh, H.-H. Lin, and H.-T. Chang, "An efficient emotion detection scheme for popular music," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '09)*, pp. 1799–1802, Taipei City, Taiwan, May 2009.

[3] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen, "A regression approach to music emotion recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 2, pp. 448–457, 2008.

[4] Y. H. Yang and H. H. Chen, "Prediction of the distribution of perceived music emotions using discrete samples," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2184–2196, 2011.

[5] B. Han, S. Rho, R. B. Dannenberg, and E. Hwang, "SMERS: music emotion recognition using support vector regression," in *Proceedings of the International Conference on Music Information Retrieval*, Kobe, Japan, 2009.

[6] L. Lu, D. Liu, and H.-J. Zhang, "Automatic mood detection and tracking of music audio signals," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 5–18, 2006.

[7] C.-Y. Chang, C.-Y. Lo, C.-J. Wang, and P.-C. Chung, "A music recommendation system with consideration of personal emotion," in *Proceedings of the International Computer Symposium (ICS '10)*, pp. 18–23, Tainan City, Taiwan, December 2010.

[8] F. C. Hwang, J. S. Wang, P. C. Chung, and C. F. Yang, "Detecting emotional expression of music with feature selection approach," in *Proceedings of the International Conference on Orange Technologies (ICOT '13)*, pp. 282–286, March 2013.

[9] K. Hevner, "Expression in music: a discussion of experimental studies and theories," *Psychological Review*, vol. 42, no. 2, pp. 186–204, 1935.

[10] M. Chouchane, S. Paris, F. Le Gland, C. Musso, and D.-T. Pham, "On the probability distribution of a moving target. Asymptotic and non-asymptotic results," in *Proceedings of the 14th International Conference on Information Fusion (Fusion '11)*, pp. 1–8, July 2011.

[11] O. Lartillot and P. Toiviainen, "MIR in Matlab (II): a toolbox for musical feature extraction from audio," in *Proceedings of the International Conference Music Information Retrieval*, pp. 127–130, 2007, https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox.

[12] C.-W. Chen, K. Lee, and H.-H. Wu, "Towards a class-based representation of perceptual tempo for music retrieval," in *Proceedings of the 8th International Conference on Machine Learning and Applications (ICMLA '09)*, pp. 602–607, December 2009.

[13] W. Chai, "Semantic segmentation and summarization of music," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 124–132, 2006.

[14] M. A. Bartsch and G. H. Wakefield, "Audio thumbnailing of popular music using chroma-based representations," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 96–104, 2005.

[15] X. Yu, J. Zhang, J. Liu, W. Wan, and W. Yang, "An audio retrieval method based on chromagram and distance metrics," in *Proceedings of the International Conference on Audio, Language and Image Processing (ICALIP '10)*, pp. 425–428, Shanghai, China, November 2010.

[16] J. O. García and C. A. R. Garcia, "Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 3140–3145, July 2003.

[17] C. Y. Lin and S. Cheng, "Multi-theme analysis of music emotion similarity for jukebox application," in *Proceedings of the International Conference on Audio, Language and Image Processing (ICALIP '12)*, pp. 241–246, July 2012.

[18] B. Shao, M. Ogihara, D. Wang, and T. Li, "Music recommendation based on acoustic features and user access patterns," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, pp. 1602–1611, 2009.

[19] W.-Q. Zhang, D. Yang, J. Liu, and X. Bao, "Perturbation analysis of mel-frequency cepstrum coefficients," in *Proceedings of the International Conference on Audio, Language and Image Processing (ICALIP '10)*, pp. 715–718, Shanghai, China, November 2010.

[20] H. G. Kim, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, Wiley, New York, NY, USA, 2005.

[21] ISO-IEC/JTCI SC29 WGII Moving Pictures Expeti Group, "Information technology—multimedia content description interface—part 4: Audio," Comittee Draft 15938-4, ISO/IEC, 2000.

[22] M. Casey, "MPEG-7 sound-recognition tools," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 737–747, 2001.

[23] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.

[24] E. Mower, M. J. Mataric, and S. Narayanan, "A framework for automatic human emotion classification using emotion profiles," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1057–1070, 2011.

[25] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, New York, NY, USA, 2nd edition, 2001.

[26] T. Jebara, R. Kondor, and A. Howard, "Probability product kernels," *Journal of Machine Learning Research*, vol. 5, pp. 819–844, 2004.

[27] F. A. Mianji and Y. Zhang, "Robust hyperspectral classification using relevance vector machine," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2100–2112, 2011.

[28] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, New York, NY, USA, 2nd edition, 2007.

[29] "The All Music Guide," http://www.allmusic.com.

[30] "Last.fm," http://cn.last.fm/home.

[31] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[32] "Pattern Recognition Toolbox," http://www.newfolderconsulting.com/prt.

Research Article

Improving Learning Performance with Happiness by Interactive Scenarios

Chi-Hung Chuang,[1] Ying-Nong Chen,[2] Luo-Wei Tsai,[3] Chun-Chieh Lee,[4] and Hsin-Chun Tsai[5]

[1] *Department of Applied Informatics, Fo Guang University, Ilan, Taiwan*
[2] *Department of Information Management, National United University, No. 1, Lienda, Miaoli 36003, Taiwan*
[3] *Department of Information and Communications Research Lab, Industrial Technology Research Institute, Hsinchu, Taiwan*
[4] *Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan*
[5] *Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan*

Correspondence should be addressed to Ying-Nong Chen; ynchen@nuu.edu.tw

Received 30 August 2013; Accepted 17 October 2013; Published 16 January 2014

Academic Editors: B.-W. Chen, S. Hsieh, S. Liou, J.-F. Wang, and C.-H. Wu

Recently, digital learning has attracted a lot of researchers to improve the problems of learning carelessness, low learning ability, lack of concentration, and difficulties in comprehending the logic of math. In this study, a digital learning system based on Kinect somatosensory system is proposed to make children and teenagers happily learn in the course of the games and improve the learning performance. We propose two interactive geometry and puzzle games. The proposed somatosensory games can make learners feel curious and raise their motivation to find solutions for boring problems via abundant physical expressions and interactive operations. The players are asked to select particular operation by gestures and physical expressions within a certain time. By doing so, the learners can feel the fun of game playing and train their logic ability before they are aware. Experimental results demonstrate that the proposed somatosensory system can effectively improve the students' learning performance.

1. Introduction

Nowadays, the education background becomes increasingly important. Therefore, more and more parents have paid attention to early childhood education because they want their children to win at the starting line. However, some problems usually occur in the ordinary course of the children's educational process, such as learning carelessness, low learning ability, lack of concentration, and difficulties in comprehending the logic of math. Some children have even suffered from learning disorders, which their parents are unaware of. The phenomenon of children with learning disabilities, also known as learning difficulties, refers to children's intelligence in the normal range, but they have difficulty in learning. The common examples are mathematics disorder, dyslexia, writing disorders, attention-deficit/hyperactivity disorder, and so forth, which are all learning disabilities. The crucial point of guiding children to successful learning is to arise their interest. According to the market research of NPD, among 2 to 17 years old children and teenager population in America, 82% are video game players, amount to about 557 million people (2009), which demonstrates that video games are very popular leisure activities in the children and teenagers. Therefore, embedding the learning process in the video games playing would be a solution for happy learning. Hogle [1] proposed some advantages of learning by video games playing.

(1) Trigger intrinsic motivation and increase interest: the nature of curiosity, expectation, control, interaction, and fantasy storyline in the video games could improve the learners' interest in learning and intrinsic motivation, and the learners would be able to keep trying in the face of difficult challenges for obtaining sense of achievements.

(2) Memory reserving: compared with the traditional learning methods, learning by video games playing could achieve higher effect of memory reserving.

(3) Practice and feedback: many learning by video games playing software provide learners chances of repeating practice and immediate feedbacks of errors, which make learners assess their learning performance and improve the achievement of learning objectives.

(4) Proving high-level thinking: learning by video games playing is the best way of learning, since the design of video games meets with the human cognitive structure, which makes learners find solutions and obtain knowledge via repeatedly solving problems, making decisions, and integrating what they learn. Then, the teaching content could be constantly implanted into learners' memories.

Somatosensory system is an arising interactive video game and multimedia technology. The users can receive abundant feedbacks in the operation since its direct responding. Therefore, it would be a good choice to use experience system as a learning platform so that the students can learn happily and enhance their attention and interest in learning. Motivated from the above, the Microsoft Kinect is used to develop a learning somatosensory system in this study. The Kinect can extract color image, 3D depth image, and voice. In the proposed system, the 3D depth image is used to detect users' actions. The Kinect uses three steps for object detection and tracking. First, it uses Light Coding method to extract 3D depth image [2]. Then, the color image and depth image are combined to find out the human skeleton and joints [3]. Finally, the regression method is applied to improve the consistency between human posture and skeleton [4]. In addition, it can detect at most six people and recognize the actions of the two simultaneously. Twenty joints of skeleton are extracted for each detected and tracked people, including their body, limbs, and fingers for interactive somatosensory operation. Based on the extracted and tracked skeletons, pose estimation, [5–8], action recognition [9–12], image segmentation [13–15], body pose reconstruction [16–19], and building rich 3D maps of environments [20, 21] could be achieved. Although the Kinect defines and extracts many human joints, however, the details of the palm are insufficient. Therefore, the gestures cannot be detected and recognized in the original Kinect system. To solve this problem, in our proposed system, the palm joint is set as region of interest (ROI) via the Kinect skeleton tracking system and then extracted for image processing and gesture recognition via OPENNI environment.

In this study, we propose a learning somatosensory system based on Kinect to make children and teenagers happily learn in the course of the games and improve the learning performance. We propose two interactive geometry and puzzle games. The proposed geometry game can make learners feel curious and raise their motivation to find solutions for boring geometry problems via abundant physical expressions and interactive operations. The players are asked to select particular operation by gestures and physical expressions within a certain time. By doing so, the learners can feel the fun of game playing and train their logic ability happily. The proposed puzzles game can train the learners' concentration ability and logical thinking via abundant physical expressions. For example, when the learners are playing the puzzle games, they would practice their cognition ability to identify and group the shape and color of puzzles. In addition, the learners would practice their physical skills moving puzzles to correct positions via physical expressions and gestures.

The rest of this paper is organized as follows. In Section 2, the basis of Kinect system and digital learning will be briefly reviewed. Then, the proposed learning based on Kinect is presented in Section 3. In Section 4, experimental results are illustrated to demonstrate the soundness and effectiveness of the proposed digital learning method. Finally, conclusions are given in Section 5.

2. Related Works

2.1. The Application of Kinect. As mentioned above, the Kinect uses Light Coding method for object detection. This method is based on the Laser Speckle theory, which is a random reflecting speckle pattern produced when the laser light projects on some object. Since the speckle pattern would never be the same in any position, the monitored space is all marked to detect object's position. In practice, the Kinect uses infrared laser light projector and sensor to analyze the shift of laser speckle pattern by projecting from one position and observing from another to construct the depth map as Figure 1.

After object detection and 3D depth map construction, the skeleton tracking system on the Kinect is applied to extract twenty joints of human body and limbs. Since the joints of human body and limbs are extracted, the relationships of the joints are used as features for action recognition in Microsoft somatosensory games. Based on the skeleton tracking system, the human can be tracked in real time as Figure 2 and other applications can also be achieved.

2.2. The Digital Learning. Recently, to enhance learning through games has attracted a lot of researchers. Sarmanho et al. [22] proposed a Kinect-based game to help students of dyslexia and dysgraphia. Kenneth et al. [23] integrated the drawing games and somatosensory system, so the students could paint via moving their limbs. Since the somatosensory system is a new user interface making the interactions between computers and users are improved. Therefore, Lien et al. [24] designed an L-shape platform for learning based on interaction with computers. As Figure 3 shows, students can easily and happily learn by moving their limbs and interacting with computers. Casas et al. [25] integrated augmented reality technology with Kinect to create an environment in which users can interact with virtual objects with limbs (as Figure 4 illustrates). In addition, in order to improve the willingness of students in learning, Smorkalov et al. [26] integrated virtual reality technology with Kinect to immerse users in the virtual world and operate the virtual characters with limbs. Tuveri et al. [27] proposed a method using gestures and positions of palms to control planetarium software, improving the effectiveness of learning.

Li [28] proposed a Protractor scheme which was a template-based and single-stroke gesture identifier that used

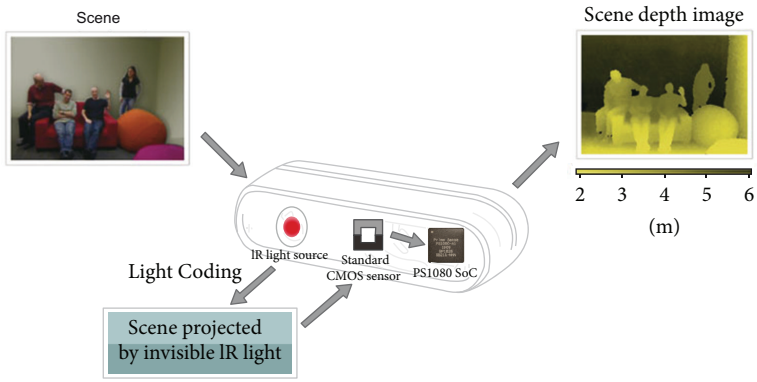


Figure 1: The procedure of object detection of Light Coding method (PrimeSense).

Figure 2: The skeleton of human body and tracking (PrimeSense).

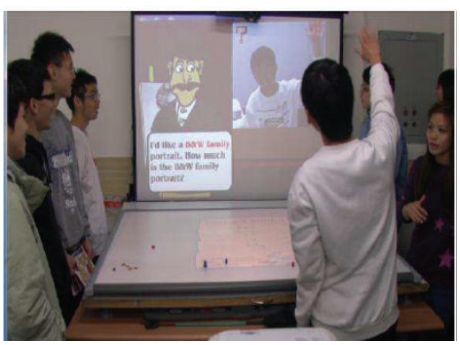


Figure 3: L-shape plate form [24].

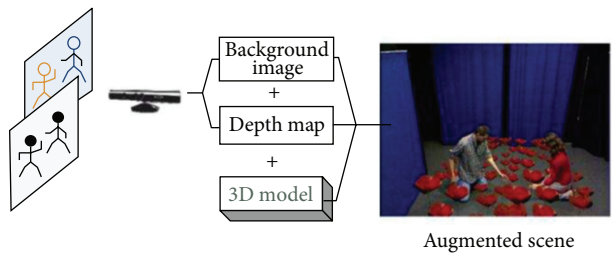


Figure 4: Integration of augmented reality and Kinect [25].

a new closed-form method to compute the similarity between gestures. Kratz and Rohs [29] designed a $3 gesture recognition system using 3D acceleration sensors. The scheme was proposed to be implemented immediately in prototyping environments and does not need any special equipment or environments. Only simple trigonometric and geometric calculations were necessary. Kratz and Rohs [30] also proposed a lightweight classifier for motion-based 3D gestures solving the difficulty of searching the optimal rotation between an input gesture and a template gesture. Hong et al. [31] designed a state based method for gesture learning and recognition. It uses spatial clustering and temporal alignment, and gestures are defined as sequence of states in spatial-temporal space. Wobbrock et al. [32] proposed a $1 recognizer system for gestures recognition. The system was easy to be implemented. Although the scheme was simple, it can provide rotation, scale, and position invariance. It also needed no sophisticated mathematical operations but competed with methods using dynamic programming and statistical recognition.

3. The Proposed Somatosensory Learning System

3.1. The Details of Kinect System. The skeleton detection technique is the core module of the somatosensory system. As shown in Figure 5, skeleton of the object was detected by the technique called Light Coding. Pose recognition is achieved by combining the 3D depth map and color information of the object. The dynamic posture correction is also performed till the object is out of the camera's range or program termination.

In the program flowchart, the object detection system is opened firstly. When the object appears on the camera's shooting range, "New User" function is called and the pose detection system starts. "New User" and the "Lost User" functions (events callback function) are called when "new object be detected" and "objects leave the detection range for a while." "New User" program calls pose detection by "Start Pose Detection" and its own "Pose Detected." "Pose Detected" check whether the number of object limit is out of bound. Once the limits is reached, it calls "Stop Pose Detection" to stop the detection module. "Pose Detected" also calls the skeleton processing unit named "Request Calibration" for the calibration and analysis of human skeleton. When "Request Calibration" is called, the object's skeleton will be analyzed.

Skeleton processing unit calls "Calibration Start" to start to skeleton correction. When the skeleton calibration is

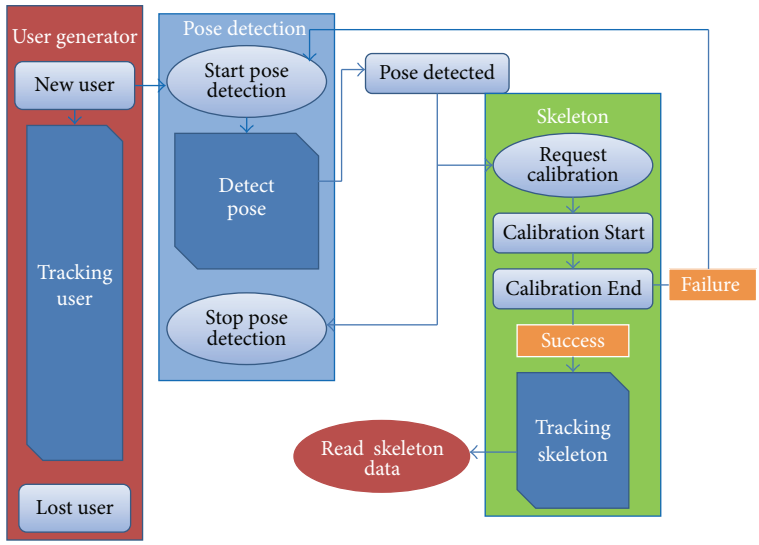


Figure 5: Skeleton detection flowcharts of the Kinect.

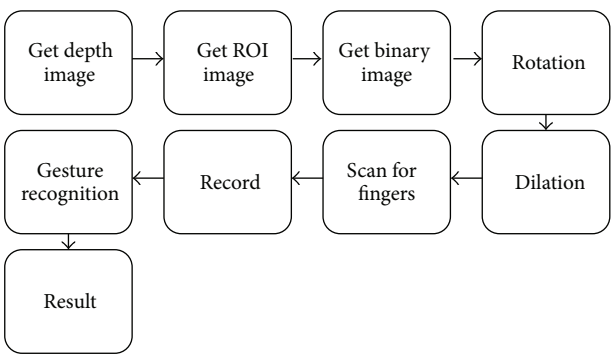


Figure 6: Finger gesture recognition system.

completed, "Calibration End" is called. However, "Calibration End" does not represent a successful identification of the object's skeleton. If it is successful, the next stage is "Start Tracking," allowing the system to start tracking the skeleton calibration data. If it fails, back to pose detection unit and redetect user gestures. When skeleton calibration and tracking skeleton are successful, users call the joint information function to get the object's joints data. The whole human skeleton is established.

3.2. The Gesture Tracking Systems and Applications. The system expands OPENNI development kit for finger gesture recognition. Users can play puzzle games produced by the system with gestures intuitively. Another application about the gesture recognition system is the Microsoft PowerPoint presentation software. The speaker can switch slides with gestures. Since the Kinect can track the palm's position, we use this information to locate the fingertips' position. With the fingertips' position, finger gesture can be extracted through the number of the fingertips and the angle between the fingertips. Detail system flowcharts can be found in Figure 6.

Each module of the system shown in Figure 6 will be described as follows.

(1) *Get ROI (Region of Interest) Image.* The rough area of the palm is extracted using OPENNI for further analysis and identification. The size of the palm is defined as follows:

$$\text{PlamSize} = \frac{1}{(\text{handDepth}) \times k_1}, \tag{1}$$

where the "handDepth" is the palm position depth and k_1 is a constant and is set as 20000 in the experiment.

(2) *Get Binary Image.* Based on the depth values of each pixel in palm image, the real palm region can be defined as follows:

$$I_{\text{palm}} = \begin{cases} 1, & \text{the depth distance of the pixels } i \le d_{\text{plam}}, \\ 0, & \text{the depth distance of the pixels } i > d_{\text{plam}}. \end{cases} \tag{2}$$

d_{plam} is the depth value of the central pixels of the palm. The upper image of the Figure 7 shows the palm detection result. The bottom image of Figure 7 is the binary image of real palm region.

(3) *Rotation.* For image normalization, each pixel (x, y) in the binary image of palm is rotated according to the following:

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta, \\ y' &= -x\cos\theta + y\sin\theta. \end{aligned} \tag{3}$$

However, the center axis of the palm needs to be shifted to fit the 2D coordinates using the following:

$$\begin{aligned} x' - h &= (x-h)\cos\theta + (y-k)\sin\theta, \\ y' - k &= -(x-h)\sin\theta + (y-k)\cos\theta. \end{aligned} \tag{4}$$

Figure 7: Binary image of the palm.

Figure 9: The image at the bottom shows the palm area rotated.

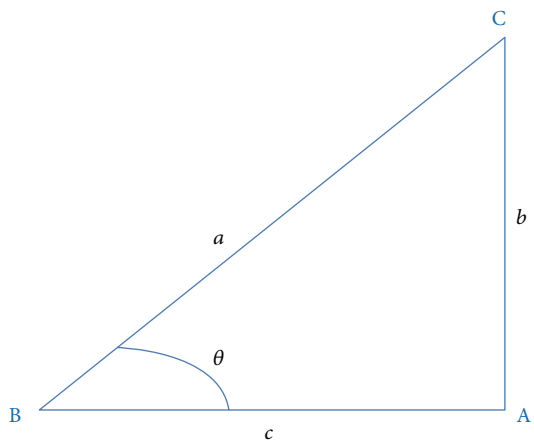


Figure 8: The illustration of the symbols used in (5).

Given the positions of the palm and wrist, the angle θ between the palm orientation and the horizontal direction can be calculated by the following:

$$\theta = \cos^{-1}\left(\frac{a^2 + c^2 - b^2}{2ac}\right). \quad (5)$$

See Figure 8 for an explanation of (5), where the points B and C represent the positions of wrist and palm, respectively.

(4) *Denoising by Dilation*. To eliminate the noises or the so-called white spots, the morphological dilation operations are employed on the rotated binary image. Note that those noises are generated by the noninteger results calculated by applying (5). There are some white spots on the rotated binary image; see Figure 9 for an illustration. Dilation is a popular morphological operation in the image processing domain and usually used to fill up the small holes spread in a binary image. Simply speaking, the pixel with value 0 (white) in the input image is set to value 1 (black) in the output image if any of its "connected neighbors" are with value 1. Usually, the 8 connected neighbors or 4 connected neighbors are used. In our experiments, the selection of 4 connected ones shows good performance.

(5) *Finding the Fingertips*. After obtaining the rotated and denoised palm area, we apply Algorithm A below to find the possible positions of all fingertips.

Algorithm A

Step 1. Scan the binary image from top to bottom to get the first pixel with value 1, and put the pixel in Queue (enqueue operation).

Step 2. Retrieve the first pixel from Queue (dequeue operation) and check the values of its 5 neighbors positioned at its right, bottom-right, bottom, bottom-left, and left. Put its neighbor pixels in Queue if the pixels are with value 1 and then set all these pixels to value 0.

Step 3. Repeat Step 2 until the Queue is empty.

Step 4. Repeat Step 1 until the binary image is scanned to its bottom.

To understand the algorithm above more clear, let us see the example shown in Figure 10. The bottom image in Figure 10 shows the results after finishing the first round of Step 3 and in this case the middle finger is detected. The grey color denotes the area eliminated in Step 2. The bottom image in Figure 11 shows the results after finishing the second round of Step 3 and in this case the ring finger is detected.

Note that the time complexity of Algorithm A is $O(n^2)$, where n is the number of pixels.

(6) *Recording the Positions of Real Fingertips*. In ideal case, there are at most 5 fingertips detected. But the erroneous protrusions due to noises could be detected as fingertips; see Figure 12 for such a possible situation.

Figure 10: The image at the bottom shows the results after finishing the first round of Step 3.

Figure 11: The image at the bottom shows the results after finishing the second round of Step 3.

After getting the candidates of fingertips and in order to eliminate the erroneous fingertips and record the real ones, Algorithm B below is then applied.

Algorithm B

Step 1. Eliminate those candidates residing on the "Nonfingertip appearing area" indicated by the area with grey color in Figure 13.

Step 2. Cluster the candidates fingertips detected in previous module according to their proximity calculated by city block distance.

Step 3. The top 5 (at most) candidates in the image, one from each clustering group, are selected as the real fingertips.

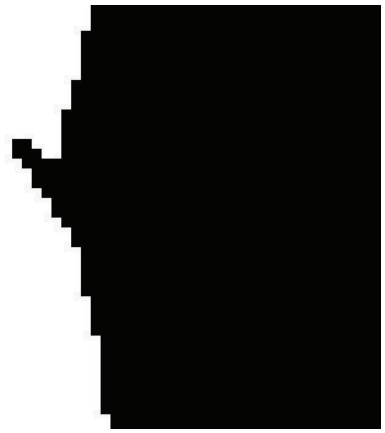


Figure 12: The protrusion at the boundary of one finger is erroneously detected as another finger.

(7) *Gesture Recognition*. The two positions of palm and wrist had been obtained through the OPENNI framework in advance and, with these two positions at hand, we can detect the positions of fingertips using the modules described previously. Therefore, the gesture can be recognized by the positions, distances, and angles between these positions. In this study, the clench gesture denotes that the user wants to grasp the object appearing at the corresponding position of the hand. And moving the hand with clench gesture indicates that the user wants to move the 2D object or rotate the 3D object. The opening gesture will break the link between the hand and its corresponding object. In addition, many other gestures can also be recognized and thus different applications can be created.

3.3. The Construction of Somatosensory Game. Based on the teaching theory and advantages illustrated by the scholars above, the somatosensory games proposed in this study integrate five elements: challenge, interaction, rules, goal, and social relationship. We propose two somatosensory games: puzzle game and geometric game.

In the puzzle game, the first phase is scoring model, students can freely explore the fun of somatosensory system and fulfill the puzzle mission. Each operation will bring back feedbacks; for example, the action of capturing images would produce corresponding sound and icons to remind students, and puzzles placed into the correct position and the wrong position will trigger a corresponding animation and sound alarm, respectively, to reward and punish students. The second phase is time keeping model; after the first phase, the students have been familiar with the operation of the game. In this phase, the students will be put into the time pressure, to improve and train their concentration. Under this time keeping model, as time goes by, warning animation will become increasingly apparent and sound will be more rapid to force the students under time pressure with high concentration completing the puzzle as Figures 14 and 15 show. By this way, the concentration and logic of the students can be effectively improved.

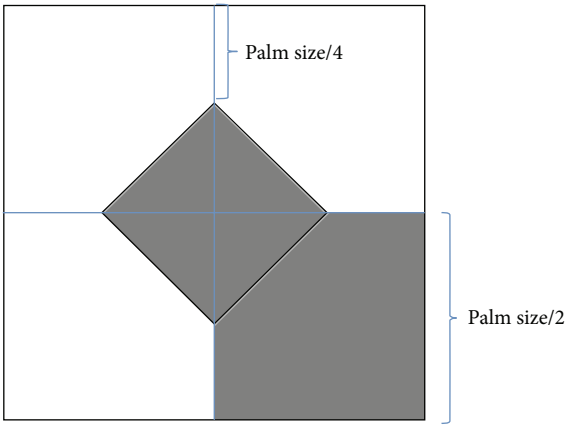


Figure 13: The grey color indicates the "Nonfingertip appearing area," and this area is derived from the right hand.

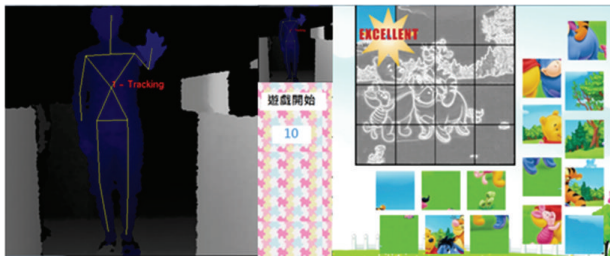


Figure 14: The procedure of puzzle game.

In the first phase of scoring model, the main purpose is guiding the users to be familiar with the operation of the somatosensory system and raise their interests. In this stage, the elements of intrinsic motivations such as interaction, rules, and goal are immersed into the users. Then, in the second stage, the main goal is to raise the attention of users to the puzzle game and their reasoning ability. In this second stage, the elements of intrinsic motivations such as challenge, risk, and goal are immersed into the users.

Next, we also proposed another geometric game and the main objective of the game is to use a virtual 3D objects with the somatosensory system that allows students to intuitively and easily view and rotate the virtual 3D objects and learn more about 3D geometric concepts through this way. The game is divided into two stages. The first stage is paper testing, the learners should be complete 25 questions within a fixed time. Each question consists of four geometric patterns, in which a geometric pattern is the target and the rest are the optional ones. Learners choose from three optional ones having the same volume with the target. If the answer is correct, they gain the bonus points. At this stage, the students evaluate their ability of geometry reasoning. Then, the second phase of gameplay consists of 25 questions. Each question would reveal a target object at the top of the screen, and there are three different candidate objects at the bottom. The user can optionally view and rotate these objects and select one of three objects within the stipulated time. As Figures 16 and 17

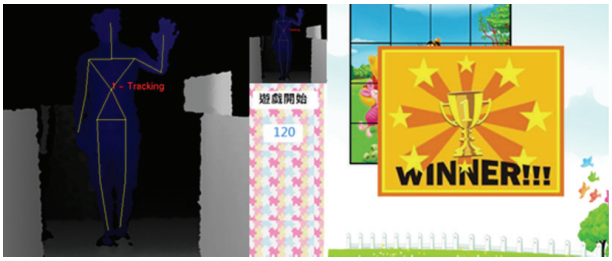


Figure 15: The completeness of puzzle game.

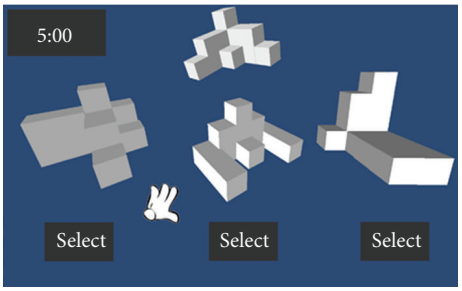


Figure 16: The illustration of geometric game.

illustrate, if the selected objects have the same volume with the target object, it is correct otherwise an error.

3.4. The Application of Somatosensory in Learning. With the growing technical of somatosensory system, it has become common in the entertainment game industry. But somatosensory can not only bring people the joy of playing games. Unlike the usual way, it is a better and more appealing man-machine interactive media and also an excellent learning tool in the field of digital learning (especially young children learning). There are already many scholars and industry investing in developing. The application of somatosensory is wide; in addition to the use of games, medical industry, education institutions, and police investigation could take advantage of this unique and fun way to get a different feeling in the learning application. When somatosensory is applied in learning, students' interest in learning, intuitive in operation would be enhanced via manipulating the virtual characters and objects. In addition, similar to the authenticity of such a learning a learning environment; students will have a simple and intuitive feeling to increase the interest in study. For example, the learning of dancing in general environment must only be to imitate the way through movies. With the help of the somatosensory system, you can create a 3D virtual environment in the meanwhile. Then the learning performance could be estimated via student's dace choreography. Mathematical learning can also be used to aid in the somatosensory system. For example, facing the geometry problems with space concept, you can manipulate 3D models through somatosensory designed for students to experience it. Students can not only gain a more intuitive and more intense sensory stimulation and logic training, but also can enhance students' interest in learning. In this study, the main consideration is intuitive, interactive, and fun. It

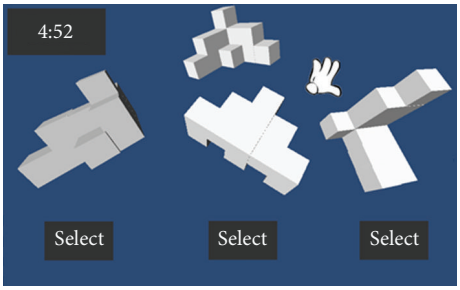


Figure 17: Another illustration of geometric game.

combines the original puzzle and somatosensory, to attempt to allow students to intuitively operate the puzzle in virtual world and to train the logical thinking of students through puzzles and enhance the students' interest in learning to achieve happy learning.

3.5. Analysis of Learning Speed and Experience. This study divided puzzle game into two phases: scoring mode and timing mode. In score mode, the main consideration is for students to familiarize with the somatosensory environment and stimulation. It encourages students to try to complete the puzzle with scoring points. The second phase timing mode is to analyze through students' learning speed and score. The effects of somatosensory on learning speed could be analyzed by estimating students' scoring points under the pressure of time. On the other hand, geometric learning games are also divided into two stages. The first stage is in the traditional way on paper test; the second stage is to allow students to manipulate virtual objects. After completion of the test, we survey the feelings and perception of somatosensory games as the basis for analysis of learners.

4. Experimental Results

In this section, experimental results conducted on the puzzle game and geometric game are illustrated to demonstrate the effectiveness of the proposed method. In the puzzle game, there are two modes in this experiment. Each mode processes one game teaching. In puzzle game, the first mode adopts scoring mode game, whose point is to make learners somatosensory game and be familiar with the operation of puzzle environment. The second mode adopts timing mode, whose purpose is to measure and observe the learners' learning ability and their attention and logical thinking on puzzle game under time pressure. There will be two times of timing mode. We record separately the learner's scores as reference of learning speed. The practical environment is a multimedia computer classroom. Besides a projector, there are Kinect camera and 25 computers. The teacher's operation teaching time is one class per unit. Six minutes in class for each learning comprehension test and 50 minutes for each class. Hence, two units consume two classes. The place of this experiment is in the multimedia computer class. Two learners use one computer. The researcher of this experiment is not the class teacher. The teacher teaches with the materials. Before teaching, the researchers explain the purpose of the research and remind the teachers' materials and the operation and notes of this practice. During the game, the researchers will aid the teachers on teaching and assist the learners to solve the problems of learning. The learners of this research are 51 G1 students, 27 boys and 24 girls; 67 G3 students, 38 boys and 29 girls; 68 G5 students, 38 boys and 30 girls. There are total 186 students. The experiment is divided into 3 levels, easy, medium, and difficult. The difficulty is based on the numbers of puzzles: 16, 64, and 81 puzzles. There are two parts of results. Part 1 is to analyze the learners' learning speed and their scores. See Figure 18.

Table 1: Results of the survey.

Learning state	Mean (M)	Standard deviation (SD)	Sample (N)
Interest/enjoyment	5.47	1.62	186
Perceived competence	4.69	1.81	186
Effort/importance	5.24	1.44	186
Pressure/tension	2.74	2.00	186

According to Figure 18, we can find that the second test scores are mostly higher than the first test scores. This is because after the first test, the learners are getting familiar with the somatosensory game. Therefore, they get better scores on the second test. Because each learner's score on these two stages is not the same, it presents the difference between learners' learning speed on somatosensory system. We can also understand the effect and influence of learners on somatosensory system with the survey on second stages. The contents of the survey are as Table 3(a). The results of survey are as Table 1. As shown in Figure 18, the second test scores are mostly higher than the first ones. The learners are familiar with the somatosensory game after the first test. Therefore, they get better scores on the second test. According to different learner's score on these two tests, the analysis of learning speed with the somatosensory system can be obtained. To get the effect and influence of learners on somatosensory system, we provide the questionnaire as shown in Table 3 to analyze the impact and feelings of learner with the somatosensory system. The results are listed in Table 1. From the result, the new stimulus brought by somatosensory system can intrigue learners' interest on this game, which makes learners concentrate on the game and develop their logical and thinking ability into the game. They also meet the challenge and gain achievement among it.

In addition, in the geometry game, there are two steps in this experiment. We use traditional paper test in the first step. The content of test sets an object as a goal which is the same size with targeted object and there are 25 questions with 3 choices. Five points for each question. The questions in the second stage are the same with those in the first stage. However, they are presented with 3D virtual objects and the learners operate it with somatosensory system as an interface, which is based on hand gestures to allow learners to rotate and see the 3D virtual objects. This research is practiced in the same place with the puzzle game. The subjects are 141 G5

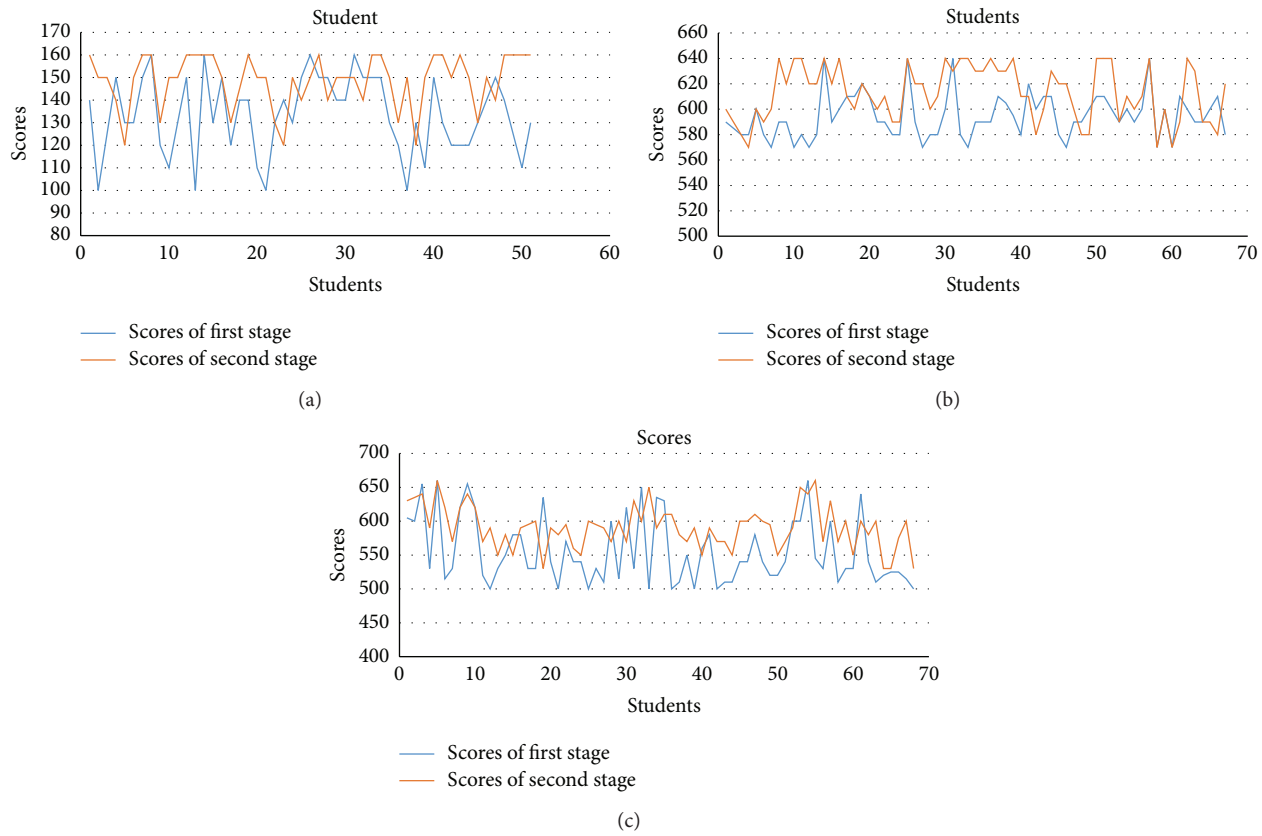


Figure 18: (a) is the result of G1 students. (b) is the result of G3 students. (c) is the result of G5 students.

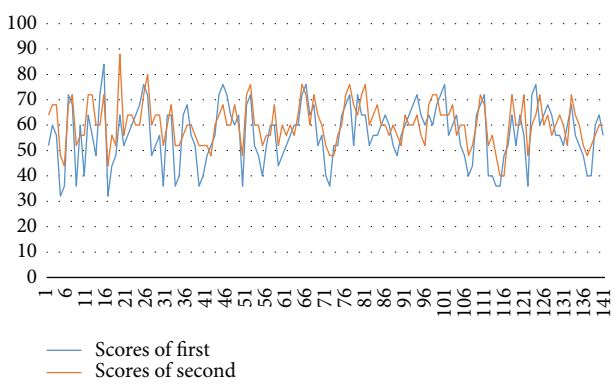


Figure 19: Distribution of test scores.

students, 72 boys and 69 girls. The aim is to observe the level of how learners enjoy and how much interest is intrigued, the learning speed, and learning status when they face new learning approaches. This part is conducted by a survey. The result of test is as Figure 19. The result of survey is as Table 2. The content of survey is as Table 3(b).

As we can see in Figure 19, learners score higher with new test approach than traditional paper test. Especially the group of lower scores, they progress enough. This is because learners can rotate the virtual objects through somatosensory system. By this kind of interactive approach, the freedom of learning has been improved and the learners understand geometry better. From the result of the survey, most learners think that this new approach is really helpful.

Table 2: Results of the survey.

	Score of average (M)	Standard deviation (SD)	Sample (N)
Great help in learning	5.68	1.64	141
Little help in learning	3.94	1.49	141
No help for learning	3.24	1.93	141

5. Conclusions

Somatosensory system is an interactive media system which has risen recently. Because of its directness, the users gain more feedback during the operation. Therefore, in this study, we proposed a somatosensory based learning system and discuss the influence and the change brought by somatosensory system. In somatosensory system, we create a puzzle game suitable for students, which is built by body skeleton and human behavior. Also, by doing this, we observe how helpful somatosensory game is and if the learners can learn happily during this somatosensory game. From the result, most learners hold a positive attitude when using this new approach to learn. Among all the factors, including interest/enjoyment, perceived competence, effort/importance, pressure/tension,

Table 3: Contents of the questionnaire in this study.

(a)

After you finished the game, please fill up the form below. Make scores of items according to what you felt in the game.							
(1) I felt very happy when I was playing the game	7	6	5	4	3	2	1
(2) I think I had done well in the game	7	6	5	4	3	2	1
(3) I did a lot of effort in the game	7	6	5	4	3	2	1
(4) I got no pressure when I was playing the game	7	6	5	4	3	2	1
(5) This game is very interesting	7	6	5	4	3	2	1
(6) I think I can do a better job next time	7	6	5	4	3	2	1
(7) I had tried to perform better in the game	7	6	5	4	3	2	1
(8) I felt nervous when I was playing the game	7	6	5	4	3	2	1
(9) I think this game makes me feel happy	7	6	5	4	3	2	1
(10) I'm satisfied to my performance in the game	7	6	5	4	3	2	1
(11) I really do a lot of efforts in the game	7	6	5	4	3	2	1
(12) I felt relaxed when I was playing the game	7	6	5	4	3	2	1
(13) Playing the puzzle game makes me very attentive	7	6	5	4	3	2	1
(14) I think I'm good at playing the puzzle game	7	6	5	4	3	2	1
(15) I think this activity is important to me	7	6	5	4	3	2	1
(16) I feel anxiety when I was playing the game	7	6	5	4	3	2	1
(17) I think playing the game makes me feel happy	7	6	5	4	3	2	1
(18) I think I'm good at playing puzzle games	7	6	5	4	3	2	1
(19) I spend a lot of energy in this game	7	6	5	4	3	2	1
(20) This activity makes me feel pressure	7	6	5	4	3	2	1

(b)

After you finished the game, please fill up the form below. Make scores of items according to what you felt in the game.							
(1) This game is helpful for learning geometric concepts	7	6	5	4	3	2	1
(2) This game is not very helpful for my learning	7	6	5	4	3	2	1
(3) I think I learned nothing in this activity	7	6	5	4	3	2	1
(4) I was very focused on playing the game	7	6	5	4	3	2	1
(5) I think this game is not very fun	7	6	5	4	3	2	1
(6) I felt nervous when I was playing the game	7	6	5	4	3	2	1
(7) I have learned a lot from this activity	7	6	5	4	3	2	1
(8) I think the learning effect from this activity is limited	7	6	5	4	3	2	1
(9) I learned nothing from this activity	7	6	5	4	3	2	1
(10) I'm satisfied to my performance in the game	7	6	5	4	3	2	1
(11) This activity has no effect on me	7	6	5	4	3	2	1
(12) I think this activity is not good	7	6	5	4	3	2	1
(13) I think learning by playing game is a good way	7	6	5	4	3	2	1
(14) I think learning by playing game is a common way	7	6	5	4	3	2	1
(15) I think learning by playing game is not good	7	6	5	4	3	2	1
(16) I think this way to learn is better than traditional way	7	6	5	4	3	2	1
(17) I think this learning way is not special for me	7	6	5	4	3	2	1
(18) I prefer traditional way to learning	7	6	5	4	3	2	1
(19) I spend a lot of energy in this game	7	6	5	4	3	2	1
(20) I'm not interested in this activity	7	6	5	4	3	2	1
(21) This activity makes me feel pressure	7	6	5	4	3	2	1

enjoyment factor is the learners gain the best. Because the specialty and fun of somatosensory game intrigues learners' interest, they learn and play simultaneously. This kind of somatosensory system provides users with different ways to interact with computers and draw more attention from users. Meanwhile, the users can operate it more fluently and interactively.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] J. G. Hogle, *Considering Games as Cognitive Tools: In Search of Effective Edutainment*, University of Georgia Department of Instructional Technology, 1996.

[2] C. Albitar, P. Graebling, and C. Doignon, "Robust structured light coding for 3D reconstruction," in *Proceedings of the 11th IEEE International Conference on Computer Vision (ICCV '07)*, October 2007.

[3] J. Shotton, A. Fitzgibbon, M. Cook et al., "Real-time human pose recognition in parts from single depth images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1297–1304, June 2011.

[4] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 415–422, November 2011.

[5] D. Grest, J. Woetzel, and R. Koch, "Nonlinear body pose estimation from depth images," in *Proceedings of the 27th DAGM (German Association for Pattern Recognition) Symposium (DAGM '05)*, pp. 285–292, September 2005.

[6] J. Charles and M. Everingham, "Learning shape models for monocular human pose estimation from the Microsoft Xbox Kinect," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV '11)*, pp. 1202–1208, November 2011.

[7] F. A. Kondori, S. Yousefi, H. Li, S. Sonning, and S. Sonning, "3D head pose estimation using the Kinect," in *Proceedings of the International Conference on Wireless Communications and Signal Processing (WCSP '11)*, November 2011.

[8] G. Sheasby, J. Warrell, Y. Zhang, N. Crook, and P. H. S. Torr, "Simultaneous human segmentation, depth and pose estimation via dual decomposition," in *Proceedings of the 4th UK Computer Vision Student Workshop (BMVW '12)*, 2012.

[9] G. Ye, Y. Liu, N. Hasler, X. Ji, Q. Dai, and C. Theobalt, "Performance capture of interacting characters with handheld Kinects," in *Proceedings of the European Conference on Computer Vision (ECCV '12)*, 2012.

[10] K. Lai, J. Konrad, and P. Ishwar, "A gesture-driven computer interface using Kinect," in *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI '12)*, pp. 185–188, 2012.

[11] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '12)*, pp. 20–27, Providence, RI, USA, 2012.

[12] S. Sempena, N. U. Maulidevi, and P. R. Aryan, "Human action recognition using dynamic time warping," in *Proceedings of the International Conference on Electrical Engineering and Informatics (ICEEI '11)*, July 2011.

[13] A. Abramov, K. Pauwels, J. Papon, F. Worgotter, and B. Dellen, "Depth-supported real-time video segmentation with the Kinect," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV '12)*, 2012.

[14] C. J. Taylor and A. Cowley, "Fast scene analysis using image and range data," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '11)*, pp. 3562–3567, 2011.

[15] A. Harati, S. Gächter, and R. Siegwart, "Fast range image segmentation for indoor 3D-SLAM," in *Proceedings of the 6th IFAC Symposium on Intelligent Autonomous Vehicles (IAV '07)*, pp. 475–480, Pierre Baudis, France, September 2007.

[16] A. Baak, M. Muller, G. Bharaj, H.-P. Seidel, and C. Theobalt, "A data-driven approach for real-time full body pose reconstruction from a depth camera," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 1092–1099, November 2011.

[17] H.-D. Yang and S.-W. Lee, "Reconstruction of 3D human body pose from stereo image sequences based on top-down learning," in *Proceedings of the International Conference on Pattern Recognition (ICPR '06)*, 2006.

[18] M. Zeng, Z. Liu, Q. Meng, Z. Bai, and H. Jia, "Motion capture and reconstruction based on depth information using Kinect," in *Proceedings of the 5th International Congress on Image and Signal Processing (CISP '12)*, pp. 1381–1385, 2012.

[19] L. Rogge, T. Neumann, M. Wacker, and M. A. Magnor, "Monocular pose reconstruction for an augmented reality clothing system," in *Proceedings of the Vision, Modeling, and Visualization Workshop (VMV '12)*, Berlin, Germany, 2012.

[20] R. A. Newcombe, S. Izadi, O. Hilliges et al., "KinectFusion: Real-time dense surface mapping and tracking," in *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '11)*, pp. 127–136, October 2011.

[21] H. Du, P. Henry, X. Ren et al., "Interactive 3D modeling of indoor environments with a consumer depth camera," in *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*, pp. 75–84, September 2011.

[22] E. S. Sarmanho, E. S. Barros, D. C. Monteiro, L. B. Marques, and D. G. de Souza, in *Proceedings of the 10th Brazilian Symposium on Computer Games and Digital Entertainment*, Federal University of Sao Carlos, Study Laboratory of Human Behavior, 2011.

[23] E. M. Kenneth, J. V. Hansen, M. Andersen, and S. Safiri, *Physical-Digital Interaction Design for Children*, The Maersk Mc-Kinney Moller Institute, University of Southern Denmark Niels, Odense, Denmark, 2012.

[24] C.-L. Lien, C.-Y. Huang, C.-Y. Wang, and G.-D. Chen, "Using Kinect to track learning behavior of students in the classroom as video Portfolio to enhance reflection learning," in *Proceedings of the 20th International Conference on Computers in Education (ICCE '12)*, 2012.

[25] X. Casas, G. Herrera, I. Coma, and M. Fernández, "A Kinect-based augmented reality system for individual with autism spectrum disorders," in *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications (GRAPP/IVAPP '12)*, pp. 440–446, Rome, Italy, 2012.

[26] A. Smorkalov, M. Fominykh, and E. Prasolova-Forland, "Virtualizing real-life lectures with academia and kinect," in *Proceedings of the IEEE Virtual Reality Conference*, 2013.

[27] E. Tuveri, S. A. Iacolina, F. Sorrentino, L. D. Spano, and R. Scateni, "Controlling a planetarium software with a Kinect or in a multi-touch table: a comparison," in *Biannual Conference of the Italian Chapter of SIGCHI*, ACM, Trento, Italy, September 2013.

[28] Y. Li, "Protractor: a fast and accurate gesture recognizer," in *Proceedings of the 28th Annual CHI Conference on Human Factors in Computing Systems (CHI '10)*, pp. 2169–2172, April 2010.

[29] S. Kratz and M. Rohs, "A % gesture recognizer: simple gesture recognition for devices equipped with 3D acceleration sensors," in *Proceedings of the 14th ACM International Conference on Intelligent User Interfaces (IUI '10)*, pp. 341–344, February 2010.

[30] S. Kratz and M. Rohs, "Protractor3D: a closed-form solution to rotation-invariant 3D gestures," in *Proceedings of the 15th ACM International Conference on Intelligent User Interfaces (IUI '11)*, pp. 371–374, February 2011.

[31] P. Hong, M. Turk, and T. S. Huang, "Gesture modeling and recognition using finite state machines," in *Proceedings of the 4th IEEE International Conference on on Face and Gesture Recognition*, Grenoble, France, 2000.

[32] J. O. Wobbrock, A. D. Wilson, and Y. Li, "Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes," in *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*, pp. 159–168, October 2007.

Hindawi Publishing Corporation
The Scientific World Journal
Volume 2013, Article ID 618649, 12 pages
http://dx.doi.org/10.1155/2013/618649

Research Article

Real-Time EEG-Based Happiness Detection System

Noppadon Jatupaiboon,[1] Setha Pan-ngum,[1] and Pasin Israsena[2]

[1] *Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand*
[2] *National Electronics and Computer Technology Center, Pathumthani 12120, Thailand*

Correspondence should be addressed to Noppadon Jatupaiboon; noppadon_j@hotmail.com

Received 3 June 2013; Accepted 15 July 2013

Academic Editors: B.-W. Chen, S. Hsieh, and C.-H. Wu

We propose to use real-time EEG signal to classify happy and unhappy emotions elicited by pictures and classical music. We use PSD as a feature and SVM as a classifier. The average accuracies of subject-dependent model and subject-independent model are approximately 75.62% and 65.12%, respectively. Considering each pair of channels, temporal pair of channels (T7 and T8) gives a better result than the other area. Considering different frequency bands, high-frequency bands (Beta and Gamma) give a better result than low-frequency bands. Considering different time durations for emotion elicitation, that result from 30 seconds does not have significant difference compared with the result from 60 seconds. From all of these results, we implement real-time EEG-based happiness detection system using only one pair of channels. Furthermore, we develop games based on the happiness detection system to help user recognize and control the happiness.

1. Introduction

The aim of human computer interaction (HCI) is to improve the interactions between human and computers. Because most computers lack of understanding of user's emotions, sometimes they are unable to respond to the user's needs automatically and correctly [1]. One of the most interesting emotions is happiness. world happiness report reflects a new worldwide demand for more attention to happiness and absence of misery as criteria for government policy [2]. Being happy is related to many positive effects including confidence, optimism, self-efficacy, likability, activity, energy, physical well-being, flexibility, creativity, and the ability to cope with stress [3]. All of these benefits are the reasons why we should be happy.

In the past decades, most of emotion recognition researches have only focused on using facial expressions and speech. However, it is easy to fake facial expressions or change tone of speech and these signals are not continuously available, and they differ from using physiological signals, which occur continuously and are hard to conceal, such as Galvanic Skin Response (GSR), Electrocardiogram (ECG), Skin Temperature (ST), and, especially, Electroencephalogram (EEG). EEG is the signal from voltage fluctuations in the brain, that is, the center of emotions [1, 4]. Emotions are thought to be related with activity in brain areas that direct our attention, motivate our behavior, and determine the significance of what is going on around us. Emotion is related with a group of structures in the center of the brain called limbic system, which includes amygdala, thalamus, hypothalamus, and hippocampus [5, 6].

Electroencephalogram (EEG) is the recording of electrical activity on the scalp. EEG measures voltage changes resulting from ionic current flows within the neurons of the brain. There are five major brain waves distinguished by their different frequency bands (number of waves per second) as shown in Figure 1. These frequency bands from low to high frequencies, respectively, are called Delta (1–3 Hz), Theta (4–7 Hz), Alpha (8–13 Hz), Beta (14–30 Hz), and Gamma (31–50 Hz). Figure 2 shows the 10–20 system of electrode placement, that is, an internationally recognized method to describe and apply the location of scalp electrodes. Each site has a letter to identify the lobe and a number to identify the hemisphere location [7, 8].

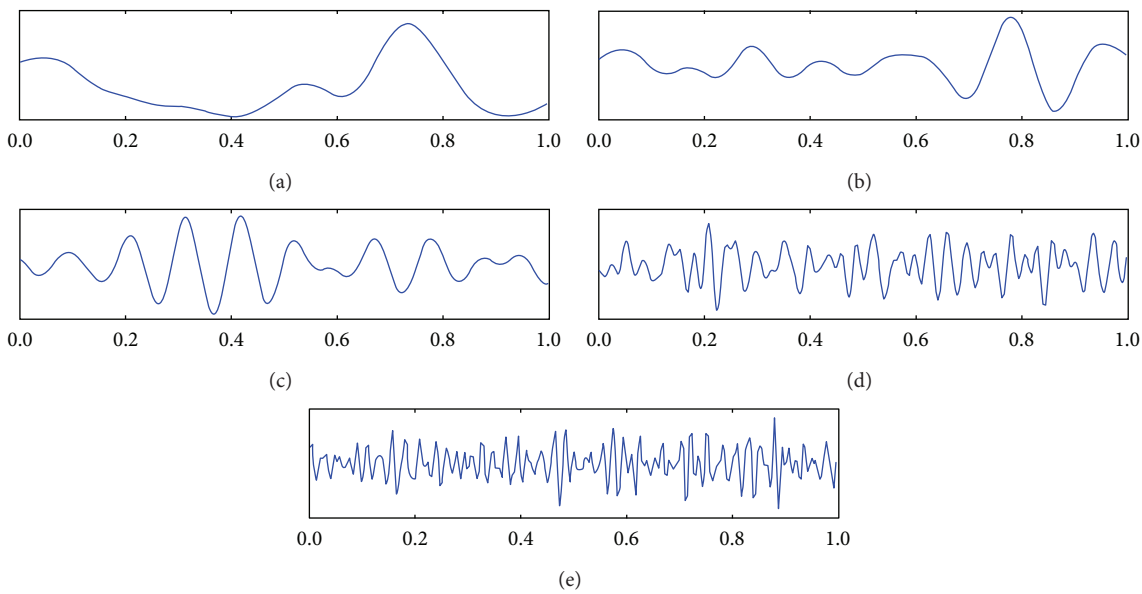


Figure 1: Brainwave: (a) Delta, (b) Theta, (c) Alpha, (d) Beta, and (e) Gamma [9].

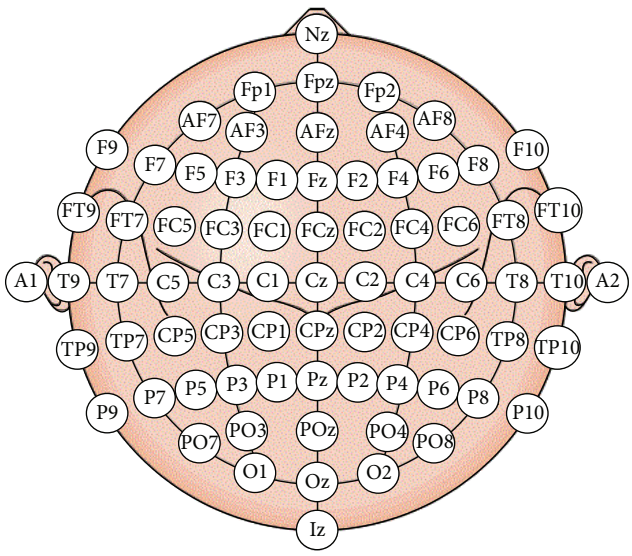


Figure 2: International 10–20 system of electrode placement [7].

2. The Literature Review

Nowadays, the EEG-based emotion recognition researches are highly active. The goal of these is to find suitable technique giving a good result that eventually can be implemented in real-time emotion recognition. The list of the EEG-based emotion recognition researches is shown in Table 1. It is difficult to compare results among them because there are a lot of factors that make different results from different researches including participant, model of emotion, stimulus, feature, temporal window, and classifier. The main six factors are described next to clarify the understanding.

2.1. Participant. The larger number of participants makes more reliable result. Moreover, we can divide the method for building emotion classification into subject-dependent and subject-independent models. The second model is harder than the first model due to interparticipants variability [10, 11]. The subject-dependent model avoids the problems related to interparticipant but a new classification model must be built for every new user. In this research, we build both subject-dependent and subject-independent models to compare the results.

2.2. Model of Emotion. The larger number of emotions makes emotion recognition harder, and some emotions may overlap. A good model of emotion should clearly separate these emotions. Several models have been proposed such as basic emotion and dimensional model. The most widely used basic emotions are the 6 basic emotions (i.e., anger, disgust, fear, joy, sadness, and surprise) that have been mostly used in facial expression recognition [12]. The common dimensional model is characterized by two main dimensions (i.e., valence and arousal). The valence emotion ranges from negative to positive, whereas the arousal emotion ranges from calm to excited [13]. This model is used in most researches because it is easier to express an emotion in terms of valence and arousal rather than basic emotions that can be confused by emotion names [14]. As shown in Figure 3, the emotions in any coordinates of the dimensional model are shown by facial expression. In this research, we use the dimensional models. The emotions used are happy and unhappy (sad). The happy emotion has positive valence and low arousal whereas the unhappy emotion has negative valence and low arousal.

2.3. Stimulus. There are various methods for emotion elicitation, which are self-eliciting, recalling, and using external stimulus such as picture, sound, and odor. The widely used

Table 1: EEG-based emotion recognition researches.

References	Year	Participant	Emotion	Stimulus	Feature	Temporal window	Classifier	Result	Real time
[10]	2006	4 subject-dependent	3 arousal classes	Picture	PSD	—	NB	58%	No
[11]	2008	26 subject-independent	4 classes (joy, anger, sadness, and pleasure)	Music	ASM	1 s	SVM	92.73%	No
[20]	2009	10 subject-dependent	2 valence classes	Picture	CSP	3 s	SVM	93.5%	No
[21]	2009	10 —	3 arousal classes	Recall	PSD	0.5 s	SVM	63%	No
[22]	2009	1 subject-dependent	3 classes (positively excited, negatively excited, and calm)	Picture	statistical features	—	QDA	66.66%	No
[23]	2009	3 subject-dependent	10 classes	Self-elicited	PSD	1 s	KNN	39.97–66.74%	No
[24]	2010	26 subject-independent	4 classes (joy, anger, sadness, and pleasure)	Music	ASM	1 s	SVM	82.29%	No
[25]	2010	6 subject-dependent	2 valence classes 2 arousal classes	Music video	PSD	—	SVM	58.8% (valence) 55.7% (arousal)	No
[26]	2010	26 subject-dependent	4 classes (calm, happy, sad, and fear)	Picture and music	SOM	2 s	KNN	84.5%	No
[28]	2010	15 —	2 classes (calm-neutral and negatively excited)	Picture	HOS	2 s	SVM	82%	No
[29]	2010	12 subject-dependent	2 valence classes 2 arousal classes	Sound	FD	—	threshold	—	Yes
[27]	2011	20 —	5 classes (happy, disgust, surprise, fear, and neutral)	video clip	Entropy	—	KNN	83.04%	No
[31]	2011	6 subject-dependent	2 valence classes	Movie clip	PSD	1 s	SVM	87.53%	No
[32]	2011	20 subject-independent	3 classes (boredom, engagement, and anxiety)	Game	PSD	—	LDA	56%	No
[33]	2011	5 subject-dependent	4 classes (joy, relax, sad, and fear)	Movie	PSD	1 s	SVM	66.51%	No
[34]	2011	11 —	3 valence classes	Picture	ASM	4 s	KNN	82%	No
[30]	2012	27 subject-independent	3 valence classes 3 arousal classes	Video	PSD and ASM	—	SVM	57.0% (valence) 52.4% (arousal)	No
[35]	2012	32 —	2 valence classes 2 arousal classes	Music video	PSD and ASM	—	NB	57.6% (valence) 62.0% (arousal)	No

Table 1: Continued.

References	Year	Participant	Emotion	Stimulus	Feature	Temporal window	Classifier	Result	Real time
[36]	2012	20 subject-dependent	5 classes (happy, angry, sad, relaxed, and neutral)	Picture	FD	—	SVM	70.5%	Yes
[37]	2012	5 subject-dependent	3 classes (positively excited, negatively excited, and calm)	Picture	HOC	—	KNN	90.77%	No
[38]	2012	4 —	2 valence classes 2 arousal classes	Video clip	ASP	—	—	66.05% (valence) 82.46% (arousal)	No
[39]	2012	32 —	2 classes (stress and calm)	Music video	PSD	—	KNN	70.1%	No
[40]	2012	36 —	3 classes	Music video	PSD	—	ANN	—	Yes
[41]	2013	11 subject-independent	2 valence classes	Picture	PSD	4 s	SVM	85.41%	No

*The feature, temporal window, and classifier shown in this table are the sets giving the best accuracy of each research.
Feature: Power Spectral Density (PSD), Spectral Power Asymmetry (ASM), Common Spatial Pattern (CSP), Higher Order Crossings (HOC), Self-Organizing Map (SOM), Higher Order Spectra (HOS), Fractal Dimension (FD), and Asymmetric Spatial Pattern (ASP).
Classifier: Support Vector Machine (SVM), Naïve Bayes (NB), Quadratic Discriminant Analysis (QDA), K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN).

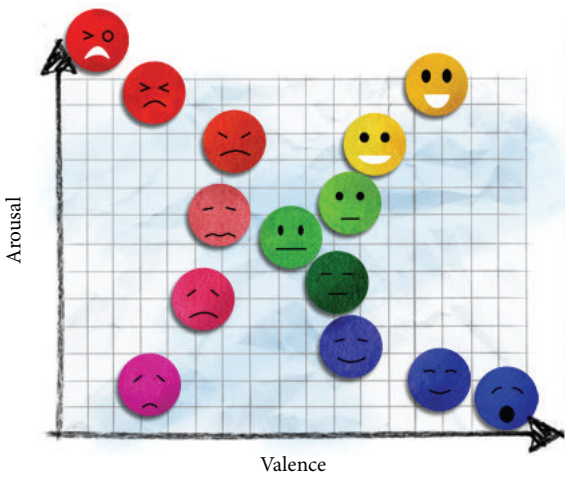


Figure 3: Dimensional model of emotion [14].

databases for emotion elicitation are International Affective Picture System (IAPS) [15] and International Digitized Sound System (IADS) [16]. These databases are generally accompanied by emotional evaluations from average judgments of several people. In this research, we choose pictures from Geneva Affective Picture Database (GAPED) [17] and sounds from classical emotion elicitation, because using visual-audio stimulus gives a better result than using either visual stimulus or audio stimulus [18].

2.4. Feature. Several signal characteristics of EEG have been used to be the features. The widely used feature is Power Spectral Density (PSD), the power of the EEG signal in focused frequency bands. In addition, others such as Spectral Power Asymmetry (ASM), Common Spatial Pattern (CSP), Higher Order Crossings (HOC), Self-Organizing Map (SOM), Higher Order Spectra (HOS), Fractal Dimension (FD), Asymmetric Spatial Pattern (ASP), and Entropy have been used as features and some give a good result. In this research, the feature we use is PSD since it gives a good performance in several researches as shown in Table 1, and it uses relatively little computation, which is suitable to implement in real-time emotion recognition.

2.5. Temporal Window. The appropriate length of temporal window depends on a type of emotion and physiological signal. Overall duration of emotions approximately falls between 0.5 and 4 seconds [42]. By using unsuitable window, the emotion may be misclassified because different emotions may be covered when too long or too short periods are measured. The existing literature does not provide suitable window size to be used to achieve optimal EEG-based emotion recognition [4]. In this research, we use temporal window 1 second.

2.6. Classifier. Several machine learning algorithms have been used as emotion classifiers such as Support Vector Machine (SVM), Naïve Bayes (NB), Quadratic Discriminant Analysis (QDA), K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), and Multilayer Perceptron (MLP). As shown in Table 1, SVM is implemented on many emotion classification researches because of many advantages. SVM is known to have good generalization properties and to be insensitive to overtraining and curse of dimensionality. The basic training principle of SVM is finding the optimal

hyperplane where the expected classification error of test samples is minimized. The optimal hyperplane is the one that maximizes the margins. Maximizing the margins is known to increase the generalization capability. SVM uses regularization parameter (C) that enables accommodation to outliers and allows errors on the training set [43]. In this research, we use Gaussian SVM to be a classifier.

Beside the aforementioned factors, there is a factor that affects classification results from different researches. We found that some researches did not separate training set and test set completely although they did cross-validation (CV). Because simple cross-validation method randomly selects some data to be test set and the rest of data to be training set, some training data and test data may be in the same trial. Although the offline result is good, it does not guarantee the online result. In online emotion recognition, the training set is used to build the classification model, and the test set is a data from real-time EEG, so the training data and the test data are absolutely separated. For reliable result that can be guaranteed when using online emotion recognition, we should separate training set and test set completely. In this research, we use leave-one-trial-out cross-validation (LOTO-CV) and leave-one-subject-out cross-validation (LOSO-CV) for evaluating subject-dependent and subject-independent models, respectively.

As shown in Table 1, most of EEG-based emotion recognition researches are not for real-time implementation. There are a few researches that implement real-time emotion recognition such as [29, 40]. Wijeratne and Perera [40] proposed real-time emotion detection system using EEG and facial expression. However, the EEG signal acquisition part was still offline due to their time constraints, so they used prerecorded EEG data instead of real-time EEG data. Liu et al. [29] proposed real-time emotion detection system using EEG. The user emotions are recognized and visualized in real time on his/her avatar. However, there is an issue in their approach that needs to be mentioned. In order to recognize an emotion, they did not use classifier and they only compared the Fractal Dimension (FD) values with predefined threshold, but they did not show how to define that threshold.

To fulfill these, we intend to implement EEG-based emotion detection system that can be truly implemented in real-time. Due to real-time processing, minimum computation time is required. We compare results among each pair of channels and different frequency bands in order to reduce insignificant channels and frequency bands. Furthermore, we develop games based on the happiness detection system to recognize and control happiness.

3. Methodology

The process of emotion classification consists of several steps as shown in Figure 4. First of all a stimulus such as picture, audio, and movie is needed. During experiment, the participant is exposed to the stimuli to elicit emotion, and EEG signal is recorded accordingly. Then artifacts that contaminate EEG signal are removed. These EEG data are analyzed and relevant features are extracted. Some parts of data are trained to build classification model and the rest of data, which are test data, are classified using this model.

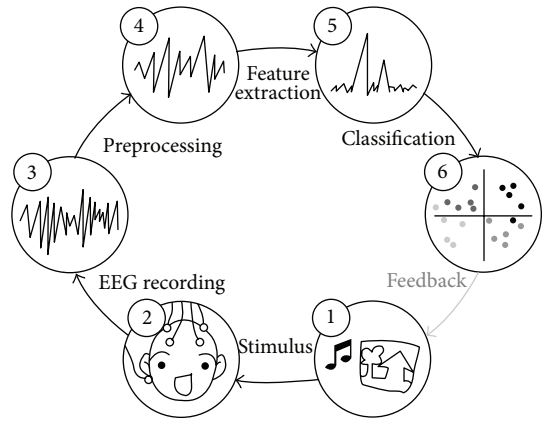


Figure 4: The process of emotion classification [19].

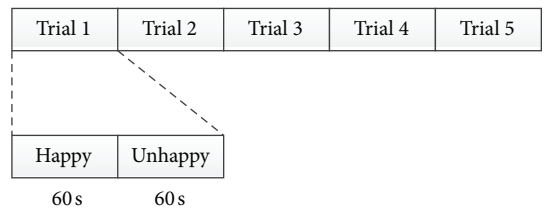


Figure 5: Procedure of experiment.

3.1. Stimulus. Both pictures and classical music were used to be the stimulus to elicit emotion. For pictures from GAPED [17], we selected the 50 highest valence scored pictures to be happy stimulus (i.e., pictures of human and animal babies as well as nature sceneries) and the 50 lowest valence scored pictures to be unhappy stimulus (i.e., pictures of human concerns and animal mistreatments). For classical music, we selected the highest and lowest valence scored pieces according to Vempala and Russo [44] to be happy and unhappy stimuli, respectively. The happy and unhappy pieces were Tritsch Tratsch Polka by Johann Strauss and Asas' Death by Edvard Grieg, respectively.

3.2. EEG Recording. We used 14-channels wireless EMOTIV [45] (i.e., AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1, and O2). The sampling rate is 128 Hz. The resolution is 16 bits (14 bits effective). Before recording EEG, we put EMOTIV on the participant's head for a while to prevent undesired emotions that can arise from unfamiliar or uncomfortable feelings. Then we described the process of recording and advised the participant to stay as still as possible to prevent artifact that can occur from moving the body. When the participant was ready, we then recorded EEG and the experiment was started. As shown in Figure 5, there were 5 trials, where each trial consisted of one happy and one unhappy stimulus. Each stimulus was composed of 10 pictures and 1 piece of classical music that played along for 60 seconds. After that, a blank screen was shown for 12 seconds to adjust participant's

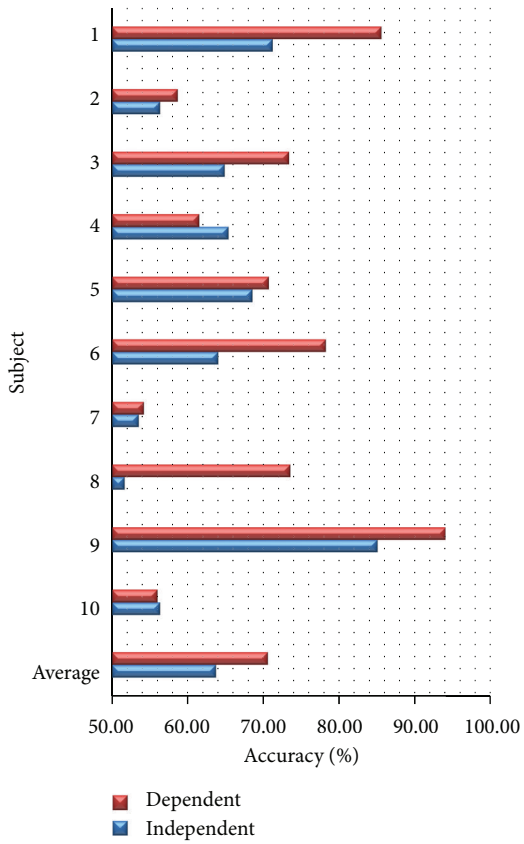


Figure 6: Accuracy from subject-dependent and subject-independent models.

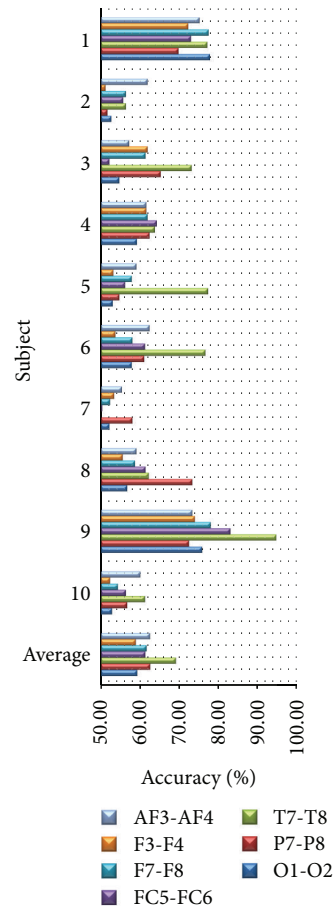


Figure 7: Accuracy from each pair of channels.

emotion to normal state and then the next stimulus was shown. When the 5 trials were completely shown, the process of recording ended. All these steps took approximately 15 minutes. There were 10 participants (i.e., 1 male and 9 females; average age is 34.60) taking part in this experiment.

3.3. Preprocessing. The EEG signal was filtered using a 5th-order sinc filter to notch out power line noise at 50 Hz and 60 Hz [45]. We removed baseline of the EEG signal for each channel so the values of the signal are distributed around 0.

3.4. Feature Extraction. The EEG signal with window 1 second was decomposed to 5 frequency bands that are Delta (0–4 Hz), Theta (4–8 Hz), Alpha (8–16 Hz), Beta (16–32 Hz), and Gamma (32–64 Hz) by Wavelet Transform as shown in Table 2. Then the PSD from each band was computed to be the feature. Since EMOTIV have 14 channels, the total features are 70. The features were normalized for each participant by scaling between 0 and 1 as shown in (1) to reduce inter-participant variability [11]:

$$\text{normalize}\left(X_i\right) = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}. \quad (1)$$

Since EEG signal from each trial has 120 seconds, there are 120 samples per trial. Due to 5 trials, there are 600 samples per participant. With 10 participants, the total samples are 6000. All samples were labeled whether happy or unhappy depending on the type of stimulus.

Table 2: EEG signal decomposition.

Frequency band	Frequency range (Hz)	Frequency bandwidth (Hz)	Decomposition level
Delta	0–4	4	A4
Theta	4–8	4	D4
Alpha	8–16	8	D3
Beta	16–32	16	D2
Gamma	32–64	32	D1

3.5. Classification. Gaussian SVM with leave-one-trail-out cross-validation (LOTO-CV) and leave-one-subject-out cross-validation (LOSO-CV) were used to compute accuracy for subject-dependent and subject-independent models, respectively. In the LOTO-CV method with 5 trials, one trial is set to be a test set and the rest to be a training set. Then the training set is built to be a classification model and the test set is classified using this model to evaluate accuracy. After that, we repeated the process using different trials as test sets, until all of the 5 trials had been test sets. The accuracy reported

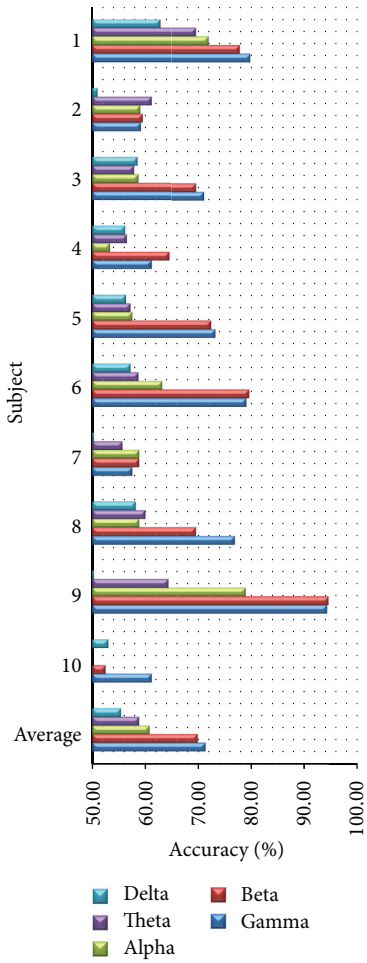


Figure 8: Accuracy from different frequency bands.

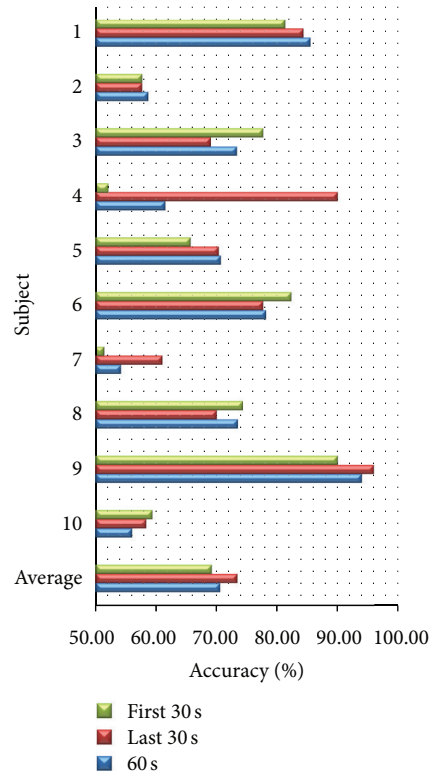


Figure 9: Accuracy from different time durations.

is the average accuracy of all 5 trials. The appropriate parameters are the set giving the best average of the 5 accuracies. In the LOSO-CV method with 10 subjects, one subject is set to be a test set and the rest to be a training set. Then the training set is built to be a classification model and the test set is classified using this model to evaluate accuracy. After that, we repeated the process using different subjects as test sets, until all of the 10 subjects had been test sets. The appropriate parameters are the set giving the best average of the 10 accuracies. The appropriate parameters C and γ of SVM were selected by grid search method. SVM implementation was done using LIBSVM [46].

4. Results and Discussion

4.1. Subject-Dependent and Subject-Independent Models. We compare subject-dependent and subject-independent accuracies using all features. As shown in Figure 6, we found that most of subject-independent accuracies are lower than subject-dependent accuracies. The average accuracies of subject-dependent model and subject-independent model are 70.55% and 63.67%, respectively. We can conclude that there are a lot of interparticipants. Different subjects may have different patterns of EEG when emotions are elicited. This conclusion is consistent with [24, 36]. As a result, we use only subject-dependent model to implement on real-time happiness detection system. Furthermore, we found that all of the older subjects (i.e., subject 2, 4, and 10; average age is 57.50) are giving low accuracies (accuracy of subject-dependent model lower than 65%). All of them confirm that they were elicited well by stimulus. We suppose as Levenson et al. [47] found that the magnitude of change in physiological signal was smaller in older than in younger subjects during emotion elicitation. So the accuracies of older subjects are low. When we exclude these older subjects, the average accuracies of subject-dependent model and subject-independent model are up to 75.62% and 65.12%, respectively.

4.2. Varying Pairs of Channels. We compare subject-dependent accuracy among each pair of channels (i.e., AF3-AF4, F3-F4, F7-F8, FC5-FC6, P7-P8, T7-T8, and O1-O2) using all frequency bands. As shown in Figure 7, we found that the highest average accuracy at 69.20% given by the pair of T7-T8 is very close to the average accuracy given by all channels. When we exclude older subjects, the average accuracy of T7-T8 is still highest at 72.90%. With PSD feature, we can conclude that temporal lobe is more effective for classifying happy and unhappy emotions than the others. This conclusion is consistent with [35, 48]. As a result, we can use this pair of channels instead of fourteen channels to reduce the number of channels and save computation time.

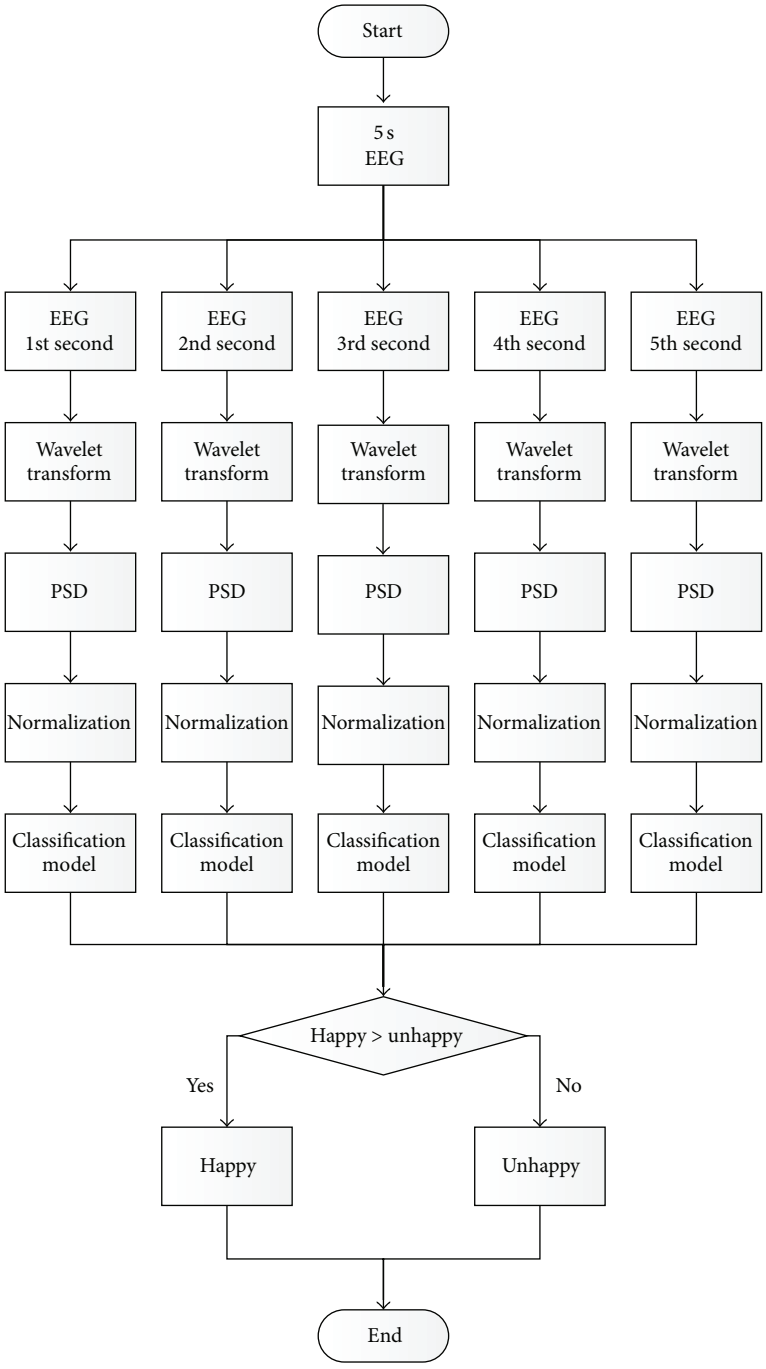


Figure 10: Flowchart of real-time happiness detection system.

4.3. Varying Frequency Bands. We compare subject-dependent accuracy among different frequency bands (i.e., Delta, Theta, Alpha, Beta, and Gamma) using all channels. As shown in Figure 8, we found that the average accuracies of Beta and Gamma are 69.83% and 71.28%, respectively, which are clearly higher than these of the other bands. When we exclude older subjects, the average accuracies of Beta and Gamma are still clearly higher than these of the other bands at 74.55% and 75.90%, respectively. With PSD feature, we can conclude that high frequency bands are more effective for classifying happy and unhappy emotions than low frequency bands. This conclusion is consistent with [20, 31, 48]. As a result, we can omit low-frequency bands such as Delta and Theta in order to save computation time.

4.4. Varying Time Durations. We compare subject-dependent accuracy from different time durations for emotion elicitation using all features. We consider accuracy from the first 30 seconds and the last 30 seconds of each stimulus. As shown in Figure 9, we found that the average accuracies

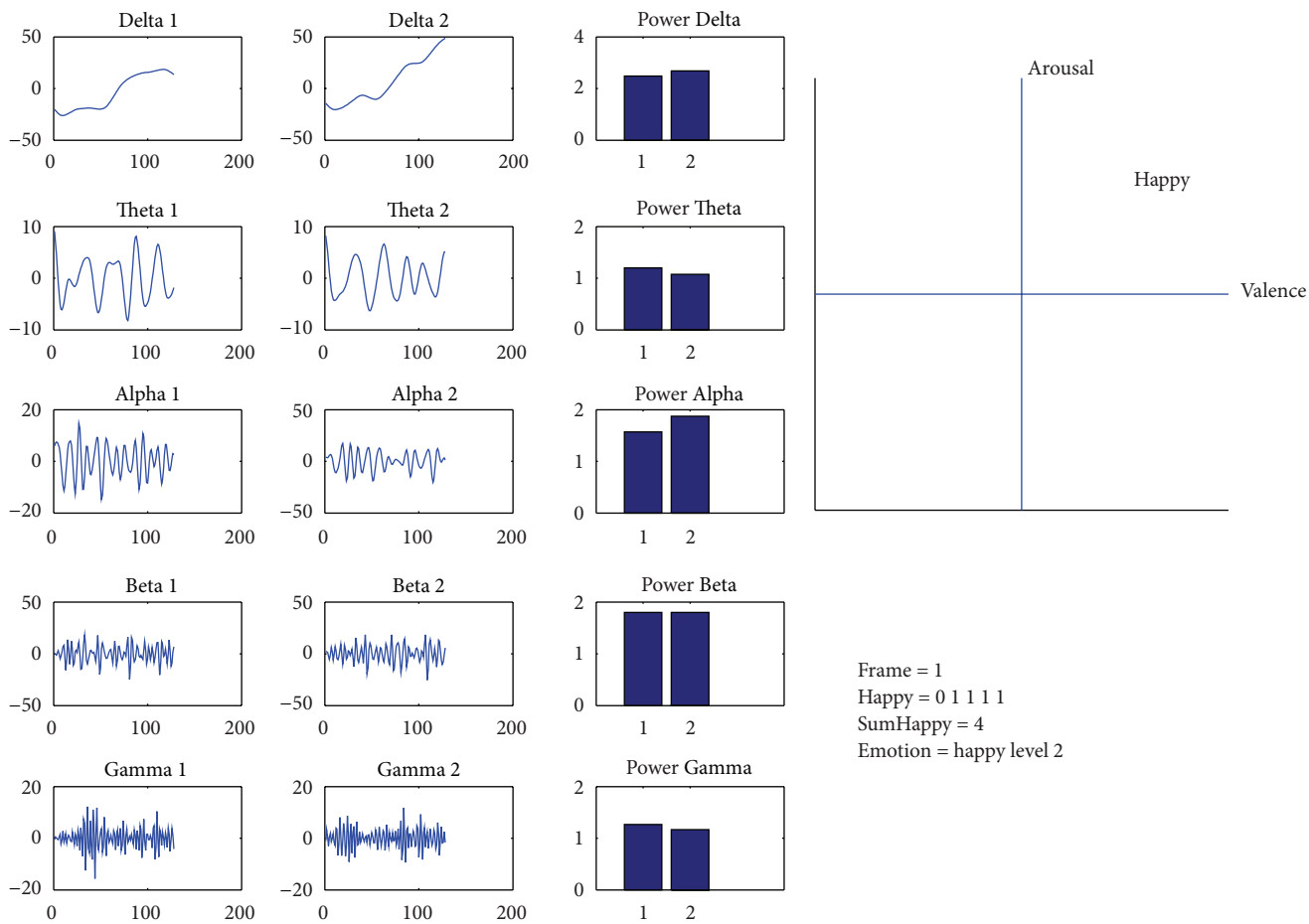
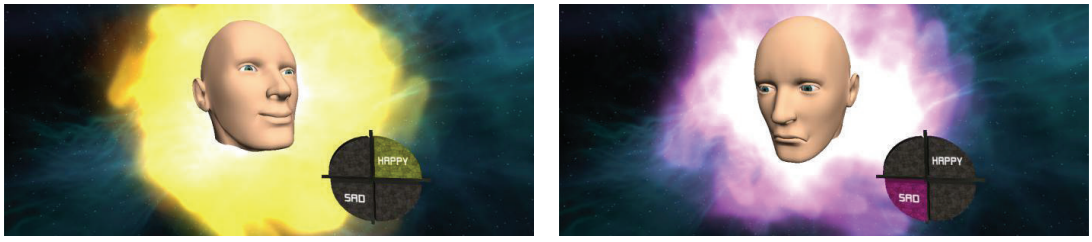


Figure 11: Screenshot of real-time happiness detection system.



(a) (b)

Figure 12: Screenshot of AVATAR game: (a) happy and (b) unhappy.

of the first 30 seconds and the last 30 seconds are 69.17% and 73.43%, respectively. When we exclude older subjects, the average accuracies of the first 30 seconds and the last 30 seconds are up to 74.67% and 75.48%, respectively. Some subjects have higher accuracy in the first 30 seconds than the last 30 seconds and some subjects have higher accuracy in the last 30 seconds than the first 30 seconds. It shows that the time duration to elicit emotion is different depending on subjects. Considering statistical significance, we found that result from the first 30 seconds does not have significant difference from the result from the last 30 seconds (P value > 0.05). Furthermore, result from the first 30 seconds does not have significant difference from the result from 60 seconds (P-value > 0.05). As a result, we may reduce time to elicit emotion from 60 to 30 seconds to save time duration for emotion elicitation.

5. Real-Time Happiness Detection System

From the results of the tests in Section 4, we implement real-time EEG-based happiness detection system using only one pair of channels. Figure 10 shows the flowchart of the happiness detection system that can be described as follows. The EEG signals with window 1 second are decomposed into

Figure 13: Screenshot of RUNNING game.

Table 3: Level of happiness.

Happy	Unhappy	Emotion
0	5	Unhappy level 3
1	4	Unhappy level 2
2	3	Unhappy level 1
3	2	Happy level 1
4	1	Happy level 2
5	0	Happy level 3

5 frequency bands (i.e., Delta, Theta, Alpha, Beta, and Gamma) by Wavelet Transform. Then we compute PSD of each band as features. With 2 channels, there are 10 features. After that, each feature is normalized by scaling between 0 and 1. Then the normalized features are inserted to classification model, built from previous experiment, to classify emotion. The selected appropriate parameters are derived from LOTO-CV method from previous experiment. The system detects the happy emotion every 5 seconds. Since emotion is classified every second, there are 5 classifications. Majority vote among classifications is used for system detection output. If the number of classifications during consecutive 5 seconds is happy more than unhappy, the detected emotion is happy. Otherwise, the detected emotion is unhappy. We divide the level of emotion from happy to unhappy depending on the number of happy classifications as shown in Table 3. The real-time happiness detection system is implemented using BCI2000 [49] and Matlab as shown in Figure 11. It is run on ASUS K45A with Intel Core i3-3110 M (2.4 GHz, 3 MB L3 Cache).

Furthermore, we develop games for recognizing and controlling happiness that consist of AVATAR and RUNNING. Both games are implemented using UINITY3D based on the real-time happiness detection system that was presented.

AVATAR. We develop AVATAR game to demonstrate real-time facial expression depending on user's emotion. When the user is happy, the program shows happy face with happy music. Conversely, when the user is unhappy, the program shows unhappy face with unhappy music as shown in Figure 12. This is the game that can help user recognize the happiness.

RUNNING. We develop RUNNING game. The aim of this game is to control the character to run as far as possible within time constraint as shown in Figure 13. The speed of character depends on how happy the user is at the moment. The happier the user is, the more speed the character has. The speed is divided into 6 levels depending on the level of happiness. If the user can sustain their happiness, the character can cover long distance. This is the game that can help user control the happiness.

6. Conclusions and Future Work

In this research we propose to use real-time EEG signal to classify happy and unhappy emotions elicited by pictures and classical music. Considering each pair of channels and different frequency bands, temporal pair of channels gives a better result than the other area does, and high frequency bands give a better result than low frequency bands do. All of these are beneficial to the development of emotion classification system using minimal EEG channels in real time. From these results, we implement real-time happiness detection system using only one pair of channels. Furthermore, we develop games to help users recognize and control the happy emotion to be what they want. In the future, we will use other physiological signals such as Galvanic Skin Response (GSR), Electrocardiogram (ECG), and Skin Temperature (ST) combined with EEG to enhance the performance of emotion recognition in the aspect of accuracy and number of emotions.

Acknowledgments

The authors would like to thank all participants for the valuable time to be a part of this research, and they would also like to thank GAPED for the effective pictures. This research has been granted by the National Science and Technology Development Agency University-Industry Research Collaboration (NUI-RC).

References

[1] A. Luneski, E. Konstantinidis, and P. D. Bamidis, "Affective medicine: a review of affective computing efforts in medical informatics," *Methods of Information in Medicine*, vol. 49, no. 3, pp. 207–218, 2010.

[2] J. Helliwell, R. layard, and J. Sachs, "World Happiness Report," http://www.earth.columbia.edu/sitefiles/file/Sachs%20Writing/2012/World%20Happiness%20Report.pdf.

[3] S. Lyubomirsky, L. King, and E. Diener, "The benefits of frequent positive affect: does happiness lead to success?" *Psychological Bulletin*, vol. 131, no. 6, pp. 803–855, 2005.

[4] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," *International Journal of Synthetic Emotions*, vol. 1, pp. 68–99, 2010.

[5] J. W. Papez, "A proposed mechanism of emotion," *Archives of Neurology and Psychiatry*, vol. 38, no. 4, pp. 725–743, 1937.

[6] P. D. MacLean, "Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (Visceral brain)," *Electroencephalography and Clinical Neurophysiology*, vol. 4, no. 4, pp. 407–418, 1952.

[7] F. Sharbrough, G. E. Chatrian, R. P. Lesser, H. Luders, M. Nuwer, and T. W. Picton, "American electroencephalographic society guidelines for standard electrode position nomenclature," *Journal of Clinical Neurophysiology*, vol. 8, no. 2, pp. 200–202, 1991.

[8] E. Niedermeyer and F. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 2004.

[9] Wikipedia, "Electroencephalography," http://en.wikipedia.org/wiki/Electroencephalography.

[10] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: arousal evaluation using EEG's and peripheral physiological signals," in *Multimedia Content Representation, Classification and Security*, B. Gunsel, A. Jain, A. M. Tekalp, and B. Sankur, Eds., vol. 4105, pp. 530–537, Springer, Berlin, Germany, 2006.

[11] Y. P. Lin, C. H. Wang, T. L. Wu, S. K. Jeng, and J. H. Chen, "Support vector machine for EEG signal classification during listening to emotional music," in *Proceedings of the 10th IEEE Workshop on Multimedia Signal Processing (MMSP '08)*, pp. 127–130, Cairns, Australia, October 2008.

[12] P. Ekman and W. Friesen, "Measuring facial movement with the facial action coding system," in *Emotion in the Human Face*, Cambridge University Press, New York, NY, USA, 2nd edition, 1982.

[13] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[14] R. Horlings, *Emotion Recognition Using Brain Activity*, Department of Mediamatics, Delft University of Technology, 2008.

[15] M. M. Bradley, P. J. Lang, and B. N. Cuthbert, *International Affective Picture System (IAPS): Digitized Photographs, Instruction Manual and Affective Ratings*, University of Florida, Gainesville, Fla, USA, 2005.

[16] M. M. Bradley and P. J. Lang, *The International Affective Digitized Sounds (IADS-2): Affective Ratings of Sounds and Instruction Manual*, University of Florida, Gainesville, Fla, USA, 2nd edition, 2007.

[17] E. S. Dan-Glauser and K. R. Scherer, "The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance," *Behavior Research Methods*, vol. 43, no. 2, pp. 468–477, 2011.

[18] T. Baumgartner, M. Esslen, and L. Jäncke, "From emotion perception to emotion experience: emotions evoked by pictures and classical music," *International Journal of Psychophysiology*, vol. 60, no. 1, pp. 34–43, 2006.

[19] D. Bos, "EEG-based emotion recognition," http://hmi.ewi.utwente.nl/verslagen/capita-selecta/CS-Oude_Bos-Danny.pdf.

[20] M. Li and B. L. Lu, "Emotion classification based on gamma-band EEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '09)*, pp. 1223–1226, Minneapolis, Minn, USA, September 2009.

[21] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *International Journal of Human Computer Studies*, vol. 67, no. 8, pp. 607–627, 2009.

[22] Z. Khalili and M. H. Moradi, "Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '09)*, pp. 1571–1575, Atlanta, Ga, USA, June 2009.

[23] O. AlZoubi, R. A. Calvo, and R. H. Stevens, "Classification of EEG for affect recognition: an adaptive approach," in *AI 2009: Advances in Artificial Intelligence*, A. Nicholson and X. Li, Eds., vol. 5866 of *Lecture Notes in Computer Science*, pp. 52–61, Springer , Berlin, Germany, 2009.

[24] Y. P. Lin, C. H. Wang, T. P. Jung et al., "EEG-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[25] S. Koelstra, A. Yazdani, M. Soleymani et al., "Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos," in *Proceeding of the International Conference on Brain Informatics (BI '10)*, pp. 89–100, Toronto, Canada, 2010.

[26] R. Khosrowabadi, H. C. Quek, A. Wahab, and K. K. Ang, "EEG-based emotion recognition using self-organizing map for boundary detection," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 4242–4245, Istanbul, Turkey, August 2010.

[27] M. Murugappan, R. Nagarajan, and S. Yaacob, "Combining spatial filtering and wavelet transform for classifying human emotions using EEG Signals," *Journal of Medical and Biological Engineering*, vol. 31, no. 1, pp. 45–51, 2011.

[28] S. A. Hosseini, M. A. Khalilzadeh, M. B. Naghibi-Sistani, and V. Niazmand, "Higher order spectra analysis of EEG signals in emotional stress states," in *Proceedings of the 2nd International Conference on Information Technology and Computer Science (ITCS '10)*, pp. 60–63, ukr, July 2010.

[29] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-time EEG-based human emotion recognition and visualization," in *Proceedings of the International Conference on Cyberworlds (CW '10)*, pp. 262–269, Singapore, October 2010.

[30] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[31] D. Nie, X. W. Wang, L. C. Shi, and B. L. Lu, "EEG-based emotion recognition during watching movies," in *Proceedings of the 5th International IEEE/EMBS Conference on Neural Engineering (NER '11)*, pp. 667–670, Cancun, Mexico, May 2011.

[32] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 41, no. 6, pp. 1052–1063, 2011.

[33] X. W. Wang, D. Nie, and B. L. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in *Neural Information Processing*, B. L. Lu, L. Zhang, and J. Kwok, Eds., vol. 7062 of *Lecture Notes in Computer Science*, pp. 734–743, Springer, Berlin, Germany, 2011.

[34] L. Brown, B. Grundlehner, and J. Penders, "Towards wireless emotional valence detection from EEG," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '11)*, pp. 2188–2191, Boston, Mass, USA, September 2011.

[35] S. Koelstra, C. Mühl, M. Soleymani et al., "DEAP: a database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[36] V. H. Anh, M. N. Van, B. B. Ha, and T. H. Quyet, "A real-time model based support vector machine for emotion recognition through EEG," in *Proceedings of the International Conference on Control, Automation and Information Sciences (ICCAIS '12)*, pp. 191–196, Ho Chi Minh City, Vietnam, November 2012.

[37] H. Xu and K. N. Plataniotis, "Affect recognition using EEG signal," in *Proceedings of the 14th IEEE International Workshop on Multimedia Signal Processing (MMSP '12)*, pp. 299–304, Banff, Canada, September 2012.

[38] D. Huang, C. Guan, K. K. Ang, H. Zhang, and Y. Pan, "Asymmetric spatial pattern for EEG-based emotion detection," in

Proceedings of the International Joint Conference on Neural Networks (IJCNN '12), pp. 1–7, Brisbane, Australia, June 2012.

[39] T. F. Bastos-Filho, A. Ferreira, A. C. Atencio, S. Arjunan, and D. Kumar, "Evaluation of feature extraction techniques in emotional state recognition," in *Proceedings of the 4th International Conference on Intelligent Human Computer Interaction (IHCI '12)*, pp. 1–6, Kharagpur, India, December 2012.

[40] U. Wijeratne and U. Perera, "Intelligent emotion recognition system using electroencephalography and active shape models," in *Proceedings of the IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES '12)*, pp. 636–641, Langkawi, Malaysia, December 2012.

[41] N. Jatupaiboon, S. Pan-ngum, and P. Israsena, "Emotion classification using minimal EEG channels and frequency bands," in *Proceedings of the 10th International Joint Conference on Computer Science and Software Engineering (JCSSE '13)*, pp. 21–24, 2013.

[42] R. W. Levenson, "Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity," in *Social Psychophysiology and Emotion: Theory and Clinical Applications*, H. L. Wagner, Ed., pp. 17–42, John Wiley & Sons, New York, NY, USA, 1988.

[43] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 4, no. 2, pp. R1–R13, 2007.

[44] N. N. Vempala and F. A. Russo, "Predicting emotion from music audio features using neural networks," in *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR '12)*, 2012.

[45] "Emotiv EEG Neuroheadset," http://emotiv.com/upload/manual/EEGSpecifications.pdf.

[46] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article 27, 2011.

[47] R. W. Levenson, L. L. Carstensen, W. V. Friesen, and P. Ekman, "Emotion, physiology, and expression in old age," *Psychology and Aging*, vol. 6, no. 1, pp. 28–35, 1991.

[48] J. A. Onton and S. Makeig, "High-frequency broadband modulations of electroencephalographic spectra," *Frontiers in Human Neuroscience*, vol. 3, article 61, 2009.

[49] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.