Emotion recognition system using short-term monitoring of physiological signals

K. H. Kim[1] **S. W. Bang**[2] **S. R. Kim**[2]

[1]Department of Biomedical Engineering, College of Health Science, Yonsei University, South Korea
[2]Human–computer Interaction Laboratory, Samsung Advanced Institute of Technology, South Korea

Abstract—*A physiological signal-based emotion recognition system is reported. The system was developed to operate as a user-independent system, based on physiological signal databases obtained from multiple subjects. The input signals were electrocardiogram, skin temperature variation and electrodermal activity, all of which were acquired without much discomfort from the body surface, and can reflect the influence of emotion on the autonomic nervous system. The system consisted of preprocessing, feature extraction and pattern classification stages. Preprocessing and feature extraction methods were devised so that emotion-specific characteristics could be extracted from short-segment signals. Although the features were carefully extracted, their distribution formed a classification problem, with large overlap among clusters and large variance within clusters. A support vector machine was adopted as a pattern classifier to resolve this difficulty. Correct-classification ratios for 50 subjects were 78.4% and 61.8%, for the recognition of three and four categories, respectively.*

Keywords—*Emotion recognition, Autonomic nervous system, Physiological signal processing, Support vector machine*

Med. Biol. Eng. Comput., 2004, **42**, 419–427

1 Introduction

RESEARCH EFFORTS in human–computer interaction are focused on the means to empower computers (robots and other machines) to understand human intention, e.g. speech recognition and gesture recognition systems (BING-HWANG and FURUI, 2000; COWIE *et al.*, 2001; TEFAS *et al.*, 2001). In spite of considerable achievements in this area during the past several decades, there are still a lot of problems, and many researchers are trying to solve them. Besides, there is another important but ignored mode of communication that may be important for more natural interaction: emotion plays an important role in contextual understanding of messages from others in speech or visual forms.

There are numerous areas in human–computer interaction that could effectively use the capability to understand emotion (COWIE *et al.*, 2001). For example, it is accepted that emotional ability is an essential factor for the next-generation personal robot, such as the Sony AIBO (ARKIN *et al.*, 2001). It can also play a significant role in 'intelligent room' (HIRSH *et al.*, 1999) and 'affective computer tutor' (PICARD, 1995).

Although limited in number compared with the efforts being made towards intention-translation means, some researchers are trying to realise man–machine interfaces with an emotion-understanding capability. Most of them are focused on facial expression recognition and speech signal analysis (COWIE *et al.*, 2001). Another possible approach for emotion recognition is physiological signal analysis. We believe that this is a more natural means of emotion recognition, in that the influence of emotion on facial expression or speech can be suppressed relatively easily, and emotional status is inherently reflected in the activity of the nervous system.

In the field of psychophysiology, traditional tools for the investigation of human emotional status are based on the recording and statistical analysis of physiological signals from both the central and autonomic nervous systems (ANDREASSI, 2000; BOUCSEIN, 1992). Researchers at IBM recently reported an emotion recognition device based on mouse-type hardware (ARK *et al.*, 1999). Picard and colleagues at the MIT Media Laboratory have been exerting their efforts to implement an 'affective computer' since the late 1990s (PICARD, 1995; PICARD *et al.*, 2001; PICARD and HEALEY, 1998; FERNANDEZ and PICARD, 1998). Although they demonstrated the feasibility of a physiological signal-based emotion recognition system, several aspects of its performance need to be improved before it can be utilised as a practical system.

First, their algorithm development and performance tests were carried out with data that reflect intentionally expressed emotion. Moreover, their data were acquired from only one subject, and, hence, their emotion recognition algorithm is user-dependent and must be tuned to a specific person. It seems natural to start from the development of a user-dependent system, as the speech

Correspondence should be addressed to Professor Kyung Hwan Kim; email: khkim@ieee.org

Paper received 3 June 2003 and in final form 17 February 2004

MBEC online number: 20043886

recognition system began with a speaker-dependent system. Nevertheless, a user-independent system is essential for practical application, so that the users do not have to be bothered with training of the system. To our knowledge, there is no previous study that has demonstrated a physiological signal-based emotion recognition system that is applicable to multiple users. Another problem with current systems is the required length of signals. At present, at least 2–5 min of signal monitoring is required for a decision (PICARD *et al.*, 2001; PICARD and HEALEY, 1998; ARK *et al.*, 1999). For practical purposes, the required monitoring time should be reduced further.

In this paper, a novel emotion recognition system based on the processing of physiological signals is presented. This system shows a recognition ratio much higher than chance probability, when applied to physiological signal databases obtained from tens to hundreds of subjects. The system consists of characteristic waveform detection, feature extraction and pattern classification stages. Although the waveform detection and feature extraction stages were designed carefully, there was a large amount of within-class variation of features and overlap among classes. This problem could not be solved by simple classifiers, such as linear and quadratic classifiers, that were adopted for previous studies with similar purposes.

We utilised a support vector machine, along with parameter determination using cross-validation, to overcome this difficulty in pattern classification. Unlike other pattern recognition problems, such as speech or character recognition, uncertainty in the class labels of feature vectors of training and test data (i.e. the ground truth on emotional status) is substantial. For example, it is quite certain that the speech waveform under investigation corresponds to '*dog*' *or* '*cat*'; however, in our case, it is impossible correctly to judge whether current multiple physiological signal waveforms represent the status of '*happiness*' *or* '*sadness*'. This absence of the 'ground truth' renders the implementation of physiological signal-based emotion recognition very difficult. Considering this and the fact that there exist many other uncontrollable sources affecting the physiological signals, the recognition ratio achieved seems to be encouraging.

2 Methods

2.1 *Subjects*

Originally, our target subjects were children aged from five to eight years. Algorithm development and performance testing of the overall system were carried out for the physiological signal databases constructed from two groups of subjects. The databases were acquired with the procedure described below in Section 2.2. The first group of subjects included 125 subjects who were from five to eight years old. Similar experiments were performed to construct another database 1 year after the construction of the first database, and 50 subjects, aged from seven to eight years old participated, as preschool children (aged five–six years) showed difficulty in inducing emotions and reporting them.

For the second database, special attention was paid so that severe contamination of the signals by motion artifact was not present, and there was no discontinuity of signal recording between the baseline measurement and the signal under stimulus. All the subjects were normal children without any history of medical, neurological or psychiatric illness. All the experiments were carriedout after consent had been obtained from the subjects, their teacher and parents.

From the database of the first group, data from half the subjects (randomly chosen) were used for the training, and data from the rest were used for the test. No special preprocessing by visual inspection was performed to eliminate severely contaminated segments from the raw signals. For the database of the second group, data from the 33 randomly chosen subjects (approximately two-thirds of all subjects) were used for the training, and data from the rest (17 subjects) were used for the test.

2.2 *Physiological studies*

2.2.1 *Selection of input signals for the emotion recognition system:* Acquisition of a high-quality database of physiological signals is vital for the emotion recognition algorithm development. An important concern is the selection of signals that are to be used as input to the emotion recognition system. It is desirable that the influence of emotion on the activity of the nervous system is effectively reflected in the physiological signals employed. Unlike the case of speech recognition or facial expression recognition, where knowledge of the correct class label of a given data point is self-evident, the acquisition of a high-quality physiological signal database with confidence in the underlying emotional status is an intricate task. It is not at all easy to judge whether the targeted emotional status is properly induced. Even if it is properly induced, the variation in physiological responses among individuals is expected to be enormous. Moreover, it is generally hard to determine whether the phenomenological changes in the physiological signals are from emotional status change or other factors, such as cognition, thought and sensory stimuli.

Because we were determined to develop a practical algorithm, there is a limitation on the range of usable signals. Although electro-encephalogram, respiration and facial electromyograms would be expected to be helpful, the attachment of electrodes to the scalp or face seems not to be tolerable for practical use, and thus we decided not to employ them. In this study, the selected input signals are skin temperature variation, electrodermal activity and heart rate, which can be derived from electrocardiogram (ECG) or photoplethysmogram. We expect that the sensors can eventually be implemented as a ring-type or wristwatch-type sensor module that can be worn for 24 h without discomfort.

The input signals reflect the activity of the autonomic nervous system. The autonomic nervous system plays a major role in maintaining the internal equilibrium of the body. It is connected to smooth muscles, the secretion glands of internal organs and cardiac muscles. The autonomic nervous system is divided into the sympathetic nervous system and the parasympathetic nervous system. These two branches of the autonomic nervous system are operated in antagonistic fashion to maintain homeostasis. It is well known that emotional stimuli can have a great effect on the activity of the autonomic nervous system (ANDREASSI, 2000; BOUCSEIN, 1992). The increase in heart rate and blood pressure and the enlargement of pupil diameter under fear stimuli are typical examples of this phenomenon. Here, we briefly explain the underlying rationale of correlation between emotion and the adopted signals.

The sino-atrial node, which acts as pacemaker of cardiovascular activity, receives inputs from both branches of the autonomic nervous system. The activity level of the sympathetic nervous system is presented to the sino-atrial node by a postganglionic fibre, and that of the parasympathetic nervous system is given by a vagal nerve. The sino-atrial node can be thought of as a spike train generator whose inter-spike interval is modulated by the integration of the activity levels of the parasympathetic and sympathetic nervous system. In other words, if we treat the heartbeat as a random point process, its rate is dependent on the activity level of the autonomic nervous system, which in turn is dependent on emotional stimuli. This information can be extracted from the change in heart rate as a function of time and thus can be extracted from the ECG. Photoplethysmography (PPG) can also be used to extract heart rate and is better suited for the simplification of the sensor module. Degradation of the signal quality of PPG due to motion artifact should be reduced before it is adopted for our purposes.

Electrodermal activity (EDA) is another signal that can easily be measured from the body surface and represents the activity of the autonomic nervous system. It is also called galvanic skin response (BOUCSEIN, 1992). It characterises changes in the electrical properties of the skin due to the activity of sweat glands and is physically interpreted as conductance. Sweat glands distributed on the skin receive input from the sympathetic nervous system only, and thus this is a good indicator of arousal level due to external sensory and cognitive stimuli. It has frequently been adopted for polygraphic lie detectors (JIANG *et al.*, 2000).

Variations in the skin temperature (SKT) mainly come from localised changes in blood flow caused by vascular resistance or arterial blood pressure. Local vascular resistance is modulated by smooth muscle tone, which is mediated by the sympathetic nervous system. The mechanism of arterial blood pressure variation can be described by a complicated model of cardiovascular regulation by the autonomic nervous system. Thus it is evident that the SKT variation reflects autonomic nervous system activity and is another effective indicator of emotional status. The variation is SKT due to emotional stimuli was studied by SHUSTERMAN and BARNEA (1995) and KATAOKA *et al.* (1998).

2.2.2 *Acquisition of emotion-specific physiological signal database:* The physiological signals were acquired using the MP100 system*. The sampling rate was fixed at 256 samples s^{-1} for all the channels. The ECG was measured from both upper arms with the two-electrode method based on lead I. PPG and SKT were measured from the little finger and the ring finger of the left hand, respectively. EDA was measured from two Ag/AgCl electrodes attached to the index and middle fingers of the right hand. Appropriate amplification and bandpass filtering were performed. One session of experiments took approximately 5 min. The first 2 min corresponded to the baseline measurement and were obtained without any emotional stimulus. The subjects were requested to be as relaxed as possible during this period. Subsequently, emotional stimulus was applied, and debriefing and recovery followed.

2.2.3 *Emotion induction protocol:* As in the cases of other pattern recognition systems, it was essential to obtain a database of physiological signals representing specific emotional statuses. To acquire a database of physiological signals in which the influence of emotional status was faithfully reflected, we developed a set of elaborate protocols for emotion induction. We concluded that visual stimulation using still images was not sufficient for effective emotion induction, and we did not use the international affective picture system (IAPS) developed by LANG *et al.* (1988), despite its being adopted for many psychophysiological studies involving emotion induction.

Our protocol utilised a multimodal (audio, visual and cognitive) approach to evoke specific targeted emotional statuses, and it was developed in collaboration with specialists from the field of cognitive and physiological psychology (YANG *et al.*, 2000). Fig. 1 illustrates one of the stimuli, the purpose of which was to induce the status of 'sadness'. It consisted of visual stimulus using controlled illumination and auditory stimulus using background music. Simultaneously, an actress narrated a sad story that was carefully prepared to evoke the sympathy of the subjects. The subjects were requested to look at a toy in front of them, so that it seemed as if the toy was telling the sad story to the subjects. The appearance of the toy was carefully chosen to induce the state of sadness effectively. These stimuli were presented in the environment of a typical living room inside an electrically shielded and soundproof chamber. We created a scenario so that the subject felt as if the toy were telling a story to them, as shown in Fig. 1. The emotion induction protocols are summarised in Table 1.

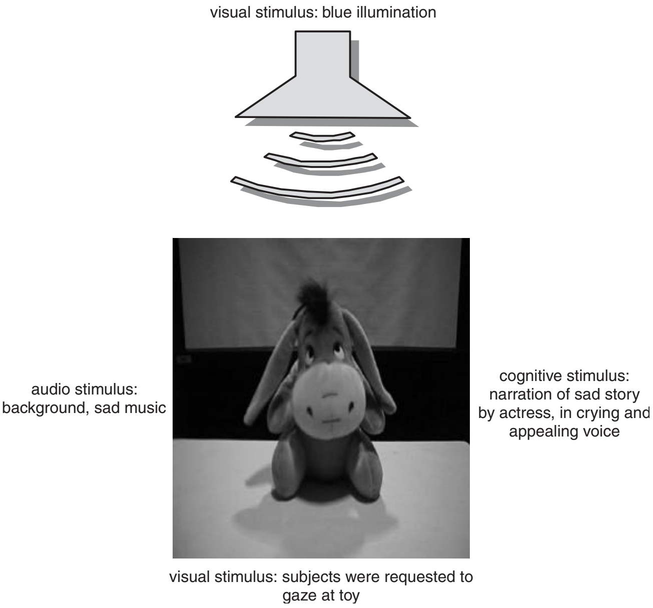
visual stimulus: blue illumination
audio stimulus: background, sad music
cognitive stimulus: narration of sad story by actress, in crying and appealing voice
visual stimulus: subjects were requested to gaze at toy

Fig. 1 *Illustration of example of emotion induction protocols. Its purpose is to induce status of 'sadness'*

A preliminary test of the protocols was performed for 80 subjects aged from seven to eight years. The test was based on the self-reports of the subjects. The self-report was obtained by a question that encouraged the subjects to state the status and strength of the emotions they felt during the applied induction protocol. They were asked to report the strength using a five-point scale. Monitoring of the activity of the subjects by

Table 1 Summary of emotion-induction protocols

Emotion	Stimulus protocol
Sadness	Story that evokes sympathy of subject told in appealing tone/crying voice; sad background music; toy with gloomy-looking appearance; blue illumination
Anger	Story that deceives subjects told in sarcastic voice; situation of feeling mortified; toy with unpleasant appearance; red illumination
Stress	Subject pressed to complete impossible mission in short time; subject compared unfavourably with other subjects; disordered environment; hard to concentrate on mission; prim looking doll; flickering illumination
Surprise	Sudden increase in volume of background music; intermittent sound of buzzer and breaking glass

*BIOPAC, Santa Barbara, CA, USA

an expert was used to provide supplementary information. The quality of the developed protocols was quantified by 'appropriateness' and 'effectiveness'. Appropriateness was defined as the percentage of subjects who reported that the given stimulus properly induced the intended emotion. Effectiveness was determined from the self-report results, where the subjects were requested to give a verbal rating of the level of strength of the emotion that he or she felt from the stimulus in five discrete ranks.

When applied to 80 subjects, the test result showed appropriateness of 85.2% and effectiveness of 82.7%. For negative emotions such as sadness or anger, the depth of stimulus was limited so as not to produce any unpleasant effects. A description of the whole experiment session and a request for active participation of the subjects were given before the stimulus presentation. After the stimulus, debriefing was performed for the recovery of the subjects' mental status.

2.3 *Preprocessing, waveform detection and feature extraction*

The first necessary step was the detection of the characteristic waveform and extraction of useful information-bearing features for pattern classification. As shown in Fig. 2, the baseline values of each component of the feature vectors were subtracted before they were given to the classifier. Here, the baseline values mean the components of feature vectors extracted from 50 s segments of signals that were acquired without stimulus.

2.3.1 *RR interval and heart rate variability:* Fig. 3 shows the block diagram of the feature extraction module for the heart rate. Heart rate variability (HRV) contains abundant information on the status of the autonomic nervous system and can be derived from ECG or PPG. Degrees of the sympathetic and parasympathetic nervous system activities can be grasped, as we described above. Time-domain features, such as mean and standard deviation (SD) of the HRV time series and its time-derivative, and a descriptor of a Poincare plot, have frequently been used as features (GARCIA-GONZALEZ and PALLAS-ARENY, 2001; WANG *et al.*, 1998). Frequency-domain features of HRV have also been considered to be significant for the exploration of the autonomic nervous system in many previous studies for cardiac function assessment and psychophysiological investigation (DRUMMOND and QUAH, 2001; MCCRATY *et al.*, 1995).

We did not employ non-linear or chaotic analyses, as they usually require long-term monitoring of signals. As our target was to extract the features that are useful for emotion recognition from short signal segments, the frequency-domain features of the HRV should be determined from short segments of signals. Although accurate spectrum estimation from a short-term signal is difficult, and special attention should be paid, in many psychophysiological studies no attention was paid to the method of spectrum estimation (FRIEDMAN and THAYER, 1998). Some efforts have been made towards accurate spectrum estimation of HRV (PINNA *et al.*, 1996) in the biomedical signal processing community, but, in our opinion, no successful guidelines are yet available, especially for the case of short signal length.

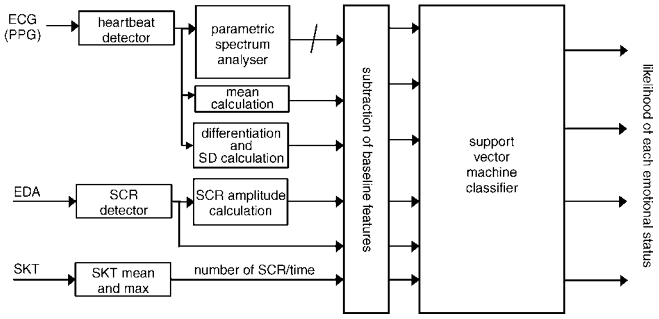
ECG
(PPG)
heartbeat detector
parametric spectrum analyser
mean calculation
differentiation and SD calculation
EDA
SCR detector
SCR amplitude calculation
SKT
SKT mean and max
number of SCR/time
subtraction of baseline features
support vector machine classifier
likelihood of each emotional status

Fig. 2 *Overall structure of emotion recognition system*

We tried to settle this problem by applying a remarkable piece of research into the model estimation of short time series, performed by BROERSEN (2000*a*; *b*). Recently, he has concluded that spectrum estimation methods based on the time-series model outperform periodogram-based methods, and he claimed that a single best time-series model, among autoregressive, moving average and autoregressive moving average models, can be selected from given samples of a short segment of signal. The selection of the best time-series model is based on an index that is determined to represent the square error of prediction. The index is derived from thorough empirical investigation of time-series behaviour. The details of the algorithm for the selection of the best time-series model, the ARMAsel algorithm, are thoroughly described in BROERSEN (2000*a*).

Fig. 4 illustrates the heartbeat detector using R-peak detection. A Teager energy operator (TEO) was used to detect the R-peak in the raw ECG signal. The output from the TEO was proportional to the product of instantaneous amplitude and frequency, and thus it was ideally suited to enhancing the R-peak in the input ECG signal (MARAGOS *et al.*, 1993). If the baseline drift was prohibitively high, a median filter was used to estimate the baseline fluctuation to generate a baseline-removed signal. After the R-peaks had been detected, the spike train could be transformed into a continuous time signal called heart rate variability (HRV) by interpolation and downsampling, as described in BERGER *et al.* (1986). From the HRV time series and its power spectrum determined by ARMAsel, frequency-domain features representing sub-band powers were extracted. Two sub-bands that are usually adopted for the spectral analysis of HRV were selected. The ranges of the low-frequency (LF) and high-frequency (HF) band were chosen as 0.03–0.15 Hz and 0.15–0.4 Hz, respectively. We did not use a very low-frequency (VLF) band, as it is difficult reliably to extract a VLF component from a short segment. Two simple time-domain features, the mean of the HRV time series and SD of its time-derivative, were also used as features. For the SD calculation, the outliers whose values belonged to the largest or smallest 10% were excluded.

2.3.2 *Electrodermal activity:* Electrodermal activity (EDA) was obtained by measurement of the voltage between two electrodes across which a low-level current was applied. Fig. 5*a* shows a typical waveform of EDA under emotional stimulation (after subtraction of mean EDA level). Important features of EDA include the DC level and the distinctive short waveforms that are indicated by arrows in Fig. 5. This is usually called the skin conductance response (SCR) and is considered to be useful as it signifies a response to internal/external stimuli. We developed a method that correctly detects the occurrence of SCR, as shown in Fig. 5*c*. After reducing the sampling rate to 20 samples s^{-1}, differentiation and subsequent convolution with a 20-point Bartlett window were performed. This procedure yielded the output waveform shown in Fig. 5*b* for the input signal shown in Fig. 5*a*.

The occurrence of the SCR was detected by finding two consecutive zero-crossings, from negative to positive and positive to negative. The amplitude of the SCR was obtained by finding the maximum value between these two zero-crossings. The mean DC level of EDA, mean values of the SCR amplitudes and duration, and number of SCR occurrences in a 50 s signal segment were extracted from the EDA as features. Detected SCRs with an amplitude smaller that 10% of the maximum

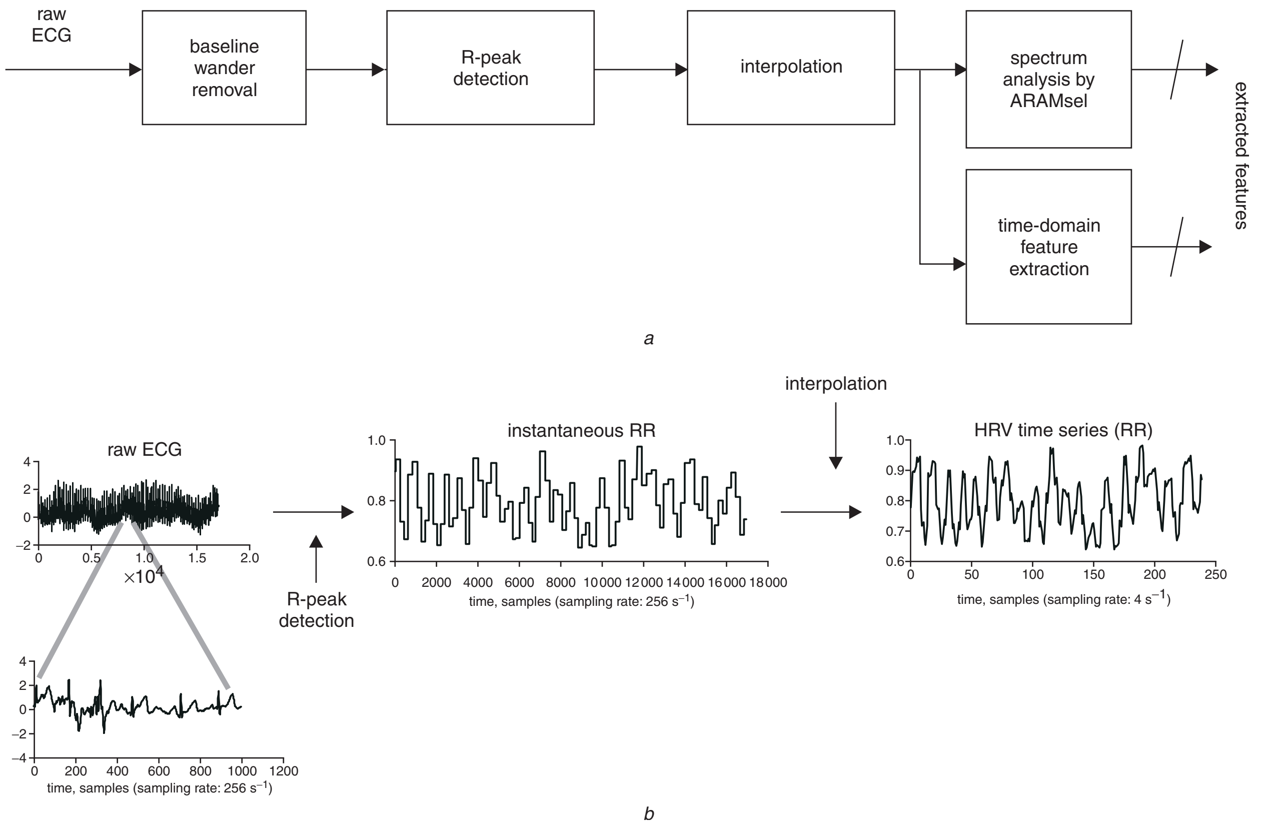
raw ECG
baseline wander removal
R-peak detection
interpolation
spectrum analysis by ARAMsel
time-domain feature extraction
extracted features
a
raw ECG
×10^4
R-peak detection
instantaneous RR
time, samples (sampling rate: 256 s^−1)
interpolation
HRV time series (RR)
time, samples (sampling rate: 4 s^−1)
time, samples (sampling rate: 256 s^−1)
b

Fig. 3 *Block diagram of information extraction module from ECG*

SCR amplitude in this segment were excluded. By this procedure, we could take into account contextual information on the level of SCR amplitude for the signal under investigation, which can vary considerably. This is advantageous compared with conventional SCR detection by visual inspection by a human supervisor, where the threshold level is determined arbitrarily, and thus objective analysis can hardly be achieved. Our method does not require explicit determination of the threshold level.

2.3.3 *Skin temperature variation:* No special signal processing was necessary for the feature extraction from the skin temperature (SKT). Although frequency-domain analysis of the time-varying SKT has been reported (SHUSTERMAN and BARNEA, 1995), here the mean and maximum values within 50 s intervals were used as the features of SKT.

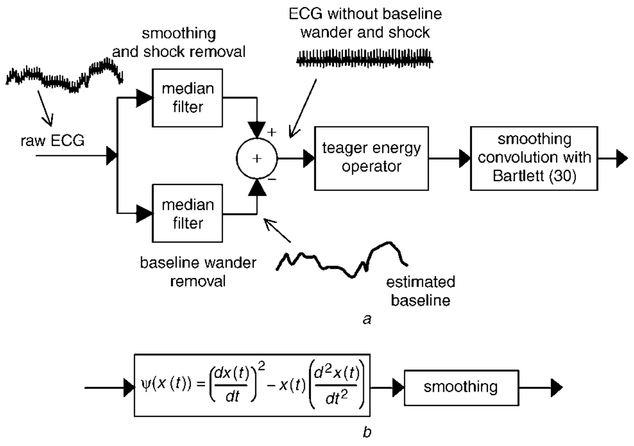
ECG without baseline wander and shock
smoothing and shock removal
median filter
raw ECG
+
−
teager energy operator
smoothing convolution with Bartlett (30)
median filter
baseline wander removal
estimated baseline
a
ψ(x(t)) = (dx(t)/dt)^2 − x(t)(d^2x(t)/dt^2)
smoothing
b

Fig. 4 *R-peak detector using TEO and baseline wander removal using median filter*

2.4 *Pattern classification using support vector machine*

Feature vectors extracted from multiple subjects under the same emotional stimulus form a distribution in high-dimensional space. As no preliminary information on the distribution of feature vectors is available, we projected them onto two-dimensional space for visualisation by a Fisher projection (DUDA *et al.*, 2001) to obtain some knowledge of the difficulty of discrimination of the feature vectors belonging to different

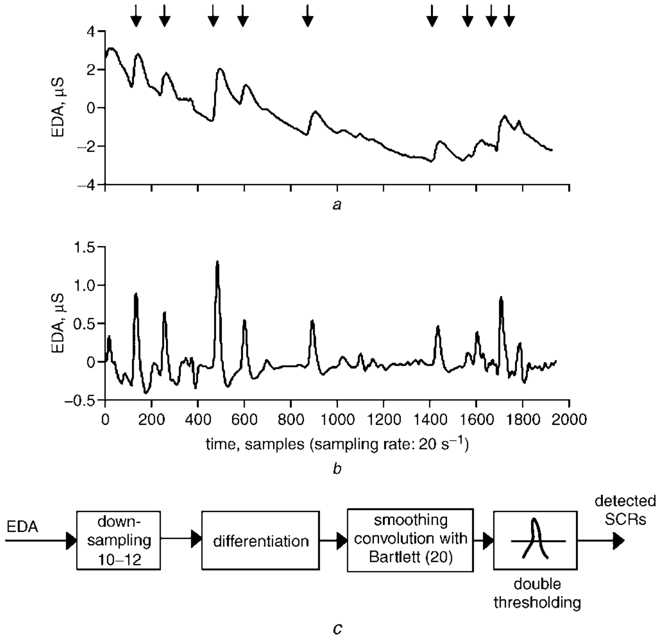

Fig. 5 *(a) Typical waveform of EDA under emotional stimulation. (b) Output signal from detection module in (a). (c) Block diagram of SCR detection module. Sampling rate of traces in (a) and (b) is 21.3 samples s^{-1}.*

emotional statuses. The projected feature vectors from the same emotional status formed a cluster with a large amount of variation, and the clusters of feature vectors from different emotional statuses significantly overlapped, as shown in Fig. 6. This is not surprising, because the feature vectors represent the state of the autonomic nervous system of multiple subjects. Inter-subject difference in emotional reaction, as well as other factors influencing the autonomic nervous system, is represented in the feature vectors.

The recently reported method of PICARD *et al.* (2001) consists of dimensionality reduction by a Fisher projection and a subsequent quadratic classifier. Their feature vectors were extracted from a single subject under the condition of deliberate expression, and it is likely that they form clusters that can be discriminated with much less effort. The combination of dimensionality reduction and a simple classifier such as the quadratic classifier was never applicable for our multi-subject problem. Furthermore, it is generally considered that quadratic classifiers show poor performance when the number of training samples is not sufficient, owing to error in the estimated covariance matrix.

Our approach utilises an efficient technique that can deal with a difficult, high-dimensional classification problem. Without dimensionality reduction, our system directly gives extracted feature vectors to the support vector machine (SVM) classifier. The SVM is based on the property that separation by a linear classifier becomes more promising after non-linear mapping onto high-dimensional space (HAYKIN, 1999) and the technique of obtaining a linear classifier with maximum generalisation performance derived from the statistical learning theory of VAPNIK (1999). Successful application of the SVM for various pattern classification problems has been recently reported (TEFAS *et al.*, 2001; CHAPELLE *et al.*, 1999). Details of the theory and learning methods of the SVM can be found in VAPNIK (1999) and BURGES (1998), and are described in the Appendix.

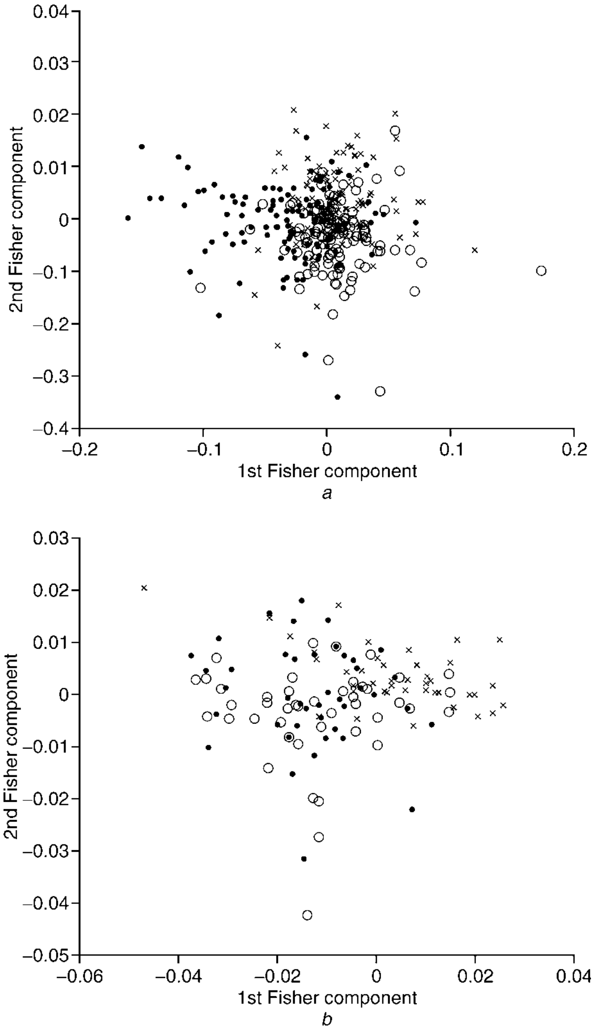

Fig. 6 *Projection of extracted feature vectors onto two-dimensional plane using Fisher projection for purpose of visualisation, for (a) first database and (b) second database.* (○) *sadness;* (×) *stress;* (•) *anger*

3 Results

First, we describe the results of each preprocessing and feature extraction module applied to relevant signals. Fig. 4*a* shows that baseline wander (low-frequency fluctuation) can be properly removed by the median filter-based method. Although identical HRV time series can be derived from both the ECG and PPG, the results shown in this Section were all obtained from the ECG.

The output from the TEO + window smoothing block is plotted in Fig. 4*b*, and it shows that the R-peaks of the ECG can be effectively detected by this procedure. From the HRV time series obtained from the interpolation and resampling after the R-peak detection, as shown in Fig. 3*b*, an optimum time-series model was obtained using the ARMAsel algorithm described above in Section 2, and sub-band powers (low frequency: 0.03–0.15 Hz; high frequency: 0.15–0.4 Hz) were calculated from the identified model. The spectrum obtained from the ARMAsel algorithm was better at increasing the separability between the stimulated and baseline statuses than the spectrum obtained from usual periodogram-based techniques, such as Welch's method. In addition, by excluding the highest and lowest 10% values of inter-beat interval (RR time), we could also enhance the separability between the stimulated and baseline statuses.

We show the result of applying the SCR detection method to a typical EDA signal in Fig. 5*b*. The most important merit of this SCR detection method is that it is not necessary to determine the threshold level. This level should be varied according to the change in EDA characteristics, dependent on the subjects, and the electrodes employed.

The feature vectors of the signals under stimulus were extracted from the last 50 s segment. Baseline feature vectors were extracted from the 50 s segments just before the stimulated segments. Figs 6*a* and *b* show the results of a Fisher projection of the feature vectors obtained from the first and second databases, respectively. These are shown for the purpose of visualisation of the difficulty of pattern classification. Three emotional categories were included. As expected, there were significant variations in the positions of data points within a single class. Overlap among different classes was also considerable. It is evident that a simple classifier such as a quadratic classifier is not sufficient for this problem. PICARD *et al.* (2001) used the combination of a Fisher projection and a quadratic classifier and reported a very high emotion recognition ratio. However, their data were acquired from a single subject, and it was assumed that emotional status was induced from voluntary endeavour of the subject, and thus it is natural that intra-cluster variation and inter-cluster overlap were much smaller compared with our situation of induced emotion in multiple subjects.

Overall performance was quantified as the percentage of correctly classified data points (correct classification ratio (CCR)). For the first year's database, the CCR was 89.7% for the training set and 55.2% for the test set, for the classification of three emotional statuses: sad, stressed and angry. For the second year's database, where the signal quality was much better, we obtained a CCR of 78.4%, for the test set in the case of the three emotional categories, and 61.8%, for the four emotional categories: sad, stressed, angry and surprised. The success rates of each emotion category for the second year's database are shown in detail in Table 2 as a contingency table.

Table 2 Contingency table showing recognition results for each emotion category for second year's database, with three and four emotional statuses. Figures denote number of subjects

	Recognition result (three emotional statuses)		
Original status	Sadness	Anger	Stress
Sadness	10	4	3
Anger	0	16	1
Stress	2	1	14

	Recognition result (four emotional statuses)			
Original status	Sadness	Anger	Stress	Surprise
Sadness	11	2	3	1
Anger	0	13	4	0
Stress	1	5	10	1
Surprise	3	4	2	8

4 Discussion

In this paper, we have shown that the realisation of user-independent emotion recognition based on physiological signals is feasible. Although our system was developed based on a biosignal database obtained from multiple subjects, the ratio of correct recognition was comparable with that of the previous systems (PICARD *et al.*, 2001; ARK *et al.*, 1999), which were developed based on data from single or few subjects. Another difference of our system compared with the previous ones is that our algorithm was developed and tested based on the biosignal database representing the induced emotional status as output of the external stimuli. It is different from previous studies where the emotion was intentionally 'tried and felt' (PICARD *et al.*, 2001) or 'acted out' (ARK *et al.*, 1999).

Our system is also better fitted for practical applications in two respects. The required signal monitoring time is significantly reduced, compared with the system of PICARD *et al.* (2001) that requires approximately 2–4 min and that of ARK *et al.* (1999) that requires 5 min. Our system does not require the attachment of electrodes (and electrode paste) to the head, face and scalp. No burdensome chest belt is necessary. We expect that, eventually, the sensor module of our system will be implemented as a wearable piece of hardware, such as a wristwatch-type device, that is perfectly suitable for everyday use. This will make an important step towards the realisation of emotional interaction between man and machine and play an important role in several applications, such as the human-friendly personal robot. For the improvement of performance and reliability, the addition of subjects to the database and further refinement and verification of emotion induction protocols may be necessary. Additional features of the physiological signals employed, for example the frequency-domain feature of SKT, may be considered to improve recognition performance (SHUSTERMAN and BARNEA, 1995).

As previously mentioned, there are numerous other factors that could affect the physiological signals, such as physical activity, cognitive workload and the physical status of subjects. In addition, although we made every effort to minimise the inclusion of these factors in our paradigms, it is probable that a considerable amount of change in physiological signals could occur owing to these confounding factors. Moreover, the basic assumption that different emotions have a more or less unique and person-independent physiological response remains questionable. This could be reflected in the fact that the recognition rate falls off with the number of emotion categories. These uncertainties could be an important cause that deteriorated the recognition ratio and troubled the model selection of the SVM.

It seems worth trying to devise a method to estimate reliably the true class label, i.e. to have knowledge of the true emotional status of a subject under consideration. Recent studies on emotion utilising advanced functional neuro-imaging techniques are encouraging (PHAN *et al.*, 2002) in this regard. If we have a means to visualise a specific brain region engaging in a specific emotional status, it will become more feasible to have 'ground truth' on the actual emotional status of the subject. At present, this is hard to obtain, and our physiological signal database is far from complete. Other techniques, such as facial muscle activity monitoring, might be helpful for this purpose, although they are not suitable for a practical emotion recognition system. A novel method for the verification of emotional status must be devised before it will be possible to say that physiological signal-based emotion recognition is a practicable and reliable way of enabling human–computer interaction with emotion-understanding capability.

5 Conclusions

We have developed a novel emotion recognition system based on the processing of physiological signals. This system shows a recognition ratio much higher than chance probability, i.e. 33.3% and 25% for three and four emotion categories, respectively, when applied to physiological signal databases obtained from tens to hundreds of subjects. The advantages of our system include the reduction of required signal monitoring time, applicability to multiple users and the use of signals that cause the minimum amount of user inconvenience. The system consists of characteristic waveform detection, feature extraction and pattern classification stages. A support vector machine was utilised as a pattern classifier to overcome the difficulty in pattern classification due to the large amount of within-class variation of features and the overlap between classes, although the features were carefully extracted. Correct classification ratios for 50 subjects were 78.43% and 61.76%, for the recognition of three and four categories, respectively.

Appendix

Pattern classification using support vector machine

Here, we briefly describe the principle of pattern classification using the SVM. A two-class classification problem is assumed for simplification. The problem of finding a linear classifier for given data points with a known class label can be described as a problem of finding a separating hyperplane $\boldsymbol{w}^T\boldsymbol{x} + b$ that satisfies

$$y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geqslant 1, \text{ for } i = 1, 2, \ldots, N \tag{1}$$

where $\boldsymbol{x}_i$ and $y_i \in \{+1, -1\}$ denote a feature vector and its given correct class label, respectively. If it is not possible to classify them with a linear classifier, as is the case with most practical problems, the problem can be described in a less strict form as follows:

$$y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geqslant 1 - \xi_i, \text{ for } i = 1, 2, \ldots, N \tag{2}$$

Here, ξ_i is called a slack variable, and it represents deviation from the ideal condition of linear separability. We can pose a problem of finding the optimum one among the separating hyperplanes by minimisation of the cost function $(1/2)\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{N}\xi_i$ subject to the constraints

$$\begin{aligned} &y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geqslant 1 - \xi_i, \text{ for } i = 1, 2, \ldots, N \\ &\xi_i \geqslant 0 \text{ for } i = 1, 2, \ldots, N \end{aligned} \tag{3}$$

The above cost function is defined so that its minimisation coincides with the maximisation of margin and minimisation of classification error under the constraint of (3).

This constrained optimisation problem can be solved by using the Lagrange multiplier method (Duda *et al.*, 2001). From the theory of the Lagrange multiplier method, it can be shown that the above problem can be expressed as a problem of finding Lagrange multipliers α_is as follows: given the training set $\{(\boldsymbol{x}_i, y_i), i = 1, 2, \ldots, N\}$, find α_is that maximise the objective function $Q(\alpha_1, \alpha_2, \ldots, \alpha_N)$, i.e.

$$\begin{aligned} \text{maximise} \quad & Q(\alpha_1, \alpha_2, \ldots, \alpha_N) \\ & = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{i=1}^{N} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_i \\ \text{subject to} \quad & \sum_{i=1}^{N} \alpha_i y_i \geqslant 0 \text{ and } 0 \leqslant \alpha_i \leqslant C \\ \text{for} \quad & i{=}1, 2, \ldots, N \end{aligned} \tag{4}$$

This is called the dual problem of the original problem that seeks to find the optimum separating hyperplane. After finding α_is, the classification of a data point $\boldsymbol{x}_{new}$ is performed as

$$f(\boldsymbol{x}_{new}) = sign\left(\sum_{i=1}^{L} y_i \alpha_i \boldsymbol{x}_{new}^T \boldsymbol{x}_i + b \right) \tag{5}$$

Here L is the number of support vectors obtained from the maximisation. The relationship between the parameter of hyperplane $\boldsymbol{w}$ and the Lagrange multiplier α_i is given as $\boldsymbol{w} = \sum_{i=1}^{L} y_i \alpha_i \boldsymbol{x}_i$. Actually, the SVM algorithm adopts a preliminary non-linear mapping to higher-dimensional feature space before the linear discrimination. The feature space is hidden from both input and output. The rationale for the non-linear mapping is taken from the Cover theorem (Haykin, 1999). It states that a non-linear mapping to high dimension increases the likelihood of linear separation (Burges, 1998). The decision function is now expressed as follows, instead of (5):

$$f(\boldsymbol{x}_{new}) = sign\left(\sum_{i=1}^{L} y_i \alpha_i \{\boldsymbol{\varphi}(\boldsymbol{x}_{new})\}^T \cdot \boldsymbol{\varphi}(\boldsymbol{x}_i) + b \right) \tag{6}$$

Here $\boldsymbol{\varphi}(\boldsymbol{x})$ denotes the non-linear mapping to high dimension. Obviously, the objective function should also be changed by substituting $\{\boldsymbol{\varphi}(\boldsymbol{x}_i)\}^T \cdot \boldsymbol{\varphi}(\boldsymbol{x}_i)$ for $\boldsymbol{x}_i^T \boldsymbol{x}_i$ in (4). Previous expressions involve computation of an inner product in high-dimensional space. In practice, an inner-product kernel $K(\boldsymbol{x}_{new}, \boldsymbol{x}_i) = \boldsymbol{\varphi}(\boldsymbol{x}_{new}) \cdot \boldsymbol{\varphi}(\boldsymbol{x}_i)$ is used instead of direct calculation of the inner product in high dimension. Not every mapping $\boldsymbol{\varphi}(\boldsymbol{x})$ can be expressed in this fashion, and the criteria are stated by Mercer's theorem (Burges, 1998). Here, we used a Gaussian kernel. Finally, the decision rule of a given data point is

$$f(\boldsymbol{x}_{new}) = sign\left(\sum_{i=1}^{L} y_i \alpha_i K(\boldsymbol{x}_{new}, \boldsymbol{x}_i) + b \right) \tag{7}$$

For the general multiclass classification problem where the number of classes is larger than 2, 'one-against-one' and 'one-against-all' approaches can be used. The former method uses $k/(k-1)/2$ classifiers, each of which is trained to separate two different classes of a total of k classes, and the best class label for a specific input vector is determined from voting. It is generally accepted that the 'one-against-one approach' gives a better result (Knerr *et al.*, 1990). Five-fold cross-validation (Duda *et al.*, 2001) is used to determine the final parameters of the classifier.

References

Andreassi, J. L. (2000): *'Psychophysiology: human behavior and physiological response'* (Lawrence Erlbaum Associates, New Jersey, 2000)

Ark, W., Dryer, D. C., and Lu, D. J. (1999): 'The emotion mouse'. 8th Int. Conf. Human-computer Interaction, pp. 818–823

Arkin, R. C., Fujita, M., Takagi, T., and Hasegawa, R. (2001): 'Ethological modeling and architecture for an entertainment robot'. IEEE Int. Conf. Robotics & Automation, pp. 453–458

Berger, R. D., Akselrod, S., Gordon, D., and Cohen, R. J. (1986): 'An efficient algorithm for spectral analysis of heart rate variability', *IEEE Trans. Biomed. Eng.*, **33**, pp. 900–904

Bing-Hwang, J., and Furui, S. (2000): 'Automatic recognition and understanding of spoken language—a first step toward natural human-machine communication', *Proc. IEEE*, **88**, pp. 1142–1165

Boucsein, W. (1992): *'Electrodermal activity'* (Plenum Press, New York, 1992)

Broersen, P. M. T. (2000*a*): 'Facts and fiction in spectral analysis', *IEEE Trans. Instrum. Meas.*, **49**, pp. 766–772

Broersen, P. M. T. (2000*b*): 'Finite sample criteria for autoregressive order selection', *IEEE Trans. Signal Process.*, **48**, pp. 3550–3558

Burges, C. J. C. (1998): 'A tutorial on support vector machines for pattern recognition', *Knowledge Discov. Data Mining*, **2**, pp. 1–43

Chapelle, O., Haffner, P., and Vapnik, V. N. (1999): 'Support vector machines for histogram-based image classification', *IEEE Trans. Neural Netw.*, **10**, pp. 1055–1064

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001): 'Emotion recognition in human-computer interaction', *IEEE Signal Process. Mag.*, **18**, pp. 32–80

Drummond, P. D., and Quah, S. H. (2001): 'The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians', *Psychophysiology*, **38**, pp. 190–196

Duda, R. O., Hart, P. E., and Stork, D. G. (2001): *'Pattern classification'*, 2nd edn (Wiley, New York, 2001)

Fernandez, R., and Picard, R. W. (1998): 'Signal processing for recognition of human frustration'. IEEE Int. Conf. Acoustic Speech. Signal Processing, pp. 3773–3776

Friedman, B. H., and Thayer, J. F. (1998): 'Autonomic balance revisited: panic anxiety and heart rate variability', *J. Psychosom. Res.*, **44**, pp. 133–151

Garcia-Gonzalez, M. A., and Pallas-Areny, R. A. (2001): 'Novel robust index to assess beat-to-beat variability in heart rate time-series analysis', *IEEE Trans. Biomed. Eng.*, **48**, pp. 617–621

Haykin, S. (1999): *'Neural networks'*, 2nd edn (Prentice Hall, New Jersey, 1999)

Hirsh, H., Coen, M. H., Mozer, M. C., Hasha, R., and Flanagan, J. L. (1999): 'Room service, AI-style', *IEEE Intell. Syst.*, **14**, pp. 8–19

Jiang, L., Qing, Z., and Wenyuan, W. (2000): 'A novel approach to analyze the result of polygraph'. IEEE Int. Conf. Systens Man Cybernetics, pp. 2884–2886

Kataoka, H., Kano, H., Yoshida, H., Saijo, A., Yasuda, M., and Osumi, M. (1998): 'Development of a skin temperature measuring system for non-contact stress evaluation'. IEEE Ann. Conf. Engineering Medicine Biology Society, pp. 940–943

Knerr, S., Personnaz, L., and Dreyfus, G. (1990): 'Single-layer learning revisited: a stepwise procedure for building and training a neural network' in Fogelman, J. (Ed.): *'Neurocomputing: algorithms, archtectures and applications'* (Springer-Verlag, Heidelberg, 1990)

Lang, P. J., Ohman, A., and Vaitl, D. (1988): 'The international affective picture system (photographic slides)'. Center for Research in Psychophysiology, University of Florida, Gainesville, USA

Maragos, P., Kaiser, J. F., and Quatieri, T. F. (1993): 'On amplitude and frequency demodulation using energy operators', *IEEE Trans. Signal Process*, **41**, pp. 1532–1550

McCraty, R., Atkinson, M., Tiller, W. A., Rein, G., and Watkins, A. D. (1995): 'The effects of emotions on short-term power spectrum analysis of heart rate variability', *Am. J. Cardiol.*, **76**, pp. 1089–1093

Phan, K. L., Wagner, T., Taylor, S. F., and Liberzon, I. (2002): 'Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI', *Neuroimage*, **16**, pp. 331–348

PICARD, R. W. (1995): *'Affective computing'* (MIT Press, Cambridge, 1995)

PICARD, R. W., and HEALEY, J. (1998): 'Digital processing of affective signals', *IEEE Int. Conf. Acoustic Speech Signal Processing*, pp. 3749–3752

PICARD, R. W., VYZAS, E., and HEALEY, J. (2001): 'Toward machine emotional intelligence: analysis of affective physiological state', *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**, pp. 1175–1191

PINNA, G. D., MAESTRI, R., and SANARICO, M. (1996): 'Effects of record length selection on the accuracy of spectral estimates of heart rate variability', *IEEE Trans. Biomed. Eng.*, **43**, pp. 754–757

SHUSTERMAN, V., and BARNEA, O. (1995): 'Analysis of skin-temperature variability compared to variability of blood pressure and heart rate', *IEEE Ann. Conf. Engineering Medicine Biology Society*, pp. 1027–1028

TEFAS, A., KOTROPOULOS, C., and PITAS, I. (2001): 'Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication', *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**, pp. 735–746

VAPNIK, V. N. (1999): 'An overview of statistical learning theory', *IEEE Trans. Neural Netw.*, **10**, pp. 988–999

WANG, F., SAGAWA, K., and INOOKA, H. (1998): 'Time domain heart rate variability index for assessment of dynamic stress', *Comput. Cardiol.*, pp. 97–100

YANG, G., LEE, K., LEE, J., CHOI, J., BANG, S., KIM, J., LEE, H., and SOHN, J. (2000): 'Development of emotion induction protocol for children', *Ann. Conf. Society for Emotion and Sensibility of Korea*, pp. 20–25

Author's biography

KYUNG HWAN KIM received the PhD from the School of Electrical & Computer Engineering, Seoul National University, in February 2001. From March 2001 to February 2004, he was a member of the research staff in the Human–Computer Interaction Laboratory, Samsung Advanced Institute of Technology, Korea. From March 2003 to December 2003, he was a Visiting Scholar at the Functional Magnetic Resonance Imaging Laboratory, Korea Advanced Institute of Science & Technology. He is now with the Department of Biomedical Engineering, Yonsei University, Korea, as an Assistant Professor. His research interests include biomedical signal processing, pattern recognition, instrumentation and neuro-imaging.