Emotion recognition based on physiological changes in music listening

Jonghwa Kim, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Kim, Jonghwa, and Elisabeth André. 2008. "Emotion recognition based on physiological changes in music listening." IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (12): 2067–83. https://doi.org/10.1109/tpami.2008.26.

Emotion Recognition Based on Physiological Changes in Music Listening

Jonghwa Kim, *Member*, *IEEE*, and Elisabeth André

Abstract—Little attention has been paid so far to physiological signals for emotion recognition compared to audiovisual emotion channels such as facial expression or speech. This paper investigates the potential of physiological signals as reliable channels for emotion recognition. All essential stages of an automatic recognition system are discussed, from the recording of a physiological data set to a feature-based multiclass classification. In order to collect a physiological data set from multiple subjects over many weeks, we used a musical induction method that spontaneously leads subjects to real emotional states, without any deliberate laboratory setting. Four-channel biosensors were used to measure electromyogram, electrocardiogram, skin conductivity, and respiration changes. A wide range of physiological features from various analysis domains, including time/frequency, entropy, geometric analysis, subband spectra, multiscale entropy, etc., is proposed in order to find the best emotion-relevant features and to correlate them with emotional states. The best features extracted are specified in detail and their effectiveness is proven by classification results. Classification of four musical emotions (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal) is performed by using an extended linear discriminant analysis (pLDA). Furthermore, by exploiting a dichotomic property of the 2D emotion model, we develop a novel scheme of emotion-specific multilevel dichotomous classification (EMDC) and compare its performance with direct multiclass classification using the pLDA. An improved recognition accuracy of 95 percent and 70 percent for subject-dependent and subject-independent classification, respectively, is achieved by using the EMDC scheme.

Index Terms—Emotion recognition, physiological signal, biosignal, skin conductance, electrocardiogram, electromyogram, respiration, affective computing, human-computer interaction, musical emotion, autonomic nervous system, arousal, valence.

1 Introduction

Resolving the absence of mutual sympathy (rapport) in interactions between humans and machines is one of the most important issues in advanced human-computer interaction (HCI) today. With exponentially evolving technology, it is no exaggeration to say that any interface that disregards human affective states in the interaction —and thus fails to pertinently react to the states—will never be able to inspire confidence. Instead, users will perceive it as cold, untrustworthy, and socially inept. In human communication, the expression and understanding of emotions helps achieve mutual sympathy. To approach this in HCI, we need to equip machines with the means to interpret and understand human emotions without the input of a user's translated intention. Hence, one of the most important prerequisites for realizing such an advanced user interface is a reliable emotion recognition system that guarantees acceptable recognition accuracy, robustness against any artifacts, and adaptability to practical applications. Developing such a system requires the following stages: modeling, analyzing, processing, training, and classifying emotional features measured from the implicit emotion channels of human communication, such as speech, facial expression, gesture, pose, physiological responses, etc. In this paper, we concentrate on identifying emotional cues in various physiological measures.

The debate on which emotion can be distinguished on the basis of physiological changes is far from being resolved in psycho and neurophysiology. Two well-known long-standing hypotheses are still under contention today, with James [1] supporting the antecedence of physiological specificity among emotional processes and Cannon [2] rejecting this claim. In neurophysiology, these opposing hypotheses can be reduced to the search for the central circuitry of emotions at the human level, that is, to finding the brain center in the central nervous system (CNS) and the neural center in the peripheral nervous system (PNS); all are involved in emotional experiences. The PNS is divided into two major parts, the somatic nervous system and the autonomic nervous system (ANS). The ANS consists of sensory neurons and motor neurons that run between the CNS and various internal organs such as the heart, lungs, viscera, and glands. For example, motor neurons of the autonomic system control the contraction of both the smooth muscle and the cardiac muscle. The ANS includes the sympathetic and parasympathetic systems.

In this paper, the expression "physiological changes" (often called "biosignals") exclusively applies to measures of the PNS functions, for example, electrodermal activity, heart and blood circulation, respiration (RSP), muscular activity, etc.

Recently, numerous studies on engineering approaches to automatic emotion recognition have been published, although research in that field is relatively new compared to the long history of emotion research in psychology and psychophysiology. In particular, many efforts have been deployed to recognize human emotions using audiovisual

• *The authors are with the Institut für Informatik, University of Augsburg, Eichleitnerstr. 30, D-86159 Augsburg, Germany.*
E-mail: {kim, andre}@informatik.uni-augsburg, de.

channels of emotion expression, that is, facial expressions, speech, and gestures. Little attention, however, has been paid so far to using physiological measures, as opposed to audiovisual emotion channels [3]. This is due to some significant limitations that come with the use of physiological signals for emotion recognition. The main difficulty lies in the fact that it is a very hard task to uniquely map physiological patterns onto specific emotional states. As an emotion is a function of time, context, space, culture, and person, physiological patterns may widely differ from user to user and from situation to situation. Above all, humans use nondiscrete labels to describe emotions. Second, recording of biosignals requires the user to be bodily connected with biosensors and sensing using surface electrodes is very sensitive to motion artifacts. Moreover, as we use various biosensors at the same time and each of them has its own specific characteristics, analyzing biosignals is itself a complex multivariate task and requires broad insight into biological processes related to neuropsychological functions. Third, obtaining the "ground truth" of physiological data for research purposes is a crucial problem. It differs from the cases of other external audiovisual channels. Labeling audiovisual corpora is relatively straightforward because they can be labeled based on objective judgments by comprehending "signs" that are associated with our common experiences in human communication and can therefore be interpreted even by perceiving and feeling them in facial expression and vocal intonation, for example. However, in the case of physiological signals, which can only be observed as a signal flow on an instrument screen, we can neither feel nor perceive emotions directly based on the signals. This leads to difficulties in data annotation as a universal data set for benchmarking research work is hard to obtain.

On the other hand, using physiological signals for emotion recognition provides some considerable advantages. We can continuously gather information about the users' affective states as long as they are connected with the biosensors. Consider extreme cases where people resort to the so-called "poker face" or simply do not say anything because they are angry. In those cases, the emotional states of the user remain internal and cannot be detected by any audiovisual recording system. Second, since the ANS activations are largely involuntary and generally cannot be easily triggered by any conscious or intentional control, we believe that physiological ANS activity would be a most robust emotional channel to combat artifacts created by human social masking. For example, it is not uncommon to observe that people smile during negative emotional experiences [4]. Such a smile is the result of social masking, where people regulate or modulate emotions interpersonally, and it should not be interpreted as the user's actual affective state. Last, experimental results have shown that some types of ANS activity are not culturally specific, that is, there is cross-cultural consistency of ANS differences between emotions. Levenson et al. [5] compared three physiological measures (heart rate, skin conductance, and finger temperature) sampled from Americans and the Minangkabau of West Sumatra and found significant levels of cross-cultural consistency in the ANS patterns among the four negative emotions, anger, disgust, fear, and sadness.

In this paper, we use four-channel biosignals to deal with all of the essential stages of an automatic emotion recognition system based on physiological measures, from data collection to the classification of four typical emotions (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal). The work in this paper is novel: in trying to recognize naturally induced musical emotions using physiological changes, in acquiring a physiological data set through everyday life recording over many weeks from multiple subjects, in finding emotion-relevant ANS specificity through various feature contents, and in designing an emotion-specific classification method. After the calculation of a great number of features (a total of 110 features) from various feature domains, we tried to identify emotion-relevant features using the backward feature selection method combined with a linear classifier. These features can be directly used to design affective human-machine interfaces for practical applications. Furthermore, we developed a novel scheme of emotion-specific multilevel dichotomous classification (EMDC) and compared its performance with direct multiclass classification. Although this new scheme is based on a very simple idea, exploiting the dichotomic structure of a 2D emotion model, it significantly improves the recognition accuracy obtained by using direct multiclass classification. Throughout the paper, we try to provide a focused spectrum for each processing stage with selected methods suitable for handling the nature of physiological changes, instead of conducting a comparison study based on a large number of pattern recognition methods.

In Section 2, we give a brief overview of related research on musical emotion and physiological ANS specificity in psychophysiology, as well as on automatic emotion recognition in engineering science. Section 3 gives the motivation and rationale for our experimental setting of musical emotion induction and is followed by a detailed explanation of all the biosensors we used. A systematic description of signal analysis methods and classification procedure using extended linear discriminant analysis (LDA) is given in Section 4. In Section 5, we present the best emotion-relevant ANS features with the recognition results we achieved. In addition, the performance of the novel EMDC scheme is tested, and its potential is proven by improved recognition accuracy. In Section 6, we discuss the problems faced during our work, including the difficulty in subject-independent recognition. We then conclude with perspectives related to future work.

2 Related Research

2.1 Physiological Differentiation of Emotions

We agree that emotion is not a phenomenon but a construct, which is systematically produced by cognitive processes, subjective feelings, physiological arousal, motivational tendencies, and behavioral reactions. Likewise, several influencing factors, including psychological processes such as attention, orientation, social interaction, and appraisal, may simultaneously impinge on the autonomous nervous system. Thus, proving that there is an ANS differentiation of emotions is an inherently difficult task.

Overall, there are a number of experiments that point to the fact that physiological activity is not an independent variable in ANS patterns but reflects experienced emotional states with consistent correlates [6], [7], [8]. In the psychophysiology literature, research on emotions with

negative valence has far outpaced research on positive emotions. For example, some reliable ANS differentiations have been observed in emotions produced by directed facial action and recalled emotional memories: heart rate acceleration in sadness, anger, and fear, heart rate deceleration in disgust, and larger skin conductance in fear and disgust than in happiness [7]. On the other side, however, there are many objections to ANS specificity. For example, Schachter and Singer [9] observed that undifferentiated arousal resulted in different reports of emotions depending on the subject's cognitive response to external events. Stemmler [10] reported that real-life fear (listening to the conclusion of Poe's *The Fall of the Usher* in a darkened room, with appropriately spooky music) led to a statistically different ANS activity than did a fear imagery task in which participants were asked to recollect and speak about a frightening personal event. Stemmler et al. [8] also asserted, concerning the various cognitive and situational factors that influence ANS activity, that a low degree of consistency of ANS specificity in the literature comes as no surprise since those influencing factors (contexts) vary widely across emotion studies and that, therefore, consistent ANS specificity among emotions could only be found if the compound of emotion-plus-context pattern is decomposed.

2.2 Music and Emotion

A primary motive for listening to music is its emotional effect, diversion, and the memories it awakens. Indeed, many studies have shown that the emotions intended by a performer are correctly recognized by listeners. Moreover, children as young as three might be able to readily recognize the intended emotions as adults do [11]. Although many scientists believe that music does not have the power to actually produce genuine emotional states though people do recognize the intended emotions, contemporary experiments have revealed that emotional reactions to music are real since music produces specific patterns of change in heart rate, blood pressure, and other autonomic bodily reactions that are linked to different emotions. Thus, research on musical emotions can be summarized in two main perspectives. Concerning the perception and production of emotions while listening to music, *emotivists* believe that music elicits emotions that are qualitatively similar to nonmusical emotions, while *cognitists* argue that the emotion is an expressive property of the music that listeners recognize in it but do not themselves experience [12]. In this section, we will briefly summarize previous research on physiological responses to music, focusing on the emotivist view of musical emotions.

It is a very old belief that music is a link between cognition and emotion and that music can influence ANS reactions both in an arousing and a calming fashion [13]. In his theory of musical emotions, Meyer [14] submits that emotions are time locked to events in the music and that a central factor of musical emotions is expectations that are derived from both general psychological principles (such as Gestalt principles of perceptual organization) and knowledge of the music style (such as tonality, harmonic progressions, and musical form). In keeping with this position, ample empirical evidence has recently been brought forward supporting music as a preeminent stimulus to evoke powerful emotions accompanied by differential changes in ANS reaction. For example, Vaitl et al. [15] attempted to find the ANS differentiation of musical emotion in live performance. While subjects were listening to the leitmotivs of a number of Wagner operas during the Bayreuth Festival (summers of 1987 and 1988), they recorded two physiological measures, electrodermal response and respiratory activity, and analyzed them using ratings for emotional arousal. Noticeable differentiations were observed in the physiological measures with respect to the leitmotivs and their musical features (for example, melody, rhythm, and continuation). A number of results also appear in clinical and therapeutic contexts. Davis and Thaut [16] found that music aroused ANS responses (vascular construction, heart rate, muscle tension, and finger skin temperature) even though subjects reported decreases in anxiety and increases in relaxation. Guzzetta [17] also reported physiological correlates of musical emotions, concluding that music is associated with lower heart rates and higher peripheral temperature.

If music is able to express the traditional basic discrete emotions (such as happiness, anger, and sadness) that are perceived when listening to music, it might also be able to produce the same emotions that we experience in our daily life. Krumhansl [13] recorded different physiological measures while listeners were hearing music that had been independently judged to be one of three emotions, that is, happiness, sadness, and fear, and analyzed them to find out what relationship existed between the physiological measures and the dynamic ratings of emotions. Interestingly, she found that the directions of the physiological changes were the same for all three emotions. The heart rate decreased, the blood pressure increased, RSP rate increased, and the skin temperature decreased, while the magnitude of the changes showed distinct patterns, depending on the emotional quality of the excerpt. For instance, happiness was linked to the largest changes in RSP, sadness involved the greatest changes in heart rate, blood pressure, and skin temperature, and fear was associated with maximal changes in the rate of blood flow. These findings convincingly support the hypothesis that music does not simply convey emotions that we can recognize but rather induces genuine emotions in the listener. However, the question of whether the ANS changes and differentiation in musical emotions correspond to those revealed in nonmusical emotions remains to be elucidated.

2.3 Approaches to Emotion Recognition Using Biosignals

A significant amount of work has been conducted by Picard et al. at the Massachusetts Institute of Technology (MIT) Laboratory, showing that certain affective states may be recognized by using physiological data, including heart rate, skin conductivity (SC), temperature, muscle activity, and RSP velocity [18], [19]. They used personalized imagery to elicit target emotions from a single subject who had two years of experience in acting and they achieved an overall recognition accuracy of 81 percent for eight emotions by using hybrid linear discriminant classification. Nasoz et al. [20] used movie clips based on the study of Gross and Levenson [21] for eliciting target emotions from 29 subjects and achieved an emotion classification accuracy of 83 percent using the Marquardt Backpropagation algorithm (MBP). In [22], the IAPS photoset [23] is used to elicit target emotions with positive and negative valence and variable arousal level from a single subject. The arousal and valence dimensions of the

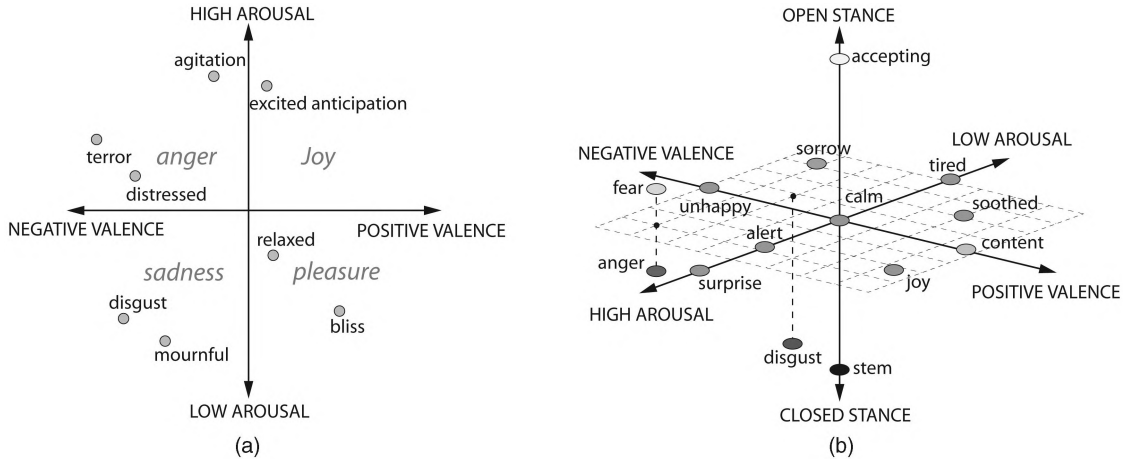

Fig. 1. Emotion models. (a) Two-dimensional model by valence and arousal. (b) Three-dimensional model by valence, arousal, and stance.

emotions were classified separately using a neural network classifier, and recognition accuracy rates of 96.6 percent and 89.9 percent, respectively, were achieved.

More recently, an interesting user-independent emotion recognition system was reported by Kim et al. [24]. They developed a set of recording protocols using multimodal stimuli (audio, visual, and cognitive) to evoke targeted emotions (sadness, stress, anger, and surprise) from 175 children aged five to eight. A classification ratio of 78.43 percent was achieved for three emotions (sadness, stress, and anger) and a ratio of 61.76 percent for four emotions (sadness, stress, anger, and surprise) by adopting support vector machines as a pattern classifier. Most interestingly, analysis steps in the system were fitted to handle relatively short lengths of the input signals (segmented in 50 seconds) compared to previous works that required longer signal lengths of about 2-6 min.

The aforementioned approaches achieved average accuracy rates of more than 80 percent, which seem to be acceptable for practical applications. It is true, however, that recognition rates are strongly dependent on the data sets that are used and on the application context. Moreover, the physiological data sets used in most of these works were gathered by using visual elicitation materials in a laboratory setting. The subjects then "tried and felt" or "acted out" the target emotions while looking at selected photos or watching movie clips that were carefully prearranged to elicit the emotions. In other words, to put it bluntly, the recognition results were achieved for specific users in specific contexts with "forced" emotional states. The emotional state or mood the subjects were in before starting the experiments, for instance, was not taken into consideration. Such individual differences can cause inconsistencies in the data sets. Another factor of the inconsistency is the uncertainty concerning the labeling of data sets due to different individual judgments (or self-reports) and the situational variables in ANS activity, as Stemmler argued in his reports [10].

Most of the aforementioned engineering approaches, however, provide evidence of the fact that the accuracy of arousal discrimination is always higher than that of valence differentiation. The reason might be the that the change in the arousal level corresponds directly to the intensity of discharge in ANS activities, such as sweat glands and blood pressure, which is straightforward to measure, while valence differentiation of emotion requires a multifactor analysis of cross-correlated ANS reactions. This finding led us to develop an emotion-specific classification scheme and to calculate a wide range of features in various analysis domains in order to extract valence-relevant features from ECG and RSP signals.

2.4 Modeling of Discrete Emotions

As all people express their emotions differently, it is not an easy task to judge or to model human emotions. Researchers often use two different methods to model emotions. One approach is to label emotions in discrete categories, that is, human judges have to choose from a prescribed list of word labels, for example, joy, sadness, surprise, anger, love, fear, etc. One problem with this method is that the stimuli may contain blended emotions that cannot be adequately expressed in words since the choice of words may be too restrictive and culturally dependent. Another way is to have multiple dimensions or scales to categorize emotions. Instead of choosing discrete labels or words, observers can indicate their impression of each stimulus on several continuous scales, for example, pleasant-unpleasant, attention-rejection, simple-complicated, etc.

Two common scales are valence and arousal. Valence represents the pleasantness of stimuli, with positive (or pleasant) at one end and negative (or unpleasant) at the other. For example, happiness has a positive valence, while disgust has a negative valence. Another dimension is arousal (activation level). For example, sadness has low arousal, whereas surprise has a high arousal level. The different emotional labels can be plotted at various positions on a 2D plane spanned by these two axes to construct a 2D emotion model [25] (see Fig. 1a). The low consistency of physiological configurations in recent research has helped support the hypothesis that ANS activation during emotions indicates the demands of a specific action tendency and action disposition, instead of reflecting emotions per se [26]. Scholsberg [27] suggested a 3D model in which he had attention-rejection in addition to the 2D model. Researchers have subsumed these associated action tendencies under the term "stance" in a 3D emotion model, that is, arousal, valence, and stance (Fig. 1b). For example, fear is associated with the action pattern of "flight," anger calls to mind the urge to "fight," and so on. However, it is not immediately obvious what elemental

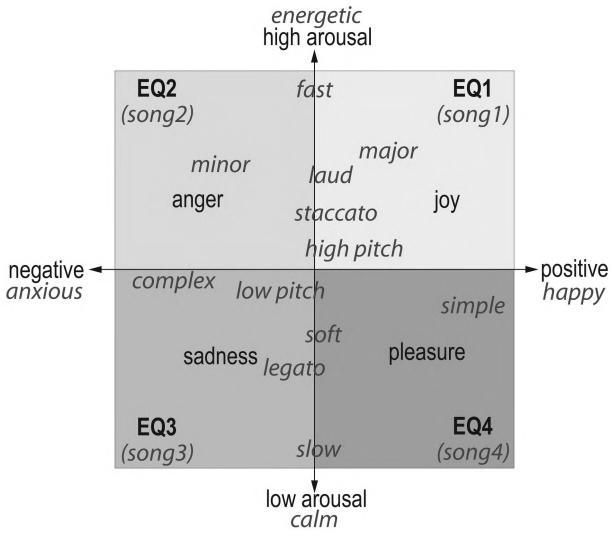

Fig. 2. Reference emotional cues in music based on the 2D emotion model. EQ1 = positive/high arousal, EQ2 = negative/high arousal, EQ3 = negative/low arousal, and EQ4 = positive/low arousal.

problem happiness solves and what action pattern or motor program is associated with this emotion. Thus, such positive emotions seem to be characterized by a lack of autonomic activation and this might be one reason why research on positive emotions has been lagging behind that on negative emotions so far. Interestingly, Fredricson and Levenson [28] reported the "undoing" effect of positive emotions, namely, that certain positive emotions help speed up recovery from the cardiovascular sequelae of negative emotions. This finding supports the idea of a symmetric process underlying the emotion system that negative emotions help the organism escape from homeostasis while positive emotions such as contentment and amusement catalyze a more rapid return to homeostatic levels.

3 Setting of Experiment

3.1 Musical Emotion Induction

To collect a database of physiological signals in which the targeted emotions corresponding to the four quadrants in the 2D emotion model (that is, EQ1, EQ2, EQ3, and EQ4 in Fig. 2) can be *naturally* reflected without any deliberate expression, we decided to use the musical induction method, that is, to record physiological signals while the subjects were listening to different pieces of music.

A well-established mechanism of emotion induction consists of triggering emotions by resorting to imagination or individual memories. Emotional reaction can be triggered by a specific cue and be evoked by an experimental instruction to imagine certain events. On the other hand, it can be spontaneously resurged in memory. Music is a pervasive element accompanying many highly significant events in human social life and particular pieces of music are often connected to significant personal memories. Following this, music can be a powerful cue in awakening emotional experiences and bringing back memories. Since listening to music is often done by an individual in isolation, the possible artifacts of social masking and social interaction can be minimized in the experiment. Furthermore, like odors, music can be treated at lower levels of the brain that are particularly resistant to modifications by later input, contrary to cortically-based episodic memory [29]. This is even the case when the listening occurs at the same time as other activities within a social setting since musical emotion cannot co-occur with social interaction in general.

The subjects were three males (one of the coauthors and two student researchers recruited from the authors' laboratory), aged 25-38, who all enjoy listening to music in their everyday life. The subjects were not paid but were allowed to perform the experiments during their regular working hours. They individually handpicked four songs that were intended to spontaneously evoke emotional memories and certain moods corresponding to the four target emotions. Fig. 2[1] shows the musical emotion model referred to for the selection of their songs. Generally, emotional responses to music vary greatly from individual to individual, depending on their unique past experiences. Moreover, cross-cultural comparisons in the literature suggest that emotional responses can be quite differentially emphasized by different musical cultures and training. This is why we advised the subjects to choose for themselves the songs they believed would help them recall their individual special memories with respect to the target emotions.

For the experiment, we prepared a quiet listening room in our institute in order to ensure that the subjects could experience the emotions evoked by the music undisturbed. For the recording, the subject had to position the sensors following the instructions posted in the room, put on the headphones, and select a song from his song list saved in the computer. When clicking on the selected song, the recording and music systems were automatically set up by preset values for each song, such as volume, treble, and bass. Most importantly, before the start of the experiment, the subjects were shown how to prepare the skin by using an antiseptic spray and a skin preparation gel for reducing electrode impedance and how to correctly position the sensors. Recording schedules were decided by the subjects themselves and the recordings took place whenever they felt like listening to music. They were also free to choose the songs they wanted to listen to. Thus, in contrast to methods used in other studies, the subjects were not forced to participate in a laboratory setting scenario and to use prespecified stimulation material. We believe that this voluntary participation of the subjects during our experiment might help obtain a high-quality data set with natural emotions.

During the three months, a total of 360 samples (90 samples for each emotion) from three subjects were collected. The signal length of each sample was between 3 and 5 minutes, depending on the duration of the songs.

3.2 Biosensors

The physiological signals were acquired using the Procomp[2] Infiniti with four biosensors: electromyogram (EMG), SC, electrocardiogram (ECG), and RSP. The sampling rates were 32 Hz for EMG, SC, and RSP, and 256 Hz for ECG. The positions and typical waveforms of the biosensors we used are illustrated in Fig. 3.

1. Metaphoric cues for song selection: song1 (positively exciting, energizing, joyful, and exuberant), song2 (noisy, loud, irritating, and discord), song3 (melancholic and sad memory), and song4 (blissful, pleasurable, slumberous, and tender).

2. This is an eight-channel multimodal Biofeedback system with 14-bit resolution and a fiber optic cable connection to the computer. www.Mind Media.nl.

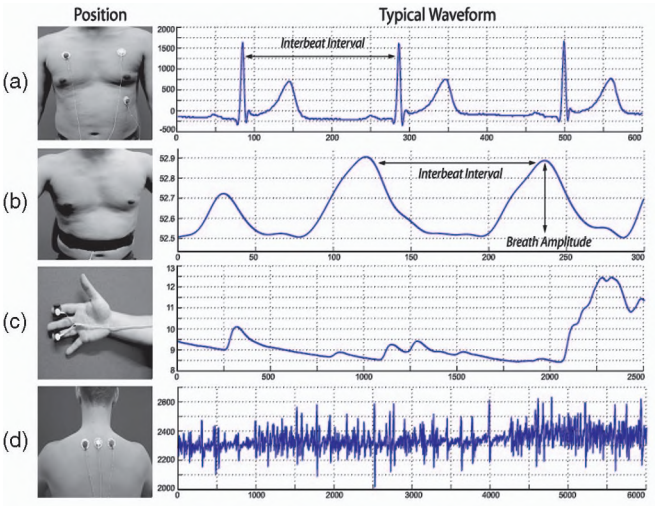

Fig. 3. Position and typical waveforms of the biosensors. (a) ECG. (b) RSP. (c) SC. (d) EMG.

3.2.1 Electrocardiogram

We used a preamplified electrocardiograph sensor (bandwidth: 0.05 Hz-1 KHz) connected with pregelled single Ag/AgCl electrodes. We cannot measure individual action potentials directly in the heart. We can, however, measure the average action potential on the skin. The mean movement of the action potential is along the "electrical axis" of the heart. The action potential starts high in the right atrium, moves to the center of the heart, and then moves down toward the apex of the heart. Therefore, the main electrical signal from the heart flows away from the upper right of the body toward the lower left of the body. Common features of the ECG signal are heart rate, interbeat interval, and heart rate variability (HRV). The heart rate reflects emotional activity. Generally, it has been used to differentiate between positive and negative emotions, with further differentiation made possible with finger temperature. HRV refers to the oscillation of the interval between consecutive heartbeats. It has been used as an indication of mental effort and stress in adults. In high-stress environments such as dispatch and air-traffic control, it is known to be a useful measure.

3.2.2 Electromyogram

We used a Myoscan-Pro sensor with an active range of 20-500 Hz and pregelled single Ag/AgCl electrodes. It can record EMG signals of up to 1,600 μV. Electromyography measures muscle activity by detecting surface voltages that occur when a muscle is contracted. Therefore, the best readings are obtained when the sensor is placed on the muscle belly and its positive and negative electrodes are parallel to the muscle fibers. Since the number of muscle fibers that are recruited during any given contraction depends on the force required to perform the movement, the intensity (amplitude) of the resulting electrical signal is proportional to the strength of contraction. In psychophysiology, EMG was often used to find the correlation between cognitive emotion and physiological reactions. In the work by Sloan [30], for example, the EMG was positioned on the face (jaw) to distinguish "smile" and "frown" by measuring the activity of zygomatic major and corrugator supercilli. In our experiment, bipolar electrodes were placed at the upper trapezius muscle (near the neck) in order to measure the mental stress of the subjects [31].

3.2.3 Respiration

A stretch sensor using a latex rubber band fixed with a Velcro RSP belt was used to capture the breathing activity of the subjects. It can be worn either thoracically or abdominally over clothing. The amount of stretch in the elastic is measured as a voltage change and recorded. The rate of RSP and depth of breath are the most common measures of RSP. Although RSP rate generally decreases with relaxation, startling events and tense situations may result in momentary RSP cessation. Negative emotions generally cause irregularity in the RSP pattern. Because RSP is closely linked to cardiac function, a deep breath can affect other measures, for example, EMG and SC measurements. In our experiment, this irregularity could be observed when the subject was talking. The RSP cycle can also be obtained by monitoring the contents of carbon dioxide (CO_2) in the inhaled/exhaled air, known as *capnography*, or by measuring the chest cavity expansion.

3.2.4 Skin Conductivity

SC is one of the measurements most often used to capture the affective state of users, especially for arousal difference. Many studies over the years have indicated that the magnitude of electrodermal change and the intensity of emotional experience are almost linearly associated in arousal dimension, [25], [32]. The SC sensor measures the skin's ability to conduct electricity. A small voltage is applied to the skin and the skin's current conduction or resistance is measured. Therefore, skin conductance is considered to be a function of the activity of the eccrine sweat glands (located in the palms of the hands and soles of the feet) and the skin's pore size. We used Ag/AgCl electrodes fixed with a two-finger band and positioned at the index and ring fingers of the nondominant hand. The SC consists of two separate components. There is a slow-moving tonic component that indicates general activity of the perspiratory glands due to temperature or other influences and a faster phasic component that is influenced by emotions and the level of arousal. For example, when a subject is startled or experiences anxiety, there will be a fast increase in the skin conductance due to increased activity in the sweat glands.

4 Methodology

The overall structure of our recognition system is illustrated in Fig. 4. After the preprocessing stage for signal segmentation and denoising, we calculated 110 features from the four-channel biosignals and selected the most significant features by using the sequential backward search method. For classification, various machine learning methods (supervised classification in our case) can be used [33]. After having tested some classifiers such as k-nearest neighbor (k-NN), multilayer perceptron (MLP), and LDA, we chose the LDA which outperformed with higher recognition accuracy in our case. It should, however, be noted that there is no single best classification algorithm and the choice of the best classification method strongly depends on the characteristics of the data set to be classified. In the work by King et al. [34], for example, this

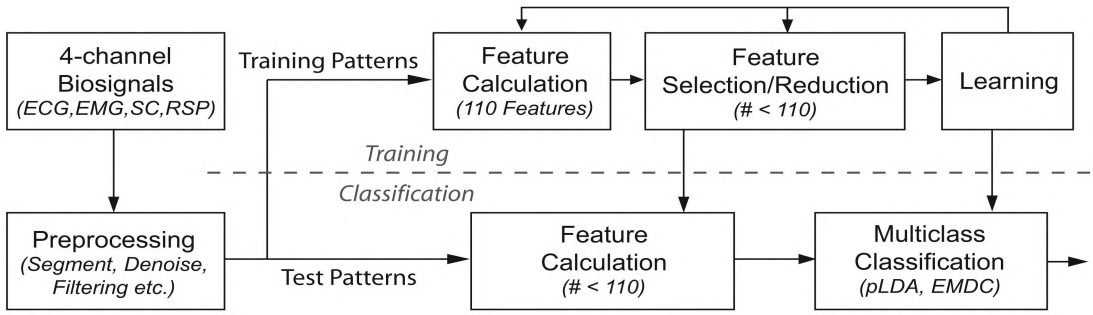

Fig. 4. Block diagram of supervised statistical classification system for emotion recognition.

conclusion was supported by a vast comparative study of about 20 different machine learning algorithms, including symbolic learning, neural networks, and statistical approaches, evaluated on 12 different real-world data sets.

4.1 Preprocessing

Different types of artifacts were observed in all of the four-channel signals, such as transient noise due to the movement of the subjects during the recording, mostly at the beginning and at the end of each recording. Thus, uniformly, for all subjects and channels, we segmented the signals into final samples of 160 seconds each, obtained by taking the middle part of each signal. It is important to note that the EMG signal generally requires additional preprocessing, such as deep smoothing or signal separation, depending on the position of the sensor, because the nature of the signal is such that all of the muscle fibers within the recording area of the sensor contract at different rates. In our case, the EMG signal contains artifacts generated by the heartbeat and RSP since we positioned the sensor at the upper trapezius muscle. Using an adaptive bandpass filter, we removed the artifacts (Fig. 5). For other signals, we used pertinent low-pass filters to remove noises without loss of information.

4.2 Measured Features

From the four-channel signals, we calculated a total of 110 features from various analysis domains, including conventional statistics in time series, frequency domain, geometric analysis, multiscale sample entropy (MSE), subband spectra, etc. For the signals with nonperiodic characteristics, such as EMG and SC, we focused on capturing the amplitude variance and localizing the occurrences (number of transient changes) in the signals. In the following sections, we describe the feature calculation methods in detail.

4.2.1 Electrocardiogram

ECG measures depolarized electrical changes of muscular contraction associated with cardiovascular activity. In general, the ECG is measured at the body surface along the axis of the heart and results from the activation, first, of the two small heart chambers, the atria, and, then, of the two larger heart chambers, the ventricles. The contraction of the ventricles produces the specific waveform known as the QRS complex (see Fig. 6).

To obtain the subband spectrum of the ECG signal, we used the typical 1,024 points fast Fourier transform (FFT) and partitioned the coefficients within the frequency range 0-10 Hz into eight nonoverlapping subbands with equal bandwidth. First, as features, the power mean values of each subband and the fundamental frequency (F0) are calculated by finding the maximum magnitude in the spectrum within the range 0-3 Hz. To capture peaks and their locations in subbands, the subband spectral entropy (SSE) is computed for each subband. Entropy plays an important role in information theory as a measure of disorganization or uncertainty in a random variable. In pattern recognition, it is generally used to measure the degree of a classifier's confidence. To compute the SSE, it is necessary to convert each spectrum into a probability mass

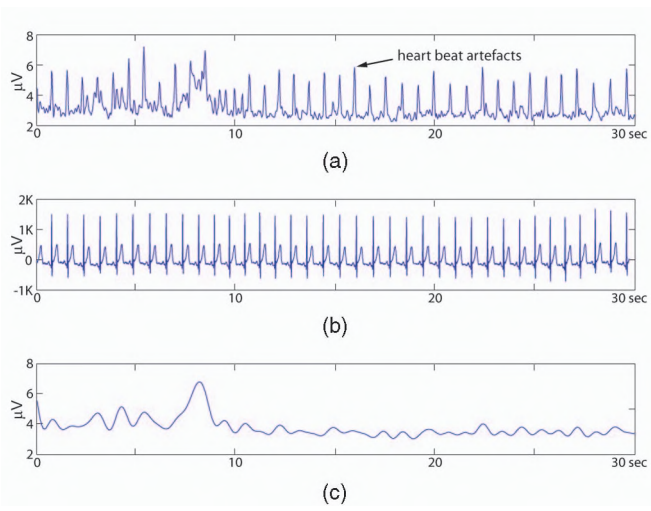

Fig. 5. Example of an EMG signal with heartbeat artifacts and denoised signal.

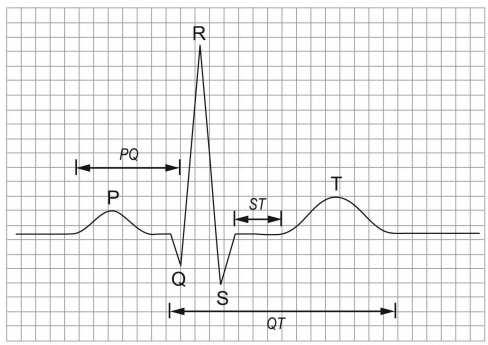

Fig. 6. QRS waveform in an ECG signal. Usual lengths: P-wave (0.08-0.10 s), QRS (0.06-0.10 s), PR-interval (0.12-0.20 s), and QT_c-interval $(QT/\sqrt{RR} \leq 0.44\ \mathrm{s})$ [35].

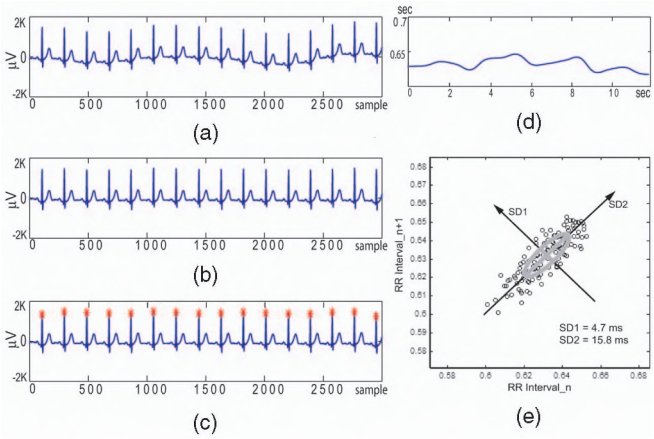

Fig. 7. Example of ECG analysis. (a) Raw ECG signal with RSP artifacts. (b) Detrended signal. (c) Detected RR interbeats. (d) Interpolated HRV time series using RR intervals. (e) Poincaré plot of the HRV time series.

function (PMF)-like form. Equation (1) is used for the normalization of the spectrum:

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i}, \qquad \text{for } i = 1 \ldots N, \tag{1}$$

where X_i is the energy of the ith frequency component of the spectrum and $\tilde{\mathbf{x}} = \{x_1 \ldots x_N\}$ is to be considered as the PMF of the spectrum. In each subband, the SSE is computed from $\tilde{\mathbf{x}}$ by

$$H_{\text{sub}} = -\sum_{i=1}^{N} x_i \cdot \log_2 x_i. \tag{2}$$

By packing the eight subbands into two bands, that is, subbands 1-3 as the low-frequency (LF) band and subbands 4-8 as the high-frequency (HF) band, the ratios of the LF/HF bands are calculated from the power mean values and the SSEs.

In biomedical engineering, the analysis of the local morphology of the QRS waveform and its time-varying properties has been a standard method for assessing cardiac health [35]. Importantly, HRV is one of the most often used measures for ECG analysis. To obtain the HRV from the continuous ECG signal, each QRS complex is detected and the RR intervals (all intervals between adjacent R waves) or the normal-to-normal (NN) intervals (all intervals between adjacent QRS complexes resulting from sinus node depolarization) are determined. We used the QRS detection algorithm of Pan and Tompkins [36] in order to obtain the HRV time series. Fig. 7 shows examples of R-wave detection and an interpolated HRV time series, referring to the increases and decreases over time in the NN intervals.

In the time domain of the HRV time series, we calculated statistical features, including the mean value, the standard deviation of all NN intervals (SDNN), the standard deviation of the first difference of the HRV, the number of pairs of successive NN intervals differing by more than 50 ms (NN50), and the proportion derived by dividing NN50 by the total number of NN intervals. By calculating the standard deviations in different distances of RR interbeats, we also added Poincaré geometry in the feature set to capture the nature of interbeat interval fluctuations. Poincaré plot geometry is a graph of each RR interval plotted against the next interval and provides quantitative information of the heart activity by calculating the standard deviations of the distances of $R - R(i)$ to lines $y = x$ and $y = -x + 2 * R - R_m$, where $R - R_m$ is the mean of all $R - R(i)$ [37]. Fig. 7e shows an example plot of the Poincaré geometry. The standard deviations SD_1 and SD_2 refer to the fast beat-to-beat variability and longer term variability of $R - R(i)$, respectively.

Entropy-based features from the HRV time series were also considered. Based on the so-called *approximate entropy* and *sample entropy* proposed in [38], an MSE was introduced [39] and successfully applied to physiological data, especially for the analysis of short and noisy biosignals [40]. Given a time series $\{X_i\} = \{x_1, x_2, \ldots, x_N\}$ of length N, the number $(n_i^{(m)})$ of similar m-dimensional vectors $y^{(m)}(j)$ for each sequence vector $y^{(m)}(i) = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}$ is determined by measuring their respective distances. The relative frequency to find the vector $y^{(m)}(j)$ within a tolerance level δ is defined by

$$C_i^{(m)}(\delta) = \frac{n_i^{(m)}}{N - m + 1}. \tag{3}$$

The approximate entropy $h_A(\delta, m)$ and the sample entropy $h_S(\delta, m)$ are defined as

$$h_A(\delta, m) = \lim_{N \to \infty} \left[H_N^{[m]}(\delta) - H_N^{(m+1)}(\delta) \right], \tag{4}$$

$$h_S(\delta, m) = \lim_{N \to \infty} -\ln \frac{C^{(m+1)}(\delta)}{C^{(m)}(\delta)}, \tag{5}$$

where

$$H_N^{(m)}(\delta) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i^{(m)}(\delta). \tag{6}$$

Because it has the advantage of being less dependent on the time-series length N, we applied the sample entropy h_S to coarse-grained versions $(y_j^{(\tau)})$ of the original HRV time series $\{X_i\}$:

$$y_j(\tau) = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le N/\tau, \quad \tau = 1, 2, 3, \ldots. \tag{7}$$

The time series $\{X_i\}$ is first divided into N/τ segments by nonoverlapped windowing with length-of-scale factor τ and, then, the mean value of each segment is calculated. Note that, for scale one, $y_j(1) = x_j$. From the scaled time series $y_j(\tau)$, we obtain the m-dimensional sequence vectors $y^{(m)}(i, \tau)$. Finally, we calculate the sample entropy h_S for each sequence vector $y_j(\tau)$. In our analysis, we used $m = 2$ and fixed $\delta = 0.2\sigma$ for all scales, where σ is the standard deviation of the original time series x_i. Note that using the fixed tolerance level δ as a percentage of the standard deviation corresponds to the initial normalizing of the time series and it thus ensures that h_S does not depend on the variance of the original time series but only on their sequential ordering.

In the frequency domain of the HRV time series, three frequency bands are of general interest: the very LF (VLF) band (0.003-0.04 Hz), the LF band (0.04-0.15 Hz), and the

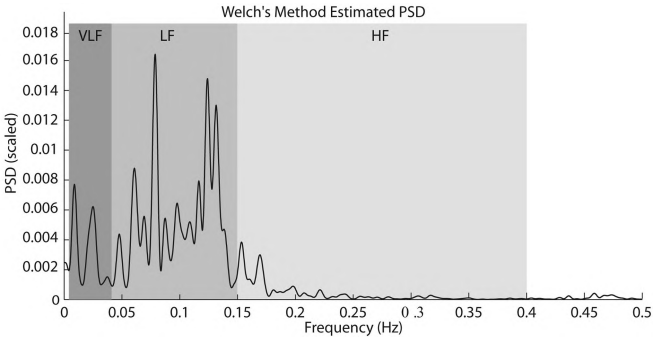

Fig. 8. Example of the heart rate spectrum in three subbands using the 1,024-point Fast Fourier transform.

HF band (0.15-0.4 Hz). From these subband spectra, we computed the dominant frequency and power of each band by integrating the power spectral densities (PSDs) obtained by using Welch's algorithm, as well as the ratio of power within the LF band to that within the HF band (LF/HF). Since parasympathetic activity dominates at HF, the LF/HF ratio is generally thought to distinguish sympathetic effects from parasympathetic effects [41]. Fig. 8 shows the heart rate spectrum from one of the subjects.

4.2.2 Respiration

RSP signal (breathing rate and intensity) is commonly acquired by measuring the physical change of the thoracic expansion with a rubber band around the chest or belly and contains fewer artifacts in general than the other sensors using electrodes, for example, ECG, EMG, SC, etc. Including the typical statistics of the raw RSP signal, we calculated similar types of features such as the ECG features, the power mean values of three subbands (obtained by dividing the Fourier coefficients within the range 0-0.8 Hz into nonoverlapped three subbands with equal bandwidth), and the set of SSEs.

In order to investigate the inherent correlation between the RSP rate and the heart rate, we considered a novel feature content for the RSP signal. Since an RSP signal exhibits a quasi-periodic waveform with sinusoidal properties, it does not seem unreasonable to conduct an HRV-like analysis for the RSP signal, that is, analysis of breathing rate variability (BRV). After detrending using the mean value of the entire signal and low-pass filtering, we calculated the BRV time series, referring to the increases and decreases over time in the peak-to-peak (PP) intervals, by detecting the peaks in the signal using the maximum ranks within each zero crossing (Fig. 9).

From the BRV time series, we calculated the mean value, SD, SD of the first difference, MSE, Poincaré analysis, etc. In the spectrum of the BRV, the peak frequency, the power of the two subbands, the LF band (0-0.03 Hz), the HF band (0.03-0.15 Hz), and the ratio of the power within the two bands (LF/HF) were calculated.

4.2.3 Skin Conductivity

The SC signal includes two types of electrodermal activity: the DC level component and the skin conductance response (SCR). The DC level in the SC signal indicates the general activity of the perspiratory glands influenced by body temperature or external temperature. The SCR is the distinctive short waveform in the SC signal and is considered

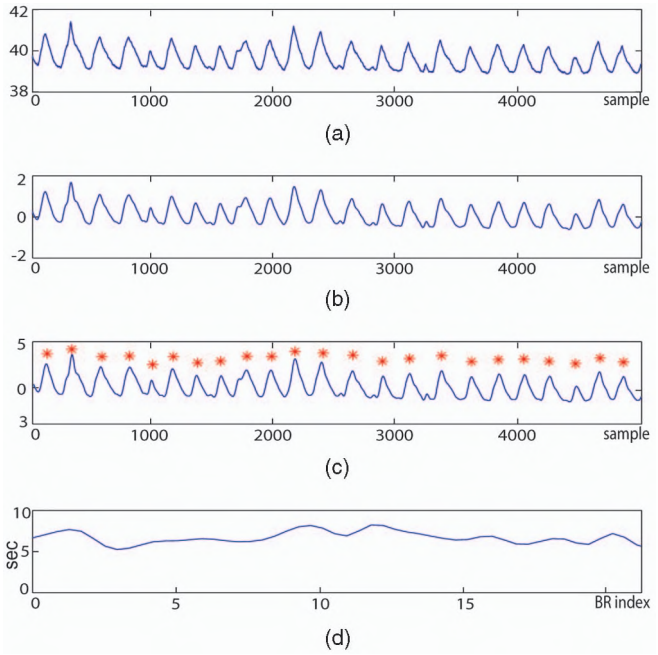

Fig. 9. BRV analysis for an RSP signal. (a) Raw RSP signal with Fs = 32Hz. (b) LP and detrended signal of (a). (c) Peak detection. (d) BRV time series referring to PP intervals.

to be useful for emotion recognition as it is linearly correlated with the intensity of arousal responding to internal/external stimuli. The mean value, standard deviation, and mean of the first and second derivations were extracted as features from the normalized SC signal and the low-passed (LP) SC signal using a cutoff frequency of 0.2 Hz. To obtain a detrended SCR waveform without DC-level components, we removed the continuous piecewise linear trend in the two LP signals, that is, the very LP (VLP) and the LP signal with a cutoff frequency of 0.08 Hz and 0.2 Hz, respectively (see Figs. 10a, 10b, 10c, 10d, and 10e).

The baseline of the SC signal was calculated and subtracted to consider only relative amplitudes. By finding two consecutive zero crossings and the maximum value between them, we calculated the number of SCR occurrences within 100 seconds from each LP and VLP signal, the mean of the amplitudes of all occurrences, and the ratio of the SCR occurrences within the LP signals (VLP/LP).

4.2.4 Electromyography

For the EMG signal, we calculated types of features similar to those of the SC signal. The mean value of the entire signal, the mean of the first and second derivations, and the standard deviation were extracted as features from the normalized and LP signals. The occurrence number of myoresponses and the ratio of that within VLP and LP signals were also added to the feature set and were determined in the same way as the SCR occurrence, but using cutoff frequencies of 0.08 Hz (VLP) and 0.3 Hz (LP) (see Figs. 10f, 10g, 10h, 10i, and 10j).

In the end, we obtained a total of 110 features from the four-channel biosignals: 53 (ECG) + 37 (RSP) + 10(SC) + 10 (EMG). See Table 4.

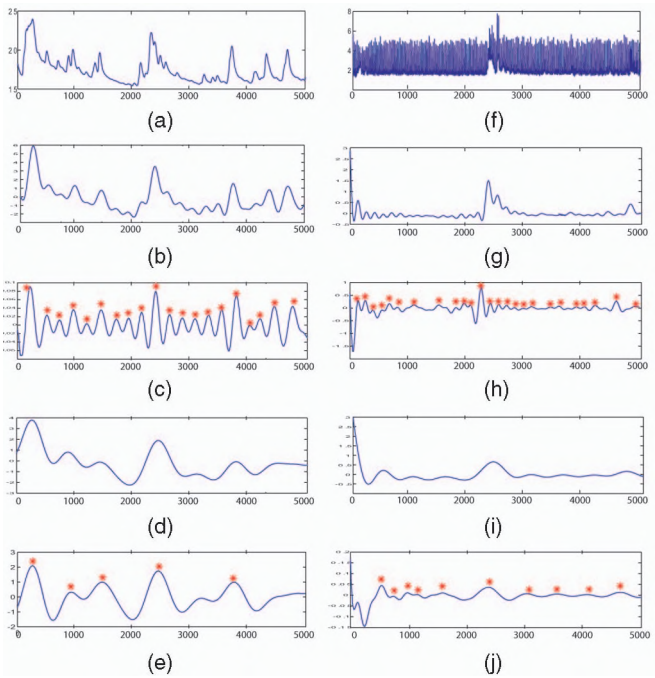

Fig. 10. Analysis examples of SC and EMG signals. (a) SC_raw signal. (b) SC_lowpassed, $fc = 0.2$ Hz. (c) SC_detrended, # occurrence. (d) SC_vlowpassed, $fc = 0.08$ Hz. (e) SC_detrended, # occurrence. (f) EMG_raw signal. (g) EMG_lowpassed, $fc = 0.3$ Hz. (h) EMG_detrended, # occurrence. (i) EMG_vlowpassed, $fc = 0.08$ Hz. (j) EMG_detrended, # occurrence.

4.3 Classification

4.3.1 Feature Selection

Compared to the works we reviewed, we calculated a relatively large amount of features within the various domains described in the previous sections. Since we calculated these features based on the signal analysis aspect exclusively, without any preliminary information on which physiological pattern might be correlated with which emotion type, there may exist garbage features within the calculated features that inherently have no bearing on the differentiation of the four emotion types. Such garbage features can ultimately reduce the performance of classifiers constructed from a limited number of training samples. If we consider the ratio between the number of features (110 features) and the fixed sample size (360 samples) in our case, we must consider that a classifier can also suffer from the "curse of dimensionality" [33]. Hence, the most essential step in our recognition system is to select salient emotion-relevant features from the given feature vectors and to map them into the emotional cues.

A large number of algorithms for feature subset selection have been proposed in the literature [42], [43], including sequential forward selection (SFS), sequential backward selection (SBS), sequential floating forward selection (SFFS), genetic algorithm (GA), etc. Most algorithms for feature selection use a criterion based on a specific classifier and are therefore useful if the classifier to be used is already known. SFS performs a heuristic-guided Depth-First search on the feature space. By starting with an empty subset, all features not yet included in the subset are sequentially incorporated in the subset and a criterion value is computed. On each iteration, the feature that yields the best value is then included in the new subset. SBS is the top-down equivalent of SFS since it begins with a complete set of the features and removes one feature on every iteration. We tested both selection methods in combination with LDA (see Section 5) as a classifier. Although SBS is computationally more demanding than SFS, we decided to use SBS in our recognition system because it outperformed SFS in the feature space. This might be explained by the fact that SBS evaluates the contribution of a given feature in the context of all other features, while SFS only evaluates the contribution of a feature in the limited context of the previously selected features. We must, however, note that the performance of all the selection methods proposed is strongly dependent on the given data set.

We did not consider integrating a dimensionality reduction method in our recognition scheme, such as principal component analysis (PCA) and Fisher projection, which are commonly used in combination with a classifier. Dimensionality reduction amounts to projecting high-dimensional data to a lower dimensional space with a minimal loss of information. This means that new features are created by the transformation of original feature values, rather than by selecting a feature subset from a given feature set. Such feature reduction methods were not suitable for the purpose of our work since we sought to determine the best emotion-relevant features that preserve their origins of analysis domain and value. We use Fisher projection exclusively to preview the distribution of the features.

4.3.2 Classifying Using Extended Linear Discriminant Analysis

In discriminant analysis, for a given data set, three scatter matrices, within-class (S_w), between-class (S_b), and mixture scatter matrices (S_m), are defined as follows:

$$S_b = \sum_{i=1}^{c} N_i(\mu_i - \overline{\mathbf{x}})(\mu_i - \overline{\mathbf{x}})^T = \mathbf{\Phi}_b \mathbf{\Phi}_b^T, \tag{8}$$

$$S_w = \sum_{i=1}^{c} \sum_{j \in C_i} (\mathbf{x}_j - \mu_i)(\mathbf{x}_j - \mu_i)^T = \mathbf{\Phi}_w \mathbf{\Phi}_w^T, \tag{9}$$

$$S_m = S_b + S_w = \sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T = \mathbf{\Phi}_m \mathbf{\Phi}_m^T, \tag{10}$$

where N is the number of all samples, N_i is the number of samples in class C_i $(i = 1, 2, \ldots, c)$, μ_i is the mean of the samples in class C_i, and $\overline{\mathbf{x}}$ is the mean of all samples, that is,

$$\mu_i = \frac{1}{N_i} \sum_{i \in C_i} \mathbf{x}_i, \tag{11}$$

$$\overline{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{c} \mathbf{x}_i = \frac{1}{N} \sum_{i=1}^{c} N_i \mu_i. \tag{12}$$

Note that the mixture scatter matrix S_m is the covariance matrix of all samples regardless of their class assignments, and all of the scatter matrices are designed to be invariant under coordinate shifts. The idea in LDA is to find an optimal transformation W that satisfies

$$\mathcal{J}(W) = \underset{W}{\operatorname{argmax}} \frac{|W^T S_b W|}{|W^T S_w W|}, \quad (13)$$

such that the separation between classes is maximized while the variance within a class is minimized (Fisher's criterion). Finding the optimal W is equivalent to finding the generalized eigenvectors satisfying $S_b W = \lambda S_w W$, for $\lambda \neq 0$. Transformation W can be obtained by applying the eigenvalue decomposition to the matrix $S_w^{-1} S_b$ if S_w is nonsingular or to the matrix $S_b^{-1} S_w$ if S_b is nonsingular and taking the rows of the transformation matrix to be the eigenvectors corresponding to the $n-1$ largest eigenvalues. It is shown in [44] that applying the singular value decomposition (SVD) on the scatter matrices of the training set is a stable way to compute the eigenvalue decomposition. Since there are at most $c-1$ nonzero generalized eigenvectors of the scatter matrix, the upper bound of the number of retained dimensions in classical LDA is $c-1$ and the dimensionality can be further reduced, for example, by incorporating in W only those eigenvectors corresponding to the largest singular values determined in the scatter SVD. Given the transformation W, classification can be performed in the transformed space based on some distance measures d such as euclidean distance. The new instance $\mathbf{v}$ is classified to

$$\underset{k}{\operatorname{argmin}}\, d\,(\mathbf{v}W, \overline{\mathbf{x}}_k W), \quad (14)$$

where $\overline{\mathbf{x}}_k$ is the centroid of the kth class and $k = 1, 2, \ldots, c$.

Note that a limitation of conventional LDA is that its objective function requires that one of the scatter matrices is nonsingular. It means that, for a given c-class p-dimensional classification problem, at least $c+p$ samples are required to guarantee that the within-class scatter matrix S_w does not become singular. To deal with the singularity problem, several extended LDA methods are proposed, such as PCA+LDA, pseudoinverse LDA (pLDA), regularized LDA, and LDA using generalized SVD (GSVD). In our work we used pLDA, a natural extension of classical LDA, applying the eigenvalue decomposition to the matrix $S_b^+ S_w$, $S_w^+ S_b$, or $S_m^+ S_b$. The pseudoinverse matrix is a generalization of the inverse matrix and exists for any $m \times n$ matrix. The computationally simplest way to get the pseudoinverse is using SVD; if $A = U\Sigma V^T$ is the singular value decomposition of A, then the pseudoinverse $A^+ = V\Sigma^+ U^T$. For a diagonal matrix such as Σ, we get the pseudoinverse by taking the reciprocal of each nonzero element on the diagonal.

5 Results

5.1 Classification Using SBS + pLDA

The confusion matrix in Table 1 presents the correct classification ratio ($\mathcal{CCR}$) of subject-dependent (Subjects A, B, and C) and subject-independent (All) classification where the features of all of the subjects are simply merged and normalized. We used the leave-one-out cross-validation method, where a single observation taken from the samples is used as the test data while the remaining observations are used for training the classifier. This is repeated such that each observation in the samples is used once as the test data.

The table shows that the $\mathcal{CCR}$ varies from subject to subject. For example, the best accuracy was 91 percent for Subject B and the lowest was 81 percent for Subject A. Not only does the overall accuracy differ from one subject to the next, but the $\mathcal{CCR}$ of the single emotions varies as well. For example, EQ2 was perfectly recognized for Subject C, while it caused the highest error rate for Subject B. It was mixed up three times with EQ1, which is characterized by opposite valence. As the confusion matrix shows, the difficulty in valence differentiation can be observed for all subjects. Most classification errors for Subjects A and B lie in false classification between EQ1 and EQ2, while an extreme uncertainty can be observed in the differentiation between EQ3 and EQ4 for Subject C. On the other hand, it is very meaningful that relatively robust recognition accuracy is achieved for the classification of emotions that are reciprocal in the diagonal quadrants of the 2D emotion model, that is, EQ1 versus EQ3 and EQ2 versus EQ4. Moreover, the accuracy is much better than that of arousal classification. The $\mathcal{CCR}$ of subject-independent classification was not comparable to that obtained for subject-dependent classification. As shown in Fig. 11, merging the features of all subjects does not refine the discriminating information related to the emotions but, rather, leads to scattered class boundaries.

We also tried to differentiate the emotions based on the two axes, arousal and valence, in the 2D emotion model. The samples of four emotions were divided into groups of negative valence (EQ2+EQ3) and positive valence (EQ1+EQ4) and into groups of high arousal (EQ1+EQ2) and low arousal (EQ3+EQ4). By using the same methods, we then performed a two-class classification of the divided samples for arousal and valence separately. Table 2 shows the results of arousal and valence classification. It turned out that emotion-relevant ANS specificity can be observed

TABLE 1
Recognition Results in Rates ($error\ 0.00 = \mathcal{CCR}\ 100$ percent) Achieved by Using pLDA with SBS and Leave-One-Out Cross Validation

Subject A ($\mathcal{CCR}$ % = ***81%***)

	EQ1	EQ2	EQ3	EQ4	*total**	*error*
EQ1	***22***	4	1	3	*30*	*0.27*
EQ2	3	***26***	1	0	*30*	*0.13*
EQ3	1	2	***23***	4	*30*	*0.23*
EQ4	3	0	1	***26***	*30*	*0.13*

Subject B ($\mathcal{CCR}$ % = ***91%***)

	EQ1	EQ2	EQ3	EQ4	*total**	*error*
EQ1	***27***	3	0	0	*30*	*0.10*
EQ2	3	***25***	1	1	*30*	*0.17*
EQ3	0	2	***28***	0	*30*	*0.07*
EQ4	0	1	0	***29***	*30*	*0.03*

Subject C ($\mathcal{CCR}$ % = ***89%***)

	EQ1	EQ2	EQ3	EQ4	*total**	*error*
EQ1	***28***	0	2	0	*30*	*0.07*
EQ2	0	***30***	0	0	*30*	*0.00*
EQ3	0	0	***24***	6	*30*	*0.20*
EQ4	0	0	5	***25***	*30*	*0.17*

All: Subject-independent ($\mathcal{CCR}$ % = ***65%***)

	EQ1	EQ2	EQ3	EQ4	*total**	*error*
EQ1	***62***	9	8	11	*90*	*0.31*
EQ2	15	***57***	13	5	*90*	*0.37*
EQ3	9	6	***58***	17	*90*	*0.36*
EQ4	8	5	21	***56***	*90*	*0.38*

*: Actual total # of samples

Number of samples: 120 for each subject and 360 for all. Subject A ($\mathcal{CCR}\% = 81\%$). *Subject B* ($\mathcal{CCR}\% = 91\%$). *Subject C* ($\mathcal{CCR}\% = 89\%$). *All: Subject-independent* ($\mathcal{CCR}\% = 65\%$).

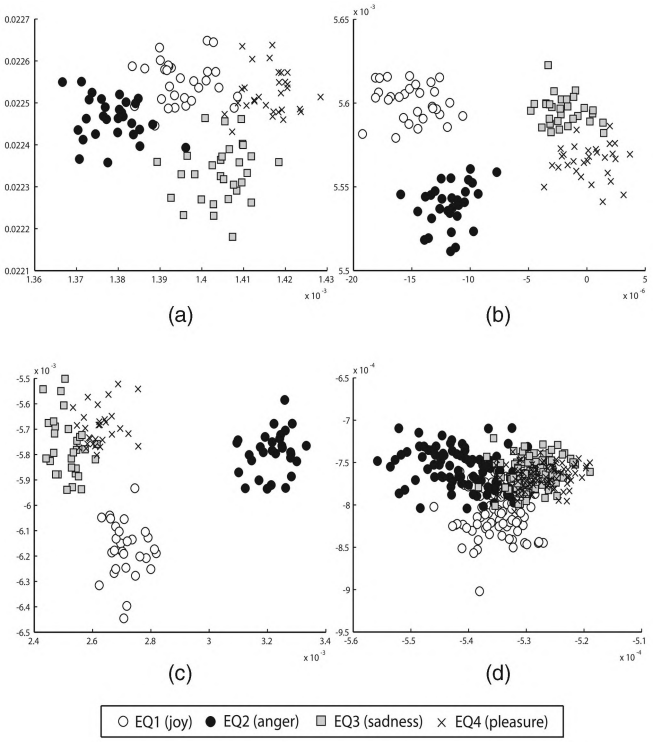

Fig. 11. Comparison of feature distributions of subject-dependent and subject-independent cases. (a) Subject A. (b) Subject B. (c) Subject C. (d) Subject independent.

more conspicuously in the arousal axis regardless of subject-dependent or independent cases. The classification of arousal achieved an acceptable $\mathcal{CCR}$ of 97-99 percent for the subject-dependent recognition and 89 percent for the subject-independent recognition, while the results for valence were 88-94 percent and 77 percent, respectively.

5.2 Finding the Best Emotion-Relevant ANS Features

In most of the literature dealing with emotion-relevant ANS specificity, a tendency analysis of physiological changes has been used to correlate ANS activity with certain emotional states, for example, EQ1 with increased heart rate or anxiety with increased SC. Even for multiclass classification problems, however, such a direction analysis of physiological changes is not sufficient to capture accompanying multimodal ANS reactions that are cross-correlated with each other when using multichannel biosensors. Therefore, we tried to first identify the significant features for each classification problem and thereby to investigate the class-relevant feature domain and interrelation between the features for a certain emotion.

In Table 3, the best emotion-relevant features, which we determined by ranking the features selected for all subjects (including Subject All) in each classification problem, are listed in detail by specifying their values and domains. One interesting result is that each classification problem respectively links together with a certain feature domain. The features obtained from the time/frequency analysis of the HRV time series are decisive for the classification of arousal and for the classification of the four emotions, while the features from the MSE domain of ECG signals are a predominant factor for correct valence differentiation. More

TABLE 2
Results of Arousal and Valence Recognition

Subject A

	high	low	*error*		pos	neg	*error*
high(60*)	58	2	*0.03*	pos(60)	56	4	*0.07*
low(60)	2	58	*0.03*	neg(60)	8	52	*0.15*
$\mathcal{CCR}$(%) :	*97 %*			$\mathcal{CCR}$(%) :	*90 %*		

Subject B

	high	low	*error*		pos	neg	*error*
high(60)	58	2	*0.03*	pos(60)	55	5	*0.08*
low(60)	0	60	*0.00*	neg(60)	2	58	*0.03*
$\mathcal{CCR}$ (%) :	*98 %*			$\mathcal{CCR}$(%) :	*94 %*		

Subject C

	high	low	*error*		pos	neg	*error*
high(60)	60	0	*0.00*	pos(60)	51	9	*0.15*
low(60)	1	59	*0.02*	neg(60)	5	55	*0.08*
$\mathcal{CCR}$ (%) :	*99 %*			$\mathcal{CCR}$(%) :	*88 %*		

All: Subject-independent

	high	low	*error*		pos	neg	*error*
high(180)	155	25	*0.14*	pos(180)	143	37	*0.21*
low(180)	16	164	*0.09*	neg(180)	45	135	*0.25*
$\mathcal{CCR}$(%) :	*89 %*			$\mathcal{CCR}$(%) :	*77 %*		

*: Actual total # of samples

particularly, the mutually sympathizing correlate between HRV and BRV (first proposed in this paper) has been clearly observed in all of the classification problems by the features from their time/frequency analysis and Poincaré domain, *_PoincareHRV* and *_PoincareBRV*. This reveals a manifest cross correlation between RSP and cardiac activity with respect to the emotional state. This is one of the most important findings for future work. In fact, in biomedicine,[3] it is commonly accepted that the respiratory mechanism mediates HF components of HRV, but its specific role in affective ANS reactions has so far not been satisfactorily explained. When inhaling, the vagus nerve is impeded and the heart rate begins to increase, whereas this pattern is reversed when exhaling, that is, the activation of the vagus nerve typically leads to a reduction in heart rate, blood pressure, or both. Apart from its influence on the heart rate, the vagus nerve is also responsible for sweating, several muscle movements in the mouth, and even for speech. It means that most physiological channels we used are innately correlated with each other and respond together as a chain reaction to emotional stimulation. For example, when the parasympathetic nerves overcompensate, a strong response from the sympathetic nervous system innervating the sinoatrial node, which occurs in cases of extreme stress or fear, the reduction in heart rate and blood pressure becomes proportionally faster to the intensity of the emotion.

Our feature analysis proves that the correlation between the heart rate and RSP is obviously captured by the features from the HRV power spectrum (*_HRVspec*), the fast/long-term HRV/BRV analysis using the Poincaré method, and the multiscale variance analysis of HRV/BRV (*_MSE*). It also demonstrates that the peaks of the HF range in the HR subband spectrum (*_SubSpectra*) provide information about

3. The influence of breathing on the flow of the sympathetic and vagus impulses to the sinoatrial node causes the so-called respiratory sinus arrhythmia (RSA). The degree of fluctuation in heart rate is also significantly controlled by regular impulses from the baroreceptors in the aorta and carotid arteries.

TABLE 3
Best Emotion-Relevant Features Extracted from Four-Channel Physiological Signals

Classes	Best Emotion-relevant Features (Ch_value_domain, $\mathcal{C}$: ECG, $\mathcal{R}$: RSP, $\mathcal{S}$: SC, $\mathcal{M}$: EMG)
Arousal	$\mathcal{C}$_std(diff)_HRVtime, $\mathcal{C}$_**sd2_PoincareHRV**, $\mathcal{C}$_powerLow_HRVspec, $\mathcal{R}$_**meanEnergy_SubSpectra**, $\mathcal{R}$_**sd2_PoincareBRV**, $\mathcal{R}$_**mean_MSE**, $\mathcal{S}$_mean_RawLowpassed, $\mathcal{S}$_std_RawLowpassed, $\mathcal{M}$_#occurrenceRatio_RawLowpassed, $\mathcal{M}$_mean_RawNormed
Valence	$\mathcal{C}$_**sd2_PoincareHRV**, $\mathcal{C}$_meanEnergy_SubSpectra, $\mathcal{C}$_ratioLH_HRVspec, $\mathcal{C}$_mean_MSE, $\mathcal{C}$_mean(diff)_MSE, $\mathcal{R}$_**meanEnergy_SubSpectra**, $\mathcal{R}$_mean(diff)_SubSpectra, $\mathcal{R}$_sd1_PoincareBRV, $\mathcal{R}$_**sd2_PoincareBRV**, $\mathcal{R}$_**mean_MSE**, $\mathcal{S}$_mean(diff)_RawNormed, $\mathcal{M}$_mean(diff)_RawNormed
Four Emotions	$\mathcal{C}$_mean_HRVtime, $\mathcal{C}$_std_HRVtime, $\mathcal{C}$_std(diff)_HRVtime, $\mathcal{C}$_mean(diff)_MSE, $\mathcal{C}$_mean_MSE, $\mathcal{C}$_mean_SSE, $\mathcal{C}$_**sd2_PoincareHRV**, $\mathcal{C}$_mean_SubSpectra, $\mathcal{R}$_**meanEnergy_SubSpectra**, $\mathcal{R}$_mean_SSE, $\mathcal{R}$_mean_BRVtime, $\mathcal{R}$_sd1_PoincareBRV, $\mathcal{R}$_**sd2_PoincareBRV**, $\mathcal{R}$_**mean_MSE**, $\mathcal{R}$_power_BRVspec, $\mathcal{S}$_std_RawLowpassed, $\mathcal{S}$_mean(diff)_RawNormed, $\mathcal{S}$_mean(diff(diff))_RawLowpassed, $\mathcal{S}$_mean_RawNormed, $\mathcal{S}$_#occurrence_RawLowpassed, $\mathcal{M}$_mean(diff)_RawNormed

: overall selected features are printed in **bold**

Arousal classes: EQ1+EQ2 versus EQ3+EQ4. Valence classes: EQ1+EQ4 versus EQ2+EQ3. Four classes: EQ1/EQ2/EQ3/EQ4.

how the sinoatrial node responds to vagal activity at certain RSP frequencies.

Table 4 shows the number of selected features using the SBS method for the three classification problems: arousal, valence, and the four emotional states. For the arousal classification, relatively few features were used, but they achieved higher recognition accuracy compared to the other class problems. If we take a look at the ratio of the number of selected features to the total feature number of each channel, it is obvious that the SC and EMG activities reflected in both the *_RawLowpassed* and *_RawNormed* domains (see Table 3) are more significant for arousal classification than the other channels. This also supports the experimental conclusions of previous research according to which the SCR is linearly correlated with the intensity of arousal. On the other hand, we observe a remarkable increase in the number of ECG and RSP features for the case of valence classification.

TABLE 4
Number of Selected Features for Each Class Problem

Arousal Classification

	$ECG_{(53)}$	$RSP_{(37)}$	$SC_{(10)}$	$EMG_{(10)}$	$\sum_{(110)}$
Subject A	18	15	3	2	38
Subject B	6	8	6	7	27
Subject C	7	9	7	2	25
All	14	13	3	5	35
Total	*45*	*45*	*19*	*16*	*125*

Valence Classification

	$ECG_{(53)}$	$RSP_{(37)}$	$SC_{(10)}$	$EMG_{(10)}$	$\sum_{(110)}$
Subject A	21	23	2	5	51
Subject B	16	11	3	3	33
Subject C	20	14	3	6	43
All	14	15	3	2	34
Total	*71*	*63*	*11*	*16*	*161*

4-Class Classification

	$ECG_{(53)}$	$RSP_{(37)}$	$SC_{(10)}$	$EMG_{(10)}$	$\sum_{(110)}$
Subject A	15	15	2	3	35
Subject B	18	12	6	5	41
Subject C	20	13	7	6	46
All	24	16	7	3	50
Total	*77*	*56*	*22*	*17*	*172*

Arousal classification. Valence classification. 4-class classification.

5.3 Emotion-Specific Multilevel Dichotomous Classification

Most common classifiers are best suited to handling two-class problems. The pLDA we used is no exception to this and assumes that the covariance matrices of each class are the same or at least close to each other for multiclass ($c > 2$) classification. Consequently, the performance of pLDA in multiclass classification could be suboptimal, depending on the difference between the covariance matrices of each class. In our work, we actually used the averaged covariance to directly solve the multiclass problem using a single pLDA classifier. One straightforward way to handle a multiclass problem by using binary classifiers is to decompose the multiple categories into a set of complementary two-class problems. Various approaches to do this have been proposed [45], [46]. The one-against-all decomposition, for example, consists of subsets grouped by opposing each class to all of the others and c binary classifiers are trained from the whole set of training samples. Alternatively, each class can be opposed to each of the other ones (one-against-one or pairwise decomposition). In this case, $c(c-1)/2$ pairwise classifiers are trained from training samples corresponding to two classes. Some methods for classifier combination exploiting the complementarity of multiple classifiers have also been proposed [47].

By taking advantage of supervised classification (where we know in advance which emotion types have to be recognized), we developed an EMDC scheme. This scheme exploits the property of the dichotomous categorization in the 2D emotion model and the fact that arousal classification yields a higher $\mathcal{CCR}$ than valence classification or direct multiclass classification. This proves true in almost all previous works and according to our results as well. Fig. 12 illustrates the EMDC scheme and provides an example of the dyadic decomposition for the eight-class problem in Fig. 1a.

First, the entire training patterns are grouped into two opposing *"superclasses"* (on the basis of valence or arousal): $\bar{C}$ consisting of all patterns in some subset of the

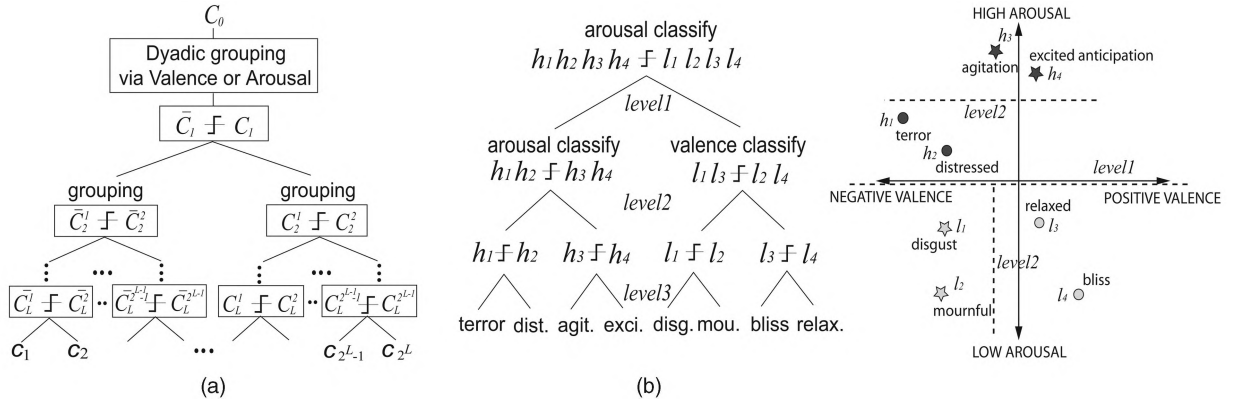

Fig. 12. Framework of emotion-specific multilevel dichotomous classification (EMDC). (a) Diagram of the decomposition process. (b) Decomposition example for an eight-class problem.

TABLE 5
Results Using the EMDC Scheme with the Best Features

Subject A (CCR % = ***94%***, *113/120*)

	EQ1 & EQ2			EQ3 & EQ4
EQ1 & EQ2		**58**		2
		EQ1	EQ2	
	EQ1	**27**	1	
	EQ2	1	**29**	
EQ3 & EQ4	2			**58**
				EQ3 / EQ4
			EQ3	**29** / 0
			EQ4	1 / **28**

Subject C (CCR % = ***94%***, *113/120*)

	EQ1 & EQ2			EQ3 & EQ4
EQ1 & EQ2		**60**		0
		EQ1	EQ2	
	EQ1	**30**	0	
	EQ2	0	**30**	
EQ3 & EQ4	1			**59**
				EQ3 / EQ4
			EQ3	**27** / 2
			EQ4	4 / **26**

Subject B (CCR % = ***98%***, *117/120*)

	EQ1 & EQ2			EQ3 & EQ4
EQ1 & EQ2		**60**		0
		EQ1	EQ2	
	EQ1	**30**	0	
	EQ2	0	**30**	
EQ3 & EQ4	1			**59**
				EQ3 / EQ4
			EQ3	**29** / 1
			EQ4	1 / **28**

Subject All (CCR % = ***70%***, *251/360*)

	EQ1 & EQ2			EQ3 & EQ4
EQ1 & EQ2		**155**		25
		EQ1	EQ2	
	EQ1	**62**	13	
	EQ2	15	**65**	
EQ3 & EQ4	16			**164**
				EQ3 / EQ4
			EQ3	**64** / 19
			EQ4	21 / **60**

class categories and C consisting of all remaining patterns, that is, $\bar{C} \cap C = \{\}$. This dyadic decomposition using one of the two axes is serially performed until one subset contains only two classes. The grouping axis can be different for each dichotomous level. Then, multiple binary classifiers for each level are trained from the corresponding dyadic patterns. Therefore, the EMDC scheme is obviously emotion specific and effective for a 2D emotion model. Note that the performance of the EMDC scheme is limited by the maximum CCR of the first-level classification and makes sense only if the CCR for one of the two superclasses is higher than that for direct multiclass classification (theoretically, this always holds true; see Table 2 for our case). Because we used four emotion classes in our experiment, we needed a two-level classification based on arousal and valence grouping for both superclasses in parallel.

Table 5 shows the dichotomous contingency table of recognition results by using the novel EMDC scheme. The best feature sets shown in Table 3 are used for the binary classification at each level. As expected, the CCRs significantly improved for all class problems. For the classification of four emotions, we obtained an average CCR of 95 percent for subject-dependent and 70 percent for subject-independent classification. Compared to the results obtained for pLDA, the EMDC scheme achieved an overall CCR improvement of about 5-13 percent in each class problem (see Table 6).

6 Discussion

We achieved an overall CCR of 95 percent, which is more than three times higher than chance probability, for four emotional states from three subjects. This should be sufficient to support the view that emotions, either produced or perceived while listening to music, exist and are accompanied by physiological differences in both the

TABLE 6
CCR Comparison between pLDA and EMDC

	Subj. A	Subj. B	Subj. C	Subj. All	*Average (ABC)*
pLDA	81%	91%	89%	65%	*87%*
EMDC	**94%**	**98%**	**94%**	**70%**	***95%***

arousal and valence dimensions such that they can eventually be recognized by the machine. At the same time, however, some issues remain in relation to the processing stages of our recognition system.

Recording physiological changes using biosensors is still invasive since the subjects, for example, have to be in physical contact with adhesive electrodes. Furthermore, most biosensors using such electrodes are very susceptible to motion artifacts, which we could observe in almost all signals of our data set. For practical HCI applications, it is therefore necessary to develop noninvasive biosensors, preferably with built-in denoising filters in wirelessly miniaturized form. We expect that today's nanotechnology will help design such hardware soon. This would then not only improve the signal quality and the usability of the technology but also reduce computational costs in the preprocessing stage.

Our analysis results based on the best emotion-relevant features are incontrovertibly useful findings, for example, the consistent tendency of the feature contents to valence and arousal differentiation separately and the proven efficiency of new feature domains that are first considered in this paper. We should, however, note that the effectiveness of the best features might not be universally guaranteed for other data sets or classifiers. First, only three subjects might not be sufficient to generalize the features. Second, the SBS and most algorithms for feature selection use a criterion based on a specific classifier and are therefore effective only if the classifier used is known in advance. In addition, such sequential algorithms may lead to suboptimal subsets due to their unidirectional property, that is, once a feature is added or removed, this action can never be reversed.

By dividing given patterns using the arousal and valence axes in the 2D emotion model, we proposed the EMDC scheme, which contributed to a significant improvement in the recognition results. The scheme may, however, still be adjusted in several ways. For instance, since it needs multiple classifiers to be trained for each level, the combination of different classifiers seems to be feasible. By taking advantage of the fact that EMDC enables us to view the classification results of each level in a multiresolution aspect (see Table 5), the scheme could be more sophisticatedly designed thanks to the parametric refining of each binary classifier depending on the level.

The reason for the great disparity of $\mathcal{CCR}$ between subject-dependent classification and independent classification can be explained in many different ways indeed. We mention that one of the main factors in the difficulty of subject-independent classification is the intricate variety of nonemotional individual contexts among the subjects, rather than an individual ANS specificity in emotion. A naive idea for improving the performance of the user-independent system for practical applications would be to first identify the user, prior to starting the recognition process, and then to classify a user's emotion in a user-dependent way. Of course, this is feasible only if the number of users is finite and the users are known to the system or if the system can cumulatively collect the data of each user in a learning phase. Although this goes beyond the subject of this paper, we tried to identify the subjects in our experiment by using the same feature set and the pLDA classifier that we used for the emotion recognition task. Surprisingly enough, we obtained perfect identification accuracy with a $\mathcal{CCR}$ of 100 percent from all emotion-dependent identifications, that is, subject identification for each emotion, EQ1, EQ2, EQ3, and EQ4, respectively, and 99.4 percent from the emotion-independent identification using all the data sets taken together. Illustrated below are the detailed results with the confusion matrix for the latter case and the person-specific features extracted by ranking overlapped features in each identification problem.

Person Identification ($\mathcal{CCR}$ % = ***99.4%***)

	Subject A	Subject B	Subject C	*total**
Subject A	***119***	0	1	*120*
Subject B	0	***120***	0	*120*
Subject C	1	0	***119***	*120*

*: Actual total # of samples

Person-specific Features
$\mathcal{C}$_meanEnergy_SubSpectra, $\mathcal{C}$_meanHR_HRVtime, $\mathcal{C}$_powerLow_HRVspec
$\mathcal{C}$_mean_MSE, $\mathcal{C}$_mean_SSE, $\mathcal{R}$_meanBR_BRVtime, $\mathcal{R}$_ratioLH_BRVspec
$\mathcal{R}$_meanEnergy_SubSpectra , $\mathcal{R}$_mean_MSE, $\mathcal{R}$_sd2_PoincareBRV
$\mathcal{S}$_mean_RawNormed, $\mathcal{S}$_mean_RawLowpassed, $\mathcal{S}$_std_RawLowpassed
$\mathcal{S}$_#occur_RawVLowpassed, $\mathcal{S}$_#occRatio_RawVLowpassed
$\mathcal{S}$_#occur_RawLowpassed, $\mathcal{M}$_mean_RawNormed, $\mathcal{M}$_#occur_RawLowpassed
$\mathcal{M}$_mean(diff(diff))_RawLowpassed, $\mathcal{M}$_#occRatio_RawVLowpassed
:(Ch_value_domain, $\mathcal{C}$: ECG, $\mathcal{R}$: RSP, $\mathcal{S}$: SC, $\mathcal{M}$: EMG)

More interestingly, it is likely that the accuracy of person identification is inversely proportional to the accuracy of subject-independent emotion classification when using the same features for both systems.

7 Conclusion

In this paper, we dealt with all the essential stages of an automatic emotion recognition system using multichannel physiological measures, from data collection to the classification process, and analyzed the results from each stage of the system. For four emotional states of three subjects, we achieved an average recognition accuracy of 95 percent, which connotes more than a prima facie evidence that there are some ANS differences among emotions. Moreover, the accuracy is higher than that in the previous works reviewed in this paper when considering the different experimental settings in the works, such as the number of target classes, the number of subjects, the naturalness of the data set, etc.

To acquire a naturalistic data set from a reliable experiment, we designed a musical induction method that was not based on any laboratory setting or any deliberate instructions for evoking certain emotions but was based instead on the voluntary participation of subjects who collected the musical induction materials according to target emotions and determined the recording schedule themselves. Hence, a recorded data set must not necessarily be annotated by a labeler or through self-judgment.

A wide range of physiological features from various analysis domains, including time, frequency, entropy, geometric analysis, subband spectra, multiscale entropy, and HRV/BRV, were proposed to search for the best emotion-relevant features and to correlate them with emotional states. The selected best features were described in detail and their effectiveness was proven by classification results. We found that SC and EMG are linearly correlated with arousal change in emotional ANS activities and that the features in ECG and RSP are dominant for valence

differentiation. Particularly, the HRV/BRV analysis revealed the cross correlation between the heart rate and RSP. The classification of the features was performed by using the SBS and the pLDA classifier for arousal, valence, and four emotion classes and achieved an average recognition accuracy of 98 percent, 91 percent, and 87 percent, respectively.

In addition, in order to further improve the accuracy of the four emotion classes, we developed a new EMDC scheme. With this scheme, we actually obtained a maximum of 13 percent improved accuracy for all subjects. However, the recognition accuracy of subject-independent classification (70 percent for four classes) was not comparable with the subject-dependent case (95 percent for four classes). The main reason can probably be ascribed to the intricate difference of nonemotional individual contexts between the subjects rather than to any inconsistency of ANS differences among emotions. To deal with the difficulty of subject-independent recognition, we briefly discussed an extended recognition system where we identified the user prior to starting the recognition process and then classified the user's emotions in a user-dependent manner. Supporting this simple idea, we showed identification results achieving an almost perfect accuracy of 99.4 percent; this was obtained by using the same features we had used for emotion recognition.

One of the most challenging issues in the near future will be to explore multimodal analysis for emotion recognition. We humans use several modalities jointly to interpret emotional states, since emotion affects almost all modes—audiovisual (facial expression, voice, gesture, posture, etc.), physiological (RSP, skin temperature, etc.), and contextual (goal, preference, environment, social situation, etc.) states in human communication. In the recent literature, findings concerning emotion recognition by combining multiple modalities have been reported, mostly by fusing features extracted from audiovisual modalities such as facial expressions and speech. However, we note that combining multiple modalities by equally weighting them does not always guarantee improved accuracy. The more crucial issue is how to *complementarily* combine the additional modalities. An essential step toward a human-like analysis and finer resolution of recognizable emotion classes would therefore be to find the innate priority among the modalities to be preferred for each emotional state. Then, an ambitious undertaking might be to decompose an emotion recognition problem into several refining processes using additional modalities, for example, arousal recognition through physiological channels, valence recognition by using audiovisual channels, and then resolving of subtle uncertainties between adjacent emotion classes, or even predicting the "stance" in a 3D emotion model by cumulative analysis of a user's context information. In this sense, the physiological channel can be considered as a "baseline channel" in designing a multimodal emotion recognition system since it provides several advantages over other external channels and an acceptable recognition accuracy, as we have presented in this paper.

Acknowledgments

This research was partially supported by the European Commission (HUMAINE NoE; FP6 IST-507422).

References

[1] W. James, *The Principles of Psychology*. Holt, 1890.
[2] W.B. Cannon, "The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory," *Am. J. Psychology,* vol. 39, pp. 106-127, 1927.
[3] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J.G. Taylor, "Emotion Recognition in Human-Computer Interaction," *IEEE Signal Processing Magazine,* vol. 18, pp. 32-80, 2001.
[4] P. Ekman, "The Argument and Evidence about Universals in Facial Expressions of Emotion," *Handbook of Social Psychophysiology,* pp. 143-164, John Wiley & Sons, 1989.
[5] R.W. Levenson, P. Ekman, P. Heider, and W.V. Friesen, "Emotion and Autonomic Nervous System Activity in the Minangkabau of West Sumatra," *J. Personality and Social Psychology,* vol. 62, pp. 972-988, 1992.
[6] J.T. Cacioppo, D.J. Klein, G.G. Bemston, and E. Hatfield, "The Psychophysiology of Emotion," *Handbook of Emotions,* M. Lewis and J. Haviland, eds., pp. 119-142, Guilford Press, 1993.
[7] R.W. Levenson, P. Ekman, and W.V. Friesen, "Voluntary Facial Action Generates Emotion-Specific Autonomic Nervous System Activity," *Psychophysiology,* vol. 27, pp. 363-384, 1990.
[8] G. Stemmler, M. Heldmann, C.A. Pauls, and T. Scherer, "Constraints for Emotion Specificity in Fear and Anger: The Context Counts," *Psychophysiology,* vol. 38, pp. 275-291, 2001.
[9] S. Schachter and J.E. Singer, "Cognitive, Social, and Physiological Determinants of Emotional State," *Psychological Rev.,* vol. 69, pp. 379-399, 1962.
[10] G. Stemmler, "The Autonomic Differentiation of Emotions Revisited: Convergent and Discriminant Validation," *Psychophysiology,* vol. 26, pp. 617-632, 1989.
[11] M.P. Kastner and R.G. Crowder, "Perception of the Major/Minor Distinction: IV. Emotional Connotations in Young Children," *Music Perception,* vol. 8, pp. 189-201, 1990.
[12] P. Kivy, *Sound Sentiment: An Essay on the Musical Emotions.* Temple Univ. Press, 1989.
[13] C.L. Krumhansl, "An Exploratory Study of Musical Emotions and Psychophysiology," *Canadian J. Experimental Psychology,* vol. 51, pp. 336-352, 1997.
[14] L.B. Meyer, *Emotion and Meaning in Music.* Univ. of Chicago Press, 1956.
[15] D. Vaitl, W. Vehrs, and S. Sternagel, "Prompts-Leitmotif-Emotion: Play It Again, Richard Wagner," *The Structure of Emotion: Psychophysiological, Cognitive, and Clinical Aspects,* pp. 169-189, Hogrefe & Huber, 1993.
[16] W.B. Davis and M.H. Thaut, "The Influence of Preferred Relaxing Music on Measures of State Anxiety, Relaxation, and Physiological Responses," *J. Music Therapy,* vol. 26, no. 4, pp. 168-187, 1989.
[17] C.E. Guzzetta, "Effects of Relaxation and Music Therapy on Patients in a Coronary Care Unit with Presumptive Acute Myocardial Infarction," *Heart and Lung: J. Critical Care,* vol. 18, no. 6, pp. 609-616, 1989.
[18] J. Healey and R.W. Picard, "Digital Processing of Affective Signals," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing,* pp. 3749-3752, 1998.
[19] R. Picard, E. Vyzas, and J. Healy, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 10, pp. 1175-1191, Oct. 2001.
[20] F. Nasoz, K. Alvarez, C. Lisetti, and N. Finkelstein, "Emotion Recognition from Physiological Signals for Presence Technologies," *Int'l J. Cognition, Technology, and Work,* special issue on presence, vol. 6, no. 1, 2003.
[21] J.J. Gross and R.W. Levenson, "Emotion Elicitation Using Films," *Cognition and Emotion,* vol. 9, pp. 87-108, 1995.
[22] A. Haag, S. Goronzy, P. Schaich, and J. Williams, "Emotion Recognition Using Bio-Sensors: First Steps Towards an Automatic System," *Proc. Ninth Int'l Conf. Reliable Software Technologies,* pp. 36-48, 2004.
[23] Center for the Study of Emotion and Attention (CSEA-NIMH), *The International Affective Picture System: Digitized Photographs,* Center for Research in Psychophysiology, Univ. of Florida, 1995.
[24] K.H. Kim, S.W. Bang, and S.R. Kim, "Emotion Recognition System Using Short-Term Monitoring of Physiological Signals," *Medical & Biological Eng. and Computing,* vol. 42, pp. 419-427, 2004.
[25] P. Lang, "The Emotion Probe: Studies of Motivation and Attention," *Am. Psychologist,* vol. 50, no. 5, pp. 372-385, 1995.

[26] J. Tooby and L. Cosmides, "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments," *Ethology and Sociobiology,* vol. 11, pp. 375-424, 1990.
[27] H. Scholsberg, "Three Dimensions of Emotion," *Psychological Rev.,* vol. 61, pp. 81-88, 1954.
[28] B.L. Fredricson and R.W. Levenson, "Positive Emotions Speed Recovery from the Cardiovascular Sequelae of Negative Emotions," *Cognition and Emotion,* vol. 12, no. 2, pp. 191-220, 1998.
[29] J.E. LeDoux, *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction,* pp. 339-351. Wiley-Liss, 1992.
[30] D.M. Sloan, "Emotion Regulation in Action: Emotional Reactivity in Experiential Avoidance," *Behavior Research and Therapy,* vol. 4, pp. 1257-1270, 2004.
[31] B. Melin and U. Lundberg, "A Biopsychosocial Approach to Work-Stress and Musculoskeletal Disorders," *J. Psychophysiology,* vol. 11, no. 3, pp. 238-247, 1997.
[32] H.G. McCurdy, "Consciousness and the Galvanometer," *Psychological Rev.,* vol. 57, pp. 322-327, 1950.
[33] A. Jain, R. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 1, pp. 4-37, Jan. 2000.
[34] R.D. King, C. Feng, and A. Shutherland, "StatLog: Comparison of Classification Algorithms on Large Real-World Problems," *Applied Artificial Intelligence,* vol. 9, no. 3, pp. 259-287, May/June 1995.
[35] M.B. Mcllory and M.D. Cheithin, *Clinical Cardiology,* fifth ed. VLANGE Medical Book, 1990.
[36] J. Pan and W. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomedical Eng.,* vol. 32, no. 3, pp. 230-323, 1985.
[37] P.W. Kamen, H. Krum, and A.M. Tonkin, "Poincare Plot of Heart Rate Variability Allows Quantitative Display of Parasympathetic Nervous Activity," *Clinical Science,* vol. 91, pp. 201-208, 1996.
[38] J. Richmann and J. Moorman, "Physiological Time Series Analysis Using Approximate Entropy and Sample Entropy," *Am. J. Physiology—Heart and Circulatory Physiology,* vol. 278, p. H2039, 2000.
[39] M. Costa, A.L. Goldberger, and C.-K. Peng, "Multiscale Entropy Analysis of Biological Signals," *Physical Rev. E,* vol. 71, no. 021906, 2005.
[40] R. Thuraisingham and G. Gottwald, "On Multiscale Entropy Analysis for Physiological Data," *Physica A,* 2005.
[41] A. Malliani, "The Pattern of Sympathovagal Balance Explored in the Frequency Domain," *News in Physiological Science,* vol. 14, pp. 111-117, 1999.
[42] A. Jain and D. Zongker, "Feature Selection: Evaluation, Application, and Small Sample Performance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 2, pp. 153-163, Feb. 1997.
[43] F.J. Ferri, P. Pudil, M. Hatef, and J. Kittler, "Comparative Study of Techniques for Large-Scale Feature Selection," *Pattern Recognition in Practice IV: Multiple Paradigms, Comparative Studies, and Hybrid Systems,* S.E. Gelsema and L.N. Kanal, eds., pp. 403-413, 1994.
[44] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 8, pp. 831-836, Aug. 1996.
[45] T. Hastie and R. Tibshirani, "Classification by Pairwise Coupling," *The Annals of Statistics,* vol. 26, no. 1, pp. 451-471, 1998.
[46] J. Friedman, "Another Approach to Polychotomous Classification," technical report, Dept. of Statistics, Stanford Univ., 1996.
[47] L. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms.* John Wiley & Sons, 2004.

Jonghwa Kim received the BS and MS degrees in electronic engineering from the Kyungwon University, Korea, in 1992 and 1994, respectively, and the diploma in sound engineering and the PhD degree in communication engineering from the Technical University of Berlin in 1999 and 2003, respectively. In 2002, he jointed the Institute of Computer Science at the University of Augsburg, Germany, as a research associate, where, since 2003, he has been an assistant professor in the Department of Applied Computer Science. He is currently involved in the European IST projects HUMAINE, CALLAS, and METABO as the leader of the emotion recognition team. His current research interests include intelligent signal processing, emotion recognition, pattern classification, biomedical signal analysis, multisource data fusion, and affective human-robot interface. He has served on many technical conference committees and as a reviewer for the IEEE Computer Society. He is a member of the IEEE and the IEEE Computer Society.

Elisabeth André received the diploma and PhD degree in computer science from the Saarland University, Saarbrücken, Germany, in 1988 and 1995, respectively. In 1988, she joined the German Research Center for Artificial Intelligence (DFKI GmbH), where she was promoted to senior researcher in 1995 and to principal researcher in 1999. Since 2001, she has been a full professor of computer science at the University of Augsburg, Germany, and the chair of the Laboratory for Multimedia Concepts and Applications. Her current research interests include affective computing, intelligent multimedia interfaces, embodied conversational agents, and the integration of vision and natural language. She is the chair of the ACL Special Interest Group on Multimedia Language Processing (SIGMEDIA) and the area editor for intelligent user interfaces of the *Electronic Transactions of Artificial Intelligence*. She also serves on the editorial boards of *Artificial Intelligence Communications*, *Cognitive Processing (International Quarterly of Cognitive Science)*, *Universal Access to the Information Society*, and *Autonomous Agents and Multi-Agent Systems*.