

# MULTI-CHANNEL BIOSIGNAL ANALYSIS FOR AUTOMATIC EMOTION RECOGNITION

Jonghwa Kim and Elisabeth André

*Institute of Computer Science, University of Augsburg, Eichleitnerstr. 30, D-86159 Augsburg, Germany*

**Keywords:** Biosignal, emotion recognition, physiological measures, skin conductance, electrocardiogram, electromyogram, respiration, affective computing, human-computer interaction, musical emotion, autonomic nervous system, arousal, valence, feature extraction, pattern recognition.

**Abstract:** This paper investigates the potential of physiological signals as a reliable channel for automatic recognition of user's emotional state. For the emotion recognition, little attention has been paid so far to physiological signals compared to audio-visual emotion channels such as facial expression or speech. All essential stages of automatic recognition system using biosignals are discussed, from recording physiological dataset up to feature-based multiclass classification. Four-channel biosensors are 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 search 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 emotion recognition results.

## 1 INTRODUCTION

In human communication, expression and understanding of emotions facilitate to complete the mutual sympathy. To approach it in human-machine interaction, we need to equip machines with the means to interpret and understand human emotions without input of user's translated intention. Hence, one of the most important prerequisites to realize such an advanced user interface is a reliable emotion recognition system which guarantees acceptable recognition accuracy, robustness against any artifacts, and adaptability to practical applications. It is about to model, analyze, process, train, and classify 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 finding emotional cues from various physiological measures.

Recently many works on engineering approaches to automatic emotion recognition have been reported. For an overview we refer to (Cowie et al., 2001). Particularly, most efforts have been taken to recognize human emotions using audiovisual channels of emotion expression, facial expression, speech, and gesture. Relatively little attention, however, has been paid so far to using physiological measures. Rea-

sons are some significant limitations resulting from 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 subtle physiological patterns onto specific emotional states. As an emotion is a function of time, context, space, culture, and person, physiological patterns may also widely differ from user to user and from situation to situation.

In this paper, we treat all essential stages of automatic emotion recognition system using physiological measures, from data collection up to classification of four typical emotions (joy, anger, sadness, pleasure) using four-channel biosignals. The work in this paper is novel in trying to recognize naturally induced musical emotions using physiological changes, in acquiring the physiological dataset through everyday life recording over many weeks from multiple subjects, in finding emotion-relevant ANS (autonomic nervous system) 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 try 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.

## 2 RELATED RESEARCH

A significant amount of work has been conducted by Picard and colleagues at MIT Lab who showed that certain affective states may be recognized by using physiological data including heart rate, skin conductivity, temperature, muscle activity and respiration velocity (Healey and Picard, 1998). They used personalized imagery to elicit target emotions from a single subject who had two years' experience in acting, and achieved overall 81% recognition accuracy in eight emotions by using hybrid linear discriminant classification (Picard et al., 2001). Nasoz et al. (Nasoz et al., 2003) used movie clips based on the study by Gross and Levenson (Gross and Levenson, 1995) for eliciting target emotions from 29 subjects and achieved best emotion classification accuracy of 83% through the Marquardt Backpropagation algorithm (MBP). More recently, interesting user-independent emotion recognition system is reported by Kim et al. (Kim et al., 2004). They developed a set of recording protocols using multimodal stimuli (audio, visual, and cognitive) to evoke targeted emotions (sadness, stress, anger, and surprise) from the 175 children aged from five to eight years. Classification ratio of 78.43% for three emotions (sadness, stress, and anger) and 61.76% for four emotions (sadness, stress, anger, and surprise) has been achieved by adopting support vector machine as pattern classifier.

Note that the recognition rates in the previous works should be strongly dependent on the datasets they used and context of subjects. Moreover, the physiological datasets used in most of the previous works are gathered by using visual elicitation materials in a lab setting. The subjects then "tried and felt" or "acted out" the target emotions while looking at selected photos or watching movie clips that are carefully prearranged to the emotions. It means, extremely speaking, that the recognition results were achieved for specific users in specific contexts with the "forced" emotional states.

## 3 MUSICAL EMOTION INDUCTION

A well established mechanism of emotion induction would be either to imagine or to recall from individual memory. 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 spontaneously be 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. This claims that certain music can be a powerful cue in bringing emotional experiences from memory back into awareness. Since music listening is often done by an individual in isolation, the possible artifacts by social masking and social interaction may be minimized in the experiment. Furthermore, like odors, music may be treated at lower levels of the brain that are particularly resistant to modifications by later input, contrary to cortically based episodic memory (LeDoux, 1992).

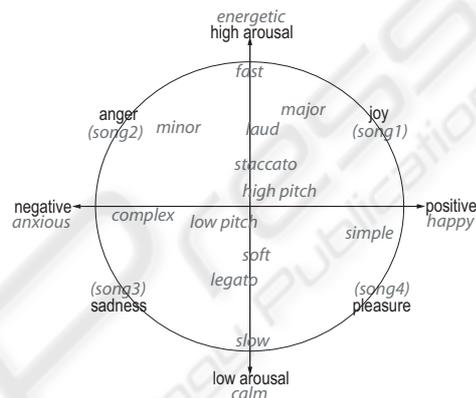


Figure 1: Reference emotional cues in music based on the 2D emotion model. Metaphoric cues for song selection: song1 (enjoyable, harmonic, dynamic, moving), song2 (noisy, loud, irritating, discord), song3 (melancholic, sad memory), song4 (blissful, pleasurable, slumberous, tender).

To collect a database of physiological signals in which the targeted emotions (joy, anger, sadness, pleasure)<sup>1</sup> can be *naturally* reflected without any deliberate expression, we decided to use musical induction method, recording physiological signals while the subjects are listening to different music songs. The subjects were three males (two students and an academic employee) between 25-38 years old and enjoyed listening to music everyday. The subjects individually handpicked four music songs by themselves that should spontaneously evoke their emotional memories and certain moods corresponding to the four target emotions. Figure 1 shows the musical emotion model referred to for the selection of their songs. Generally, emotional responses to music would vary greatly from individual to individual depending on their unique past experiences. Moreover,

<sup>1</sup>We note that these four expression words are used to cover each quadrant in the 2D emotion model, i.e. joy should represent *all* emotions with high arousal and positive valence, anger with high arousal and negative, sadness with low arousal and negative, and pleasure with low arousal and positive valence.

cross-cultural comparisons in literature suggest that the emotional responses can be quite differentially emphasized by different musical cultures and training. This is why we advised the subjects to choose themselves the songs that 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 the subjects to unaffectedly feel the emotions from the music. For the recording, the subject needs to position himself the sensors by instruction posters in the room, to apply the headphones, and to select a song from his song list saved in the computer. When he does mouse-click just at the start of recording, the recording and music systems are automatically setting up by preset values for each song. Recording schedule was determined by the subjects themselves too, at any time when they will listen to music and which song they choose. It means, different from methods used in other studies, that the subjects were not forced to participate in a lab setting scenario and to use prespecified stimulation materials. We believe that this voluntary participation of the subjects during our experiment might be a help to obtain a high-quality dataset with natural emotions. The physiological signals are acquired us-

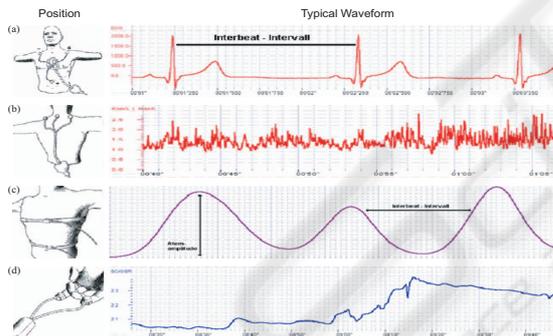


Figure 2: Position and typical waveforms of the biosensors: (a) ECG, (b) EMG, (c) RSP, (d) SC.

ing the Procomp Infiniti™ (www.mindmedia.nl) with four biosensors, electromyogram (EMG), skin conductivity (SC), electrocardiogram (ECG), and respiration (RSP). The sampling rates are 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. 2. We used pre-gelled single Ag/AgCl electrodes for ECG and EMG sensors and standard single Ag/AgCl electrodes fixed with two finger bands for SC sensor. A stretch sensor using latex rubber band fixed with velcro respiration belt is used to capture breathing activity of the subjects. It can be worn either thoracically or abdominally, over clothing.

During the three months, a total of 360 samples (90 samples for each emotion) from three subjects is collected. Signal length of each sample is between 3-5 minutes depending on the duration of the songs.

## 4 METHODOLOGY

Overall structure of our recognition system is illustrated in Figure 3.

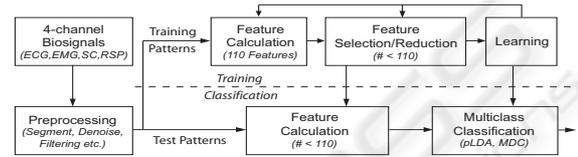


Figure 3: Block diagram of supervised statistical classification system for emotion recognition.

### 4.1 Preprocessing

Different types of artifacts were observed in all the four channel signals, such as transient noise due to movement of the subjects during the recording, mostly at the begin and end of the each recording. Thus, uniformly for all subjects and channels, we segmented the signals into final samples with fixed length of 160 seconds by cutting out from the middle part of each signal. Particularly to the EMG signal, we needed to pay closer attention because the signal contains artifacts generated by respiration and heart beat (Fig. 4). It was due to the position of EMG sensor at the nape of the neck. For other signals we used pertinent lowpass filters to remove the artifacts without loss of information.

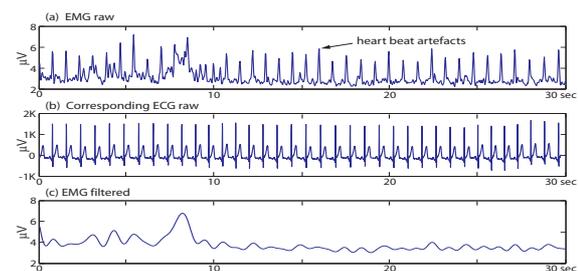


Figure 4: Example of EMG signal with heart beat artifacts and denoised signal.

### 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, subband spectra, etc. For the signals with non-periodic characteristics, such as EMG and SC, we focused on capturing the amplitude variance and localizing the occurrences (number of transient changes) in the signals.

#### 4.2.1 Electrocardiogram (ECG)

To obtain subband spectrum of the ECG signal we used the typical 1024 points fast Fourier transform (FFT) and partitioned the coefficients within the frequency range 0-10 Hz into eight non-overlapping subbands with equal bandwidth. First, as features, power mean values of each subband and fundamental frequency ( $F_0$ ) are calculated by finding maximum magnitude in the spectrum within the range 0-3 Hz. To capture peaks and their locations in subbands, subband spectral entropy (SSE) is computed for each subband. To compute the SSE, it is necessary to convert each spectrum into a probability mass function (PMF) like form. Eq. 1 is used for the normalization of the spectrum.

$$x_i = \frac{X_i}{\sum_{i=1}^N X_i}, \quad \text{for } i = 1 \dots N \quad (1)$$

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

$$H_{\text{sub}} = - \sum_{i=1}^N x_i \cdot \log_2 x_i \quad (2)$$

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

To obtain the HRV (heart rate variability) 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 (Pan and Tompkins, 1985) in order to obtain the HRV time series. Figure 5 shows example of R wave detection and interpolated HRV time series, referring to the increases and decreases over time in the NN intervals.

In time-domain of the HRV, we calculated statistical features including mean value, standard deviation of all NN intervals (SDNN), standard deviation of first difference of the HRV, the number of pairs

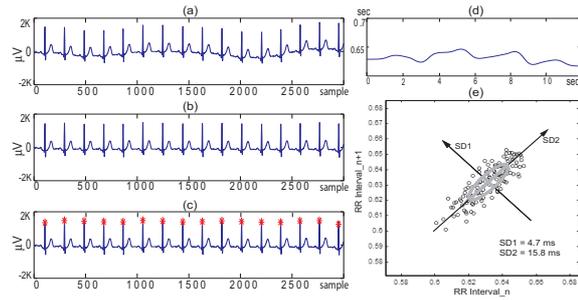


Figure 5: Example of ECG Analysis: (a) raw ECG signal with respiration artifacts, (b) detrended signal, (c) detected RR interbeats, (d) interpolated HRV time series, (e) Poincaré plot of the HRV time series.

of successive NN intervals differing by greater than 50 ms (NN50), 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 (RR) 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 the  $R - R(i)$  to the lines  $y = x$  and  $y = -x + 2 * R - R_m$ , where  $R - R_m$  is the mean of all  $R - R(i)$ , (Kamen et al., 1996). Figure 5.(e) 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 are also considered. Based on the so-called *approximate entropy* and *sample entropy* proposed in (Richmann and Moorman, 2000), a multiscale sample entropy (MSE) has been introduced (Costa et al., 2005) and successfully applied to physiological data, especially for analysis of short and noisy biosignal. Given a time series  $\{X_i\} = \{x_1, x_2, \dots, x_N\}$  of length  $N$ , the number ( $n_i^{(m)}$ ) of similar  $m$ -dimensional vectors  $y^{(m)}(j)$  for each sequence vectors  $y^{(m)}(i) = \{x_i, x_{i+1}, \dots, x_{i+m-1}\}$  is determined by measuring their respective distances. The relative frequency to find the vector  $y^{(m)}(j)$  within a tolerance level  $\delta$  is defined by

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

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

$$h_A(\delta, m) = \lim_{N \rightarrow \infty} [H_N^{(m)}(\delta) - H_N^{(m+1)}(\delta)], \quad (4)$$

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

where

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

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

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

The time series  $\{X_i\}$  is first divided into  $N/\tau$  segments by non-overlapped windowing with length of scale factor  $\tau$  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  $\sigma$  is the standard deviation of the original time series  $x_i$ . Note that using the fixed tolerance level  $\delta$  as a percentage of the standard deviation corresponds to initial normalizing of the time series and it thus enables that  $h_S$  does not depend on the variance of the original time series, but only on their sequential ordering.

In frequency-domain of the HRV time series, three frequency bands are of interest in general; very-low frequency (VLF) band (0.003-0.04 Hz), low frequency (LF) band (0.04-0.15 Hz), and high frequency (HF) band (0.15-0.4 Hz). From these subband spectra, we computed dominant frequency and power of each band by integrating the power spectral densities (PSD) obtained by using Welch's algorithm, and ratio of power within the low-frequency and high-frequency band (LF/HF). Since the parasympathetic activity dominates at high frequency, the ratio of LF/HF is generally thought to distinguish sympathetic effects from parasympathetic effects (Malliani, 1999).

#### 4.2.2 Respiration (RSP)

Including the typical statistics of the raw RSP signal we calculated similar types of features like the ECG features, power mean values of three subbands (obtained by dividing the Fourier coefficients within the range 0-0.8 Hz into non-overlapped three subbands with equal bandwidth), and the set of subband spectral entropies (SSE).

In order to investigate inherent correlation between respiration rate and heart rate, we considered a novel feature content for the RSP signal. Since RSP signal exhibits quasi periodic waveform with sinusoidal property, it is not unreasonable to process HRV like analysis for the RSP signal, i.e. to estimate breathing rate variability (BRV). After detrending with mean value of the entire signal and low-pass filtering, we calculated the BRV by detecting the peaks in the signal using the maxima ranks within each zero-crossing. From the BRV time series, similar to the ECG signal, we calculated mean value, SD, SD of first difference, MSE, Poincaré analysis, etc. In the spectrum of the BRV, peak frequency, power of two subbands, low-frequency band (0-0.03Hz) and high-frequency band (0.03-0.15 Hz), and the ratio of power within the two bands (LF/HF) are calculated.

#### 4.2.3 Skin Conductivity (SC)

The mean value, standard deviation, and mean of first and second derivations are extracted as features from the normalized SC signal and the low-passed SC signal using 0.2 Hz of cutoff frequency. To obtain a detrended SCR (skin conductance response) waveform without DC-level components, we removed continuous, piecewise linear trend in the two low-passed signals, i.e., very low-passed (VLP) with 0.08 Hz and low-passed (LP) signal with 0.2 Hz of cutoff frequency, respectively (see Fig. 6 (a)-(e)).

The baseline of the SC signal was calculated and subtracted to consider only relative amplitudes. By

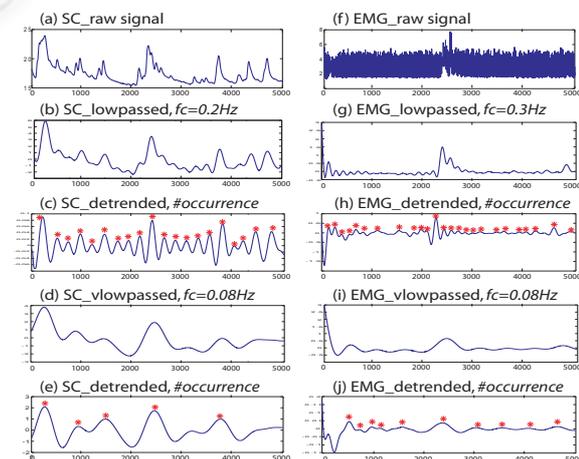


Figure 6: Analysis Examples of SC and EMG signals.

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, mean of the amplitudes of all occur-

rences, and ratio of the SCR occurrences within the low-passed signals (VLP/LP).

#### 4.2.4 Electromyography (EMG)

For the EMG signal we calculated similar types of features as in the case of the SC signal. From normalized and low-passed signals, the mean value of entire signal, the mean of first and second derivations, and the standard deviation are extracted as features. The occurrence number of myo-responses and ratio of that within VLP and LP signals are also added in feature set by similar manner used for detecting the SCR occurrence but with 0.08 Hz (VLP) and 0.3 Hz (LP) of cutoff frequency (see Fig. 6.(f)-(j)).

Finally we obtained a total of 110 features from the 4-channel biosignals; 53 (ECG) + 37 (RSP) + 10 (SC) + 10 (EMG).

## 5 CLASSIFICATION RESULT

For classification we used the pseudoinverse linear discriminant analysis (pLDA) (Ye and Li, 2005), combined with the sequential backward selection (SBS) (Kittler, 1986) to select significant feature subset. The pLDA is a natural extension of classical LDA by applying eigenvalue decomposition to the scatter matrices, in order to deal with the singularity problem of LDA. Table 1 with confusion matrix presents the correct classification ratio ( $CCR$ ) of subject-dependent (Subject A, B, and C) and subject-independent (All) classification where the features of all the subjects are simply merged and normalized. We used leave-one-out cross-validation where a single observation taken from the entire samples is used as the test data and 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. In the Table, it turned out that the  $CCR$  is depending on subject to subject. For example, best accuracy was 91% from subject B and lowest was 81% from subject A. Not only the overall accuracy but the  $CCR$  of the single emotions differs from subject to subject. On the other side, it is very meaningful that relatively robust recognition accuracy is achieved for the classification between emotions that are reciprocal with each other regarding the diagonal quadrants in the 2D emotion model, i.e., joy vs. sadness and anger vs. pleasure. Moreover, the accuracy is much better than that of arousal classification.

The  $CCR$  of subject-independent classification was not comparable to that obtained for subject-dependent classification. As shown in Figure 7, merg-

Table 1: Recognition results in rates ( $error\ 0.00 = CCR\ 100\%$ ) achieved by using pLDA with SBS and leave-one-out cross validation.

# of samples: 120 for each subject and 360 for All.

| Subject A ( $CCR\ \% = 81\%$ ) |     |       |         |          |        |       |
|--------------------------------|-----|-------|---------|----------|--------|-------|
|                                | joy | anger | sadness | pleasure | total* | error |
| joy                            | 22  | 4     | 1       | 3        | 30     | 0.27  |
| anger                          | 3   | 26    | 1       | 0        | 30     | 0.13  |
| sadness                        | 1   | 2     | 23      | 4        | 30     | 0.23  |
| pleasure                       | 3   | 0     | 1       | 26       | 30     | 0.13  |

| Subject B ( $CCR\ \% = 91\%$ ) |     |       |         |          |        |       |
|--------------------------------|-----|-------|---------|----------|--------|-------|
|                                | joy | anger | sadness | pleasure | total* | error |
| joy                            | 27  | 3     | 0       | 0        | 30     | 0.10  |
| anger                          | 3   | 25    | 1       | 1        | 30     | 0.17  |
| sadness                        | 0   | 2     | 28      | 0        | 30     | 0.07  |
| pleasure                       | 0   | 1     | 0       | 29       | 30     | 0.03  |

| Subject C ( $CCR\ \% = 89\%$ ) |     |       |         |          |        |       |
|--------------------------------|-----|-------|---------|----------|--------|-------|
|                                | joy | anger | sadness | pleasure | total* | error |
| joy                            | 28  | 0     | 2       | 0        | 30     | 0.07  |
| anger                          | 0   | 30    | 0       | 0        | 30     | 0.00  |
| sadness                        | 0   | 0     | 24      | 6        | 30     | 0.20  |
| pleasure                       | 0   | 0     | 5       | 25       | 30     | 0.17  |

| All: Subject-independent ( $CCR\ \% = 65\%$ ) |     |       |         |          |        |       |
|---|-----|-------|---------|----------|--------|-------|
|   | joy | anger | sadness | pleasure | total* | error |
| joy   | 62  | 9     | 8       | 11       | 90     | 0.31  |
| anger   | 15  | 57    | 13      | 5        | 90     | 0.37  |
| sadness                                       | 9   | 6     | 58      | 17       | 90     | 0.36  |
| pleasure                                      | 8   | 5     | 21      | 56       | 90     | 0.38  |

\*: Actual total # of samples

ing 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 are divided into groups of negative valence (anger/sadness) and positive valence (joy/pleasure) and into groups of high arousal (joy/anger) and low arousal (sadness/pleasure). By using the same methods, we then performed two-class classification of the divided samples for arousal and valence separately. Finally, it turned out that emotion-relevant ANS specificity can be observed more conspicuously in the arousal axis regardless of subject-dependent or independent case. Classification of arousal achieved an acceptable  $CCR$  of 97-99% for the subject-dependent recognition and 89% for the subject-independent recognition, while the results for valence were 88-94% and 77%, respectively.

Table 2: Best emotion-relevant features extracted from four channel physiological signals. Arousal classes: joy+anger/sadness+pleasure, Valence classes: joy+pleasure/anger+sadness, Four classes: joy/anger/sadness/pleasure.

| Classes       | Best Emotion-relevant Features <small>(Ch_value_domain, C: ECG, R: RSP, S: SC, M: EMG)</small>  |
|---------------|---|
| Arousal       | <i>C_std(diff)_HRVtime</i> , <i>C_sd2_PoincareHRV</i> , <i>C_powerLow_HRVspec</i> , <b><i>R_meanEnergy_SubSpectra</i></b> , <b><i>R_sd2_PoincareBRV</i></b><br><b><i>R_mean_MSE</i></b> , <i>S_mean_RawLowpassed</i> , <i>S_std_RawLowpassed</i> , <i>M_#occurrenceRatio_RawLowpassed</i><br><i>M_mean_RawNormed</i>  |
| Valence       | <i>C_sd2_PoincareHRV</i> , <i>C_meanEnergy_SubSpectra</i> , <i>C_ratioLH_HRVspec</i> , <i>C_mean_MSE</i> , <i>C_mean(diff)_MSE</i><br><b><i>R_meanEnergy_SubSpectra</i></b> , <i>R_mean(diff)_SubSpectra</i> , <i>R_sd1_PoincareBRV</i> , <b><i>R_sd2_PoincareBRV</i></b> , <b><i>R_mean_MSE</i></b><br><i>S_mean(diff)_RawNormed</i> , <i>M_mean(diff)_RawNormed</i>   |
| Four Emotions | <i>C_mean_HRVtime</i> , <i>C_std_HRVtime</i> , <i>C_std(diff)_HRVtime</i> , <i>C_mean(diff)_MSE</i> , <i>C_mean_MSE</i> , <i>C_mean_SSE</i><br><i>C_sd2_PoincareHRV</i> , <i>C_mean_SubSpectra</i> , <b><i>R_meanEnergy_SubSpectra</i></b> , <i>R_mean_SSE</i> , <i>R_mean_BRVtime</i><br><i>R_sd1_PoincareBRV</i> , <b><i>R_sd2_PoincareBRV</i></b> , <b><i>R_mean_MSE</i></b> , <i>R_power_BRVspec</i> , <i>S_std_RawLowpassed</i><br><i>S_mean(diff)_RawNormed</i> , <i>S_mean(diff)(diff)_RawLowpassed</i> , <i>S_mean_RawNormed</i> , <i>S_#occurrence_RawLowpassed</i><br><i>M_mean(diff)_RawNormed</i> |

: overall selected features are printed in bold

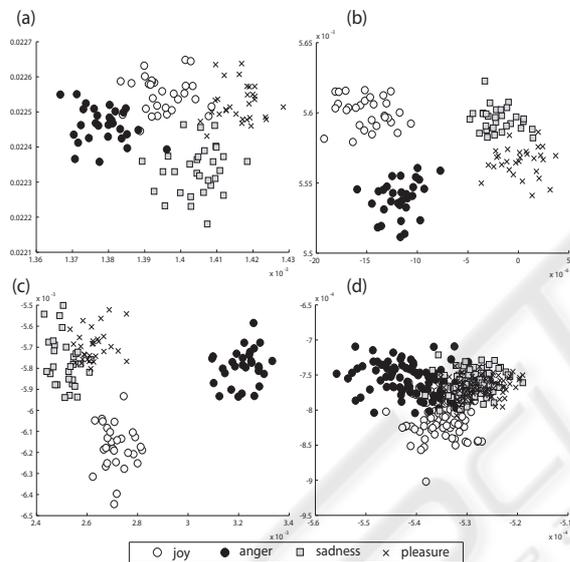


Figure 7: Comparison of feature distributions of subject-dependent and subject-independent case. (a) Subject A, (b) Subject B, (c) Subject C, (d) Subject-independent.

## 6 BEST EMOTION-RELEVANT FEATURES

In Table 2, the best emotion-relevant features, that 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 certain feature domain. The features obtained from time/frequency analysis of HRV time series are decisive for arousal and four emotions classification, while the features from MSE domain of

ECG signals are a predominant factor for correct valence differentiation. Particularly, mutually sympathizing correlate between HRV and BRV which is firstly proposed in this paper has been clearly observed in all the classification problems by the features from their time/frequency analysis and Poincaré domain, *\_PoincareHRV* and *\_PoincareBRV*. This reveals a manifest cross-correlation between respiration and cardiac activity with respect to emotional state. Furthermore, the correlation between heart rate and respiration is obviously captured by the features from HRV power spectrum (*\_HRVspec*), the fast/long-term HRV/BRV analysis using Poincaré method, and the multiscale variance analysis of HRV/BRV (*\_MSE*), and that the peaks of high frequency range in HR subband spectrum (*\_SubSpectra*) provide information about how the sinoatrial node responds to vagal activity at certain respiration frequency.

In addition, we analyzed the number of selected features for the three classification problems, arousal, valence, and four emotion states. For the arousal classification, relatively few features are used but achieved higher recognition accuracy compared to the other class problems. After the ratio of number of selected features to the total feature number of each channel, it was obvious that the SC and EMG activities reflected in both *\_RawLowpassed* and *\_RawNormed* domains (see Table 2) are more significant for arousal classification than the other channels. This supports also the experimental elucidation in previous works that the SCR is linearly correlated with the intensity of arousal. On the other side, we could observe a remarkable increase of number of the ECG and RSP features for the case of valence classification.

## 7 CONCLUSIONS

In this paper, we treated all essential stages of automatic emotion recognition system using multichannel physiological measures, from data collection up to classification process, and analyzed the results from each stage of the system. For four emotional states of three subjects, we achieved average recognition accuracy of 91% which connotes more than a prima facie evidence that there exist some ANS differences among emotions.

A wide range of physiological features from various analysis domains including time, frequency, entropy, geometric analysis, subband spectra, multiscale entropy, and HRV/BRV has been proposed to search the best emotion-relevant features and to correlate them with emotional states. The selected best features are specified in detail and their effectiveness is 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 heart rate and respiration.

As we humans use several modalities jointly to interpret emotional states since emotion affects almost all modes, one most challenging issue in near future work is to explore multimodal analysis for emotion recognition. Toward the human-like analysis and finer resolution of recognizable emotion classes, an essential step would be therefore to find innate priority among the modalities to be preferred for each emotional state. In this sense, physiological channel can be considered as a “baseline channel” in designing a multimodal fashion of emotion recognition system, since it provides several advantages over other external channels and acceptable recognition accuracy, as we presented in this work.

## ACKNOWLEDGEMENTS

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

## REFERENCES

- Costa, M., Goldberger, A. L., and Peng, C.-K. (2005). Multiscale entropy analysis of biological signals. *Phys. Rev.*, E 71(021906).
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001).

- Emotion recognition in human-computer interaction. *IEEE Signal Processing Mag.*, 18:32–80.
- Gross, J. J. and Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9:87–108.
- Healey, J. and Picard, R. W. (1998). Digital processing of affective signals. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, pages 3749–3752, Seattle, WA.
- Kamen, P. W., Krum, H., and Tonkin, A. M. (1996). Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity. *Clin. Sci.*, 91:201–208.
- Kim, K. H., Bang, S. W., and Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, 42:419–427.
- Kittler, J. (1986). *Feature Selection and Extraction*, pages 59–83. Academic Press, Inc.
- LeDoux, J. E. (1992). *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, pages 339–351. New York: Wiley-Liss.
- Malliani, A. (1999). The pattern of sympathovagal balance explored in the frequency domain. *News Physiol. Sci.*, 14:111–117.
- Nasoz, F., Alvarez, K., Lisetti, C., and Finkelstein, N. (2003). Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology, and Work - Special Issue on Presence*, 6(1).
- Pan, J. and Tompkins, W. (1985). A real-time qrs detection algorithm. *IEEE Trans. Biomed. Eng.*, 32(3):230–232.
- Picard, R., Vyzas, E., and Healy, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Trans. Pattern Anal. and Machine Intell.*, 23(10):1175–1191.
- Richmann, J. and Moorman, J. (2000). Physiological time series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039.
- Ye, J. and Li, Q. (2005). A two-stage linear discriminant analysis via qr-decomposition. *pami*, 27(6).