

# Emotion-specific Dichotomous Classification and Feature-level Fusion of Multichannel Biosignals for Automatic Emotion Recognition

Jonghwa Kim and Elisabeth André

**Abstract**— Endowing the computer with the ability to recognize human emotional states is the most important prerequisites for realizing an affect-sensitive human-computer interaction. In this paper, we deal with all the essential stages of an automatic emotion recognition system using multichannel physiological measures, from data collection to the classification process. Particularly we develop two different classification methods, feature-level fusion and emotion-specific classification scheme. Four-channel biosensors were used to measure electromyogram, electrocardiogram, skin conductivity, and respiration changes while subjects were listening to music. A wide range of physiological features from various analysis domains is proposed to correlate them with emotional states. Classification of four musical emotions is performed by using feature-level fusion combined with an extended linear discriminant analysis (pLDA). Furthermore, by exploiting a dichotomic property of the 2D emotion model, we developed a novel scheme of emotion-specific multilevel dichotomous classification (EMDC) and compare its performance with direct multiclass classification using the pLDA feature-level fusion. Improved recognition accuracy of 95% and 70% for subject-dependent and subject-independent classification, respectively, is achieved by using the EMDC scheme.

## I. INTRODUCTION

Emotional intelligence (understanding and expression of emotions) is indispensable in human communication and facilitates successful interpersonal social interaction. To approach this in human-computer interaction (HCI), the first step is 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 affect-sensitive HCI is a reliable emotion recognition system which guarantees acceptable recognition accuracy, robustness against any artifacts, and adaptability to practical applications.

Recently, many works on engineering approaches to automatic emotion recognition have been reported. For an overview we refer to [?]. In particular, many efforts have been deployed to recognize human emotions using audiovisual channels of emotion expression, i.e. facial expressions, speech, and gestures. Little attention, however, has been paid so far to using physiological measures, as opposed to audiovisual emotion channels. This is due to some significant limitations that come with the use of multichannel biosignals 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. When using multiple biosensors at the same time, analyzing biosignals is itself a complex multivariate task and requires broad insight into biological processes related to neuropsychological functions. To classify the multichannel variables, we first need to design fusion method for multichannel sensory data and to develop an emotion-specific classification scheme. Most of machine learning algorithms are generalized method based on statistics or linear regression of given data and most suitable for binary classification problems. Therefore they might not be able to capture characteristics of input variables in order to efficiently solve multiclass problems.

In this paper, we deal with all the essential stages of an automatic emotion recognition system, from data collection to the classification, based on four-channel physiological signals: electromyogram (EMG), electrocardiogram (ECG), skin conductivity (SC), and respiration changes (RSP). Generally, fusion of multisensory data can be performed at least at three levels: data, feature, and decision level. When observations are of same type, the data-level fusion might be probably the most appropriate way where we simply combine raw multisensory data. Decision-level fusion is the approach applied most often for *multimodal* sensory data containing time scale differences between modalities. Since, in this paper, we use *multichannel* biosignals that are measured in synchronized time scale and unique dimension, feature-level fusion is the most convincing way to classify by single classifier. Furthermore we develop a novel scheme of emotion-specific multilevel dichotomous classification (EMDC) and compared its performance with 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.

## II. RELATED RESEARCH

A significant amount of work has been conducted by Picard and colleagues at MIT Lab showing that certain affective states may be recognized by using physiological data including heart rate, skin conductivity, temperature, muscle activity and respiration velocity [?][?]. Nasoz et al. [?] used movie clips based on the study by Gross and Levenson [?] for eliciting target emotions from 29 subjects

This work was partially supported by European Projects HUMAINE and METABO

J. Kim and E. André are with Faculty of Applied Computer Science, University of Augsburg, D-86159 Augsburg, Germany  
{kim, andre}@informatik.uni-augsburg.de

and achieved an emotion classification accuracy of 83% using the Marquardt Backpropagation algorithm (MBP). More recently, an interesting user-independent emotion recognition system was reported by Kim et al. [?]. 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% was achieved for three emotions (sadness, stress, and anger) and a ratio of 61.76% for four emotions (sadness, stress, anger, and surprise) by adopting support vector machines as pattern classifier.

The physiological datasets used in most of the aforementioned approaches were 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 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. All the works used simple feature-level fusion to mix features from each sensor and then to classify by using common single classifier.

### III. SETTING OF EXPERIMENT

#### A. 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 (i.e. EQ1, EQ2, EQ3, and EQ4 in Fig.??) can be *naturally* reflected without any deliberate expression, we decided to use the musical induction method, i.e. to record physiological signals while the subjects were listening to different pieces of music.

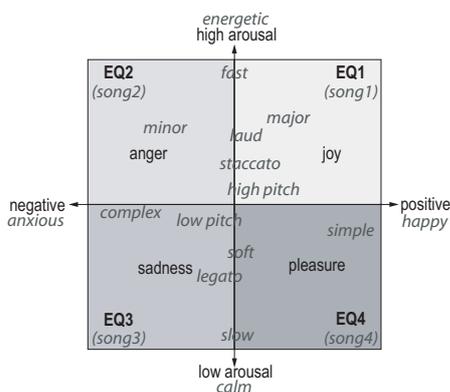


Fig. 1. Reference emotional cues in music based on the 2D emotion model. EQ1=positive/high arousal, EQ2=negative/high arousal, EQ3=negative/low arousal, EQ4=positive/low arousal

The subjects were three males aged 25-38 and who all enjoy listening to music in their everyday life. They individually handpicked four songs that were intended to spontaneously evoke emotional memories and certain moods corresponding to the four target emotions. Figure ??<sup>1</sup> shows

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

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 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 themselves the songs they believed would help them recall their individual special memories with respect to the target emotions. 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 lab setting scenario and to use prespecified stimulation material. *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 3-5 minutes depending on the duration of the songs.*

#### B. Used Biosensors

The physiological signals were acquired using the Procomp<sup>2</sup> Infniti<sup>TM</sup> with four biosensors, electromyogram (EMG), skin conductivity (SC), electrocardiogram (ECG), and respiration (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. ??.

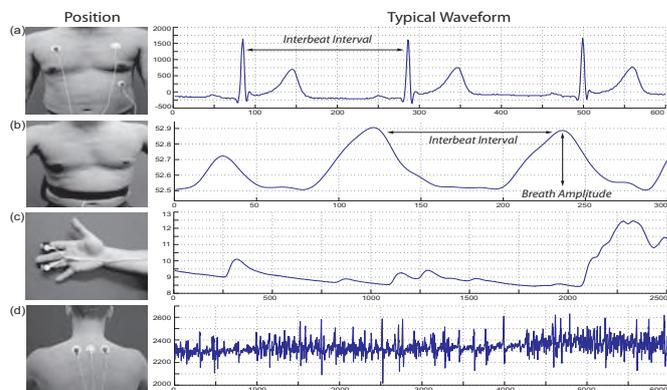


Fig. 2. Position and typical waveforms of the biosensors: (a) ECG, (b) RSP, (c) SC, (d) EMG.

### IV. FEATURE EXTRACTION

#### A. Signal Segmentation

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 beginning and at the end of each recording. Thus, uniformly for all subjects and channels, we segmented the signals into final samples of a 160 seconds each, obtained by taking the middle part of each signal.

<sup>2</sup>This is an 8 channel multi-modal Biofeedback system with 14 bit resolution and a fiber optic cable connection to the computer. www.MindMedia.nl

## B. 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.

1) *Electrocardiogram (ECG)*: To obtain the 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 (F0) 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. ?? 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  $i^{\text{th}}$  frequency component of the spectrum and  $\tilde{x} = \{x_1 \dots x_N\}$  is to be considered as the PMF of the spectrum. In each subband the SSE is computed from  $\tilde{x}$  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 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.

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 [?] in order to obtain the HRV time series. Figure ?? shows examples of R wave detection and 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 mean value, standard deviation of all NN intervals (SDNN), standard deviation of the first difference of the HRV, the number of pairs of successive NN intervals differing by more 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 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  $R - R(i)$ , [?]. Figure ??(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.

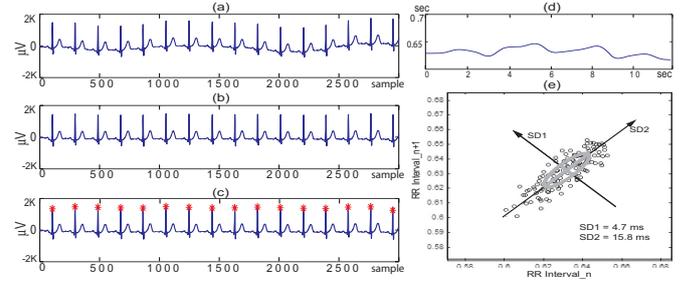


Fig. 3. Example of ECG Analysis: (a) raw ECG signal with respiration artifacts, (b) detrended signal, (c) detected RR interbeats, (d) interpolated HRV time series using RR intervals, (e) Poincaré plot of the HRV time series.

Entropy-based features from the HRV time series were also considered. Based on the so-called *approximate entropy* and *sample entropy* proposed in [?], a multiscale sample entropy (MSE) was introduced [?] 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 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  $\{X_i\}$ ,

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

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  $x_i$ .

In the frequency-domain of the HRV time series, three frequency bands are of general interest: the very-low frequency (VLF) band (0.003-0.04 Hz), the low frequency (LF) band (0.04-0.15 Hz), and the high frequency (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 (PSD) obtained by using Welch's algorithm, as well as the ratio of power within the low-frequency band to that within the high-frequency band (LF/HF).

2) *Respiration (RSP)*: 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 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 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, i.e. analysis of breathing rate variability (BRV). After detrending using the mean value of the entire signal and lowpass 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 maxima ranks within each zero-crossing. 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, peak frequency, power of the two subbands, the low-frequency band (0-0.03Hz) and the high-frequency band (0.03-0.15 Hz), and the ratio of the power within the two bands (LF/HF) were calculated.

3) *Skin Conductivity (SC)*: The mean value, standard deviation, and mean of first and second derivations were extracted as features from the normalized SC signal and the low-passed 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 low-passed signals, i.e., the very low-passed (VLP) and the low-passed (LP) signal with a cutoff frequency of 0.08 Hz and 0.2 Hz, respectively (see Fig. ?? (a)-(e)).

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 low-passed signals (VLP/LP).

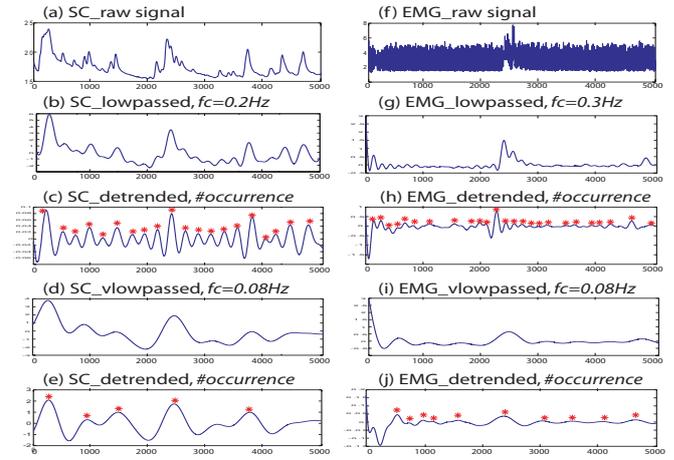


Fig. 4. Analysis Examples of SC and EMG signals

4) *Electromyography (EMG)*: 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 low-passed signals. The occurrence number of myo-responses 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 with 0.08 Hz (VLP) and 0.3 Hz (LP) (see Fig. ??.(f)-(j)).

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

## V. CLASSIFICATION RESULT

### A. pLDA Feature-level Fusion

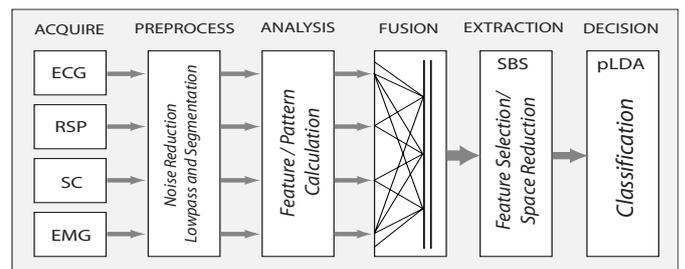


Fig. 5. Feature-level fusion for four-channel biosensor data, combined with SBS and pLDA

Fig. ?? shows the feature-level fusion model for multi-channel biosensor data. For classification we used the pseudoinverse linear discriminant analysis (pLDA) [?], combined with the sequential backward selection (SBS) [?] 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 ?? with confusion matrix presents the correct classification ratio ( $CCR$ ) of subject-dependent (Subject A, B,

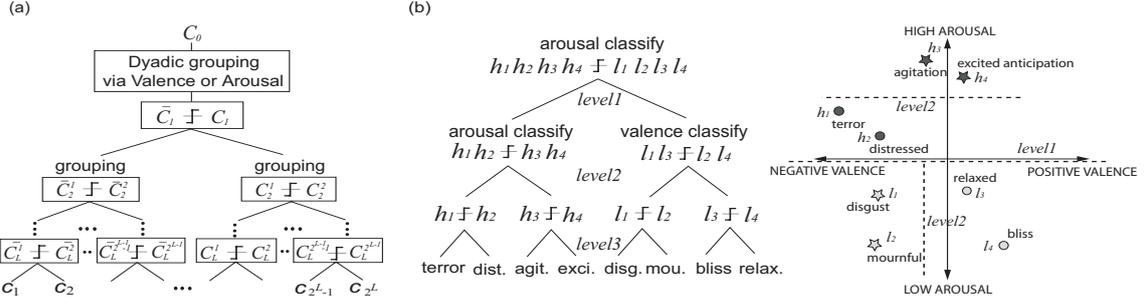


Fig. 6. Framework of emotion-specific multilevel dichotomous classification (EMDC). (a) Diagram of decomposition process, (b) Decomposition example for an eight-class problem

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.

TABLE I

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 ( <i>CCR</i> % = 81%)						
	EQ1	EQ2	EQ3	EQ4	<i>total</i> *	<i>error</i>
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 ( <i>CCR</i> % = 91%)						
	EQ1	EQ2	EQ3	EQ4	<i>total</i> *	<i>error</i>
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 ( <i>CCR</i> % = 89%)						
	EQ1	EQ2	EQ3	EQ4	<i>total</i> *	<i>error</i>
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 ( <i>CCR</i> % = 65%)						
	EQ1	EQ2	EQ3	EQ4	<i>total</i> *	<i>error</i>
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

The table shows that the *CCR* varies from subject to subject. For example, the best accuracy was 91% for subject B and the lowest was 81% for subject A. Not only does the overall accuracy differ from one subject to the next, but the *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 three times mixed up 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 Subject 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, i.e., EQ1 vs. EQ3 and EQ2 vs. EQ4. Moreover, the accuracy is much better than that of arousal classification.

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. It turned out that emotion-relevant ANS specificity can be observed more conspicuously in the arousal axis regardless of subject-dependent or independent cases. 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.

### B. Emotion-specific Multilevel Dichotomous Classification

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

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 class categories and  $C$  as all remaining patterns, i.e.,  $\bar{C} \cap C = \{\}$ . 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 from 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 a maximum  $CCR$  of 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). 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 II  
RESULTS USING EMDC SCHEME WITH THE BEST FEATURES

Subject A ( $CCR$ % = 94%, 113/120)					
		EQ1 & EQ2		EQ3 & EQ4	
EQ1 & EQ2	58		2		
		EQ1 EQ2			
	EQ1 EQ2	27 1 1 29			
EQ3 & EQ4	2		58		
				EQ3 EQ4	
			EQ3 EQ4	29 0 1 28	
Subject B ( $CCR$ % = 98%, 117/120)					
		EQ1 & EQ2		EQ3 & EQ4	
EQ1 & EQ2	60		0		
		EQ1 EQ2			
	EQ1 EQ2	30 0 0 30			
EQ3 & EQ4	1		59		
				EQ3 EQ4	
			EQ3 EQ4	29 1 1 28	
Subject C ( $CCR$ % = 94%, 113/120)					
		EQ1 & EQ2		EQ3 & EQ4	
EQ1 & EQ2	60		0		
		EQ1 EQ2			
	EQ1 EQ2	30 0 0 30			
EQ3 & EQ4	1		59		
				EQ3 EQ4	
			EQ3 EQ4	27 2 4 26	
Subject All ( $CCR$ % = 70%, 251/360)					
		EQ1 & EQ2		EQ3 & EQ4	
EQ1 & EQ2	155		25		
		EQ1 EQ2			
	EQ1 EQ2	62 13 15 65			
EQ3 & EQ4	16		164		
				EQ3 EQ4	
			EQ3 EQ4	64 19 21 60	

TABLE III  
 $CCR$  COMPARISON BETWEEN pLDA AND EMDC

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

Table ?? shows the dichotomous contingency table of recognition results by using the novel EMDC scheme. As expected, the  $CCR$ s significantly improved for all class problems. For the classification of four emotions, we obtained an average  $CCR$  of 95% for subject-dependent and 70% for subject-independent classification. Compared to the results obtained for pLDA, the EMDC scheme achieved an overall  $CCR$  improvement of about 5%-13% in each class problem (see Table ??).

## VI. 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. By analyzing a wide range of physiological features from various analysis domains, 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.

By fusing the multichannel features at the feature-level, we achieved an average recognition accuracy of 98%, 91%, and 87% for arousal, valence, and four emotion classes respectively. In order to further improve the accuracy of the four emotion classes, we developed a new classification scheme (EMDC). Although this new scheme is based on a very simple idea, it significantly improves the recognition accuracy obtained by common feature-level fusion. We actually achieved an average recognition accuracy of 95% improved which also 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, even when considering the different experimental settings in the works, such as the number of target classes, number of subjects, naturalness of dataset, etc.

## VII. ACKNOWLEDGMENTS

This research was partially supported by the European Commission (HUMAINE NoE; FP6-IST, METABO; FP7-ICT).

## REFERENCES

- [1] 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 Mag.*, vol. 18, pp. 32–80, 2001.
- [2] J. Healey and R. W. Picard, "Digital processing of affective signals," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, Seattle, WA, 1998, pp. 3749–3752.
- [3] R. Picard, E. Vyzas, and J. Healy, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [4] F. Nasoz, K. Alvarez, C. Lisetti, and N. Finkelstein, "Emotion recognition from physiological signals for presence technologies," *International Journal of Cognition, Technology, and Work - Special Issue on Presence*, vol. 6(1), 2003.
- [5] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol. 9, pp. 87–108, 1995.
- [6] K. H. Kim, S. W. Bang, and S. R. Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical & Biological Engineering & Computing*, vol. 42, pp. 419–427, 2004.
- [7] J. Pan and W. Tompkins, "A real-time qrs detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. 32, no. 3, pp. 230–233, 1985.
- [8] P. W. Kamen, H. Krum, and A. M. Tonkin, "Poincare plot of heart rate variability allows quantitative display of parasympathetic nervous activity," *Clin. Sci.*, vol. 91, pp. 201–208, 1996.
- [9] J. Richmann and J. Moorman, "Physiological time series analysis using approximate entropy and sample entropy," *Am. J. Physiol. Heart Circ. Physiol.* 278, vol. H2039, 2000.
- [10] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev.*, vol. E 71, no. 021906, 2005.
- [11] R. Thuraisingham and G. Gottwald, "On multiscale entropy analysis for physiological data," *Physica A*, 2005.
- [12] J. Ye and Q. Li, "A two-stage linear discriminant analysis via qr-decomposition," *pami*, vol. 27, no. 6, June 2005.
- [13] J. Kittler, *Feature Selection and Extraction*. Academic Press, Inc, 1986, pp. 59–83.