THE ORIGINAL VERSION OF THIS PAPER WAS FIRST PUBLISHED IN THE *PROCEEDINGS OF THE 17TH EUROPEAN SIGNAL PROCESSING CONFERENCE (EUSIPCO)*, GLASGOW, SCOTLAND, AUGUST 2009, PP. 2357-2361.
CORRECTIONS APPEAR IN RED. THANKS TO DANIELLE CUMMINGS OF TEXAS A&M UNIVERSITY FOR POINTING OUT MISTAKES.

CLASSIFICATION OF USER STATES WITH PHYSIOLOGICAL SIGNALS: ON-LINE GENERIC FEATURES VS. SPECIALIZED FEATURE SETS

F. Hönig J. Wagner†, A. Batliner* E. Nöth**

* Lehrstuhl für Mustererkennung,
Universität Erlangen-Nürnberg,
Martensstr. 3, 91058 Erlangen, Germany
phone: +49 9131 852 7872, fax: +49 9131 30 38 11
email: hoenig@informatik.uni-erlangen.de
web: http://www5.informatik.uni-erlangen.de/hoenig

† Lehrstuhl für Multimedia-Konzepte und Anwendungen,
Universität Augsburg,
Eichleitnerstr. 30, 86159 Augsburg, Germany
phone: +49 821 598 2347, fax: +49 821 598 2349
email: johannes.wagner@informatik.uni-augsburg.de
web: http://mm-werkstatt.informatik.uni-augsburg.de

ABSTRACT

For on-line classification of user states such as emotions or stress levels, we present a new, generic, and efficient physiological feature set. In contrast to common approaches using features specifically tailored to each physiological signal, we break up feature extraction into a simple, signal-specific pre-processing step, and the calculation of a comprehensive set of signal-independent features. This systematizes feature design for each physiological signal and facilitates the transfer to other signals. The time complexity of the approach is independent of the size of the analysis window and of the frequency with which feature vectors are computed for classification. We also provide a variant of the feature set that has low memory requirements. Thus, our approach is well suited for implementing real-time applications. We evaluate the proposed features with an emotion and a stress classification task, showing that they are competitive w.r.t. the performance of classifications using signal-tuned state-of-the-art features.

1. INTRODUCTION

In Human-Computer-Interfaces, there is great potential in accounting for and appropriately reacting to the current cognitive, affective, or emotional state of the user. User states are signalled in various channels, for example speech, gesture, or facial expression. Of particular interest are *physiological signals*, i. e. recordings of body functions such as heart rate or skin conductivity. They cannot be controlled up to the same extent by the user and are often not consciously perceived. Cognitive and affective processes affect physiological signals which therefore are an intriguing candidate for acquiring information on user states.

It has been shown that under favourable conditions, emotional states can be recognized automatically from physiological signals by means of pattern recognition techniques with high reliability [7, 4, 9]. Most of the existing studies, however, are based on off-line processing of relatively long, pre-defined segments of clear recordings covering steady and pronounced instances of the user states to be recognized. It remains yet to be answered whether acceptable performance across subjects can be obtained under the less restricted conditions of applications where a classification system must possibly deal with artefacts, half-blown, changing, and mixed user states; moreover, decisions have to be delivered in real-time.

Extending the work of Vyzas et al. [8], we take a step into the direction of more realistic conditions by proposing a feature set that can be computed on-line with a low and steady computational load. This is critical because in applications, the classification module can be required to deliver a continuously updated response with minimal delay. As a reliable classification of the user state from physiological data requires a relatively large context, the amounts of data to be processed in real-time can pose a problem for commonly used feature extraction approaches. Also, the efficient computation makes our feature set suitable for the limited hardware capabilities of embedded applications. In this context, we also present a variant of our features that has very modest memory requirements.

Another goal in developing the feature set was to simplify and systematize the feature design process. This is achieved by decoupling signal-specific processing from the actual feature extraction: per signal, previous knowledge can be integrated into a relatively simple pre-processing step; the computation of the actual features is then carried out by a generic, comprehensive feature extraction module. Thus, the signal-specific part of the implementation is minimized, and the transfer to other signals and modalities facilitated. This also speeds up the developing cycle in physiological signal research: new signals can be added as input and tested for their usefulness quickly and with low effort.

We evaluate our approach on two databases with different tasks (four-class emotion recognition and stress/non-stress classification) and compare the proposed features with state-of-the-art, signal-tuned features that are computed off-line. For computing these specialized features we use the Augsburg Biosignal Toolbox (AuBT)[1] which has been developed within the HUMAINE[2] research project. These features contain statistical measures such as mean and standard deviation of the preprocessed signals as well as signal-specific, physiologically relevant quantities like heart rate variability (HRV) or energy of HRV in certain frequency bands. For the signals electrocardiogram (ECG), electromyogram (EMG), skin conductivity (SC) and respiration (Resp), 84, 21, 67, and 21 specially tailored features are provided, respectively. They are a superset of the features used in [9] which have proven to be both well suited for emotion classification and competitive to those used in [7].

2. FEATURES

2.1 Derived Signals

As indicated above, we use an identical, generic feature extraction module for all signals. For applying signal-specific knowledge and capturing important information that is contained in the signals but present only in a very implicit representation, we compute characteristic measures from selected signals over time. These *derived signals* are then added to the recorded signals and processed by the generic feature extraction module just as the recorded signals.

Given the recorded signals ECG, EMG, SC, skin temperature (Temp), blood volume pulse (BVP), and Resp, we derive four additional signals. The instantaneous heart rate is computed as a step-function from ECG and BVP, yielding the signals ECG-HR and

This work was funded by the EU within HUMAINE (IST-2002-507422) and by the German Federal Ministry of Education and Research within SmartWeb (Grant 01 IMD 01 F). The responsibility for the content lies with the authors.

[1] http://mm-werkstatt.informatik.uni-augsburg.de/aubt
[2] http://emotion-research.net

BVP-HR. Similarly, the respiration rate (Resp-rate) is derived from the respiratory signal. Finally, the pulse transit time (PTT) computed from ECG and BVP (measured at a finger), which can be regarded as a surrogate parameter of the systolic blood pressure [3], is added. These derived signals can be seen as a more stationary and more explicit representation of important properties of the original signals facilitating subsequent feature extraction.

In the following, the generic features that are extracted from each original and derived signal are outlined. We present two variants of our approach: the set of the *moving features* and the set of the *sliding features*. Both are very efficient in terms of time complexity and thus suited for on-line classification. The sliding features, which are an approximation of the moving features, additionally have very low memory requirements, offering themselves for applications where few memory is to be spent.

2.2 Moving Features

We denote the n-th sample of the current signal by x_n, the sampling frequency by f_s, the length of the analysis window in samples by w, the step size (distance between successive analysis windows in samples) by s, and the frequency with which feature vectors are required for classification by $f = f_s/s$. The value of the i-th feature for an analysis window ending at x_n is $m_{i,n}$.

To efficiently implement on-line computation, each feature is computed *recursively*, using only its previous value $m_{i,n-1}$, the current sample x_n and a few auxiliary variables where required. As a consequence, the time complexity is *constant* in both w and f. It depends only on the sampling frequency and is $O(f_s)$. In contrast, common approaches using all samples of the current analysis window for calculating a feature have a time complexity of at least $O(f \cdot w) = O(f_s \cdot \frac{w}{s})$. Thus, recursive computation is to be preferred in terms of efficiency if w is considerably larger than s, i. e. if the overlap of the analysis windows is large. This is the case in on-line physiological signal classification, as a relatively large context is required for a robust decision.

Prototypical for the computations is the moving average $\mu_{w,n}$, the mean of the w previous values of x_n,

$$\mu_{w,n} = \frac{1}{w}\sum_{i=1}^{w} x_{n-w+i},$$

which is why this feature set is called *moving features*. A recursive formulation is given by

$$\mu_{w,n} = \mu_{w,n-1} - x_{n-w}/w + x_n/w. \qquad (1)$$

To store and access the necessary sample history efficiently, a ring-buffer of size w is used. When using the update rule given by Equation (1) repeatedly to compute $\mu_{w,n}$ with finite-precision floating-point arithmetic, rounding errors can accumulate for large w and render the result useless with time. Our solution is to update the computed value periodically by keeping a second accumulator $s_n = s_{n-1} + x_n$. Each w steps, the recursively computed estimate of $\mu_{w,n}$ is substituted by s_n/w and s_n is zeroed. Thus, a reasonable degree of numerical stability can be achieved for all features while increasing the computational effort only by a constant factor less than 2. Apart from $m_{0,n} = \mu_{w,n}$, also $m_{1,n} = |\mu_{w,n}|$ and $m_{2,n} = (\mu_{w,n})^2$ are used as features. A similar scheme is used for the other computed quantities.

$\mu_{w,n}$ is the mean value of the signal when weighted with a rectangular window of length w. The mean values when using a triangular and bell-shaped window are given by

$$\begin{aligned}
\mu^{(2)}_{w,n} &= \mu^{(2)}_{w,n-1} - \mu_{w_2,n-w_1}/w_1 + \mu_{w_2,n}/w_1,\\
&\quad w_1 = \lfloor w/2\rfloor + 1, w_2 = w + 1 - w_1 \quad \text{and}\\
\mu^{(3)}_{w,n} &= \mu^{(3)}_{w,n-1} - \mu^{(2)}_{w_4,n-w_3}/w_3 + \mu^{(2)}_{w_4,n}/w_3,\\
&\quad w_3 = \lfloor w/3\rfloor + 1, w_4 = w + 1 - w_3.
\end{aligned}$$

A robust estimate of the derivative of x is given by the slope of the regression line for the last w values. A recursive formulation[3] is

$$\begin{aligned}
\delta^{(reg)}_{w,n} &= \delta^{(reg)}_{w,n-1} + \frac{w-1}{2}(x_{n-w} + x_n) -\\
&\quad (w-1)\mu_{w-1,n-1}.
\end{aligned}$$

Among other used estimates of the derivative, a smoothed version is given by the auxiliary variable

$$\begin{aligned}
\mu^{(4)}_{w,n} &= \mu^{(4)}_{w,n-1} - \mu^{(3)}_{w_6,n-w_5}/w_5 + \mu^{(3)}_{w_6,n}/w_5,\\
&\quad w_5 = \lfloor w/4\rfloor + 1, w_6 = w + 1 - w_5,
\end{aligned}$$

and

$$\delta^{(4)}_{w,n} = \mu^{(4)}_{w-1,n} - \mu^{(4)}_{w-1,n-1}.$$

Non-linear measures are the standard deviation $\sigma_{w,n}$ and the root mean energy $\sqrt{E_{w,n}}$ determined from

$$\begin{aligned}
E_{w,n} &= E_{w,n-1} - x^2_{n-w}/w + x^2_n/w,\\
\sigma^2_{w,n} &= E_{w,n} - \mu^2_{w,n}.
\end{aligned}$$

Minimum and maximum of the last w values are approximated, e. g. $min_{w,n}$ by

$$\begin{aligned}
min'_{w,o,n} &= \begin{cases} x_n & n \equiv o \bmod w\\ \min(x_n, min'_{w,o,n-1}) & \text{else}\end{cases},\\
min_{w,n,k} &= \min_{i=0\ldots k-1}(min'_{w,\lfloor w\cdot i/k\rfloor,n})
\end{aligned}$$

which yields a value between $min_{w,n}$ and $min_{\lceil w\cdot(k+1)/k\rceil,n}$ (our current choice is $k = 10$). Further, the median is approximated by

$$med_{w,n} = \begin{cases} med_{w,n-1} - c_1 \cdot \sigma_{w,n} & x_n < med_{w,n-1}\\ med_{w,n-1} & x_n = med_{w,n-1}\\ med_{w,n-1} + c_1 \cdot \sigma_{w,n} & x_n > med_{w,n-1}\end{cases}$$

with $c_1 = \frac{8}{\pi w}$ chosen such that $med_{w,w/2} = 0$ for $x_n = \mathrm{sgn}(n)$ and $\alpha \to \infty$. Other non-linear quantities use the discrete derivative $\triangle_n = x_n - x_{n-1}$ and are computed as the moving average of $|\triangle_n|$, $|\max(\triangle_n, 0)|$, $|\min(\triangle_n, 0)|$, $\triangle^2_n$, $\max^2(\triangle^2_n, 0)$ and $\min^2(\triangle^2_n, 0)$ analogously to Equation (1). In total, 50 moving features are computed.

2.3 Sliding Features

The features described above have storage requirements proportional to the sample count w of the analysis window, or a space complexity of $O(w)$. For circumstances where this is not affordable, we present approximations of the features which do not depend on the sample history, thus requiring only an amount of memory *constant* in w, i. e. they have a space complexity of $O(1)$.

The approach is based on the sliding average $\mu_{\alpha,n}$, which is why the feature set is called *sliding features*:

$$\begin{aligned}
\mu_{\alpha,n} &= \alpha \cdot \mu_{\alpha,n-1} + (1-\alpha)\cdot x_n\\
&= (1-\alpha)\sum_{i=0}^{\infty} \alpha^i x_{n-i}, \qquad (2)
\end{aligned}$$

with $0 < \alpha < 1$. $\mu_{\alpha,n}$ is the mean of the signal when weighted with an infinite, exponentially decaying window with a time constant $\tau = -1/ln(\alpha)$. As a measure for the effective window length, we

[3] For a compact notation, we give the slope multiplied by $(w^3 - w)/12$.

use the standard deviation of a rectangular and exponential weighting window and set $\alpha = 1 - 2\sqrt{3}/w$, which yields an exponentional window with the standard deviation of a rectangular window of size w. This rectangular window contains approx. 97 % of the mass of the exponential window. An update rule using (2) for recursive computation does not suffer from numerical instability. However, due to the fact that the weighting window never actually reaches zero, outlier values of a signal can corrupt the mean value for a long time. Therefore, $\mu_{\alpha,n}$ is periodically substituted by a value that would result if the exponentially decaying window function was set to zero at 99 % of its mass. Again, a second accumulator is used, increasing the computational effort by less than a constant factor of 2.

In analogy to Equation (2), approximations can be calculated for most moving features. With the auxiliary variable

$$\begin{aligned} \mu^{(2)}_{\alpha,n} &= \alpha_1 \cdot \mu^{(2)}_{\alpha_1,n-1} + (1-\alpha_1) \cdot \mu_{\alpha_1,n}, \\ &\quad \alpha_1 = 2\alpha - 1, \end{aligned}$$

smoothed derivatives are given by

$$\begin{aligned} \delta^{(2)}_{\alpha,n} &= \mu^{(2)}_{\alpha_2,n} - \mu^{(2)}_{\alpha_2,n-1}, \quad \alpha_2 = (\alpha+1)/2, \text{ and} \\ \delta^{(3)}_{\alpha,n} &= \alpha_1 \cdot \delta^{(3)}_{\alpha,n-1} + (1-\alpha_1) \cdot (\mu^{(2)}_{\alpha,n} - \mu^{(2)}_{\alpha,n-1}). \end{aligned}$$

For minimum and maximum, we use an approximation, e. g.

$$min_{\alpha,n} = \begin{cases} x_n & x_n < min_{\alpha,n-1} \\ \alpha \cdot min_{\alpha,n-1} & \\ \quad +(1-\alpha) \cdot x_n & \text{else} \end{cases}.$$

The median is approximated by

$$med_{\alpha,n} = \begin{cases} med_{\alpha,n-1} & \\ \quad -c_2 \cdot \sigma_{\alpha,n} & x_n < med_{\alpha,n-1} \\ med_{\alpha,n-1} & x_n = med_{\alpha,n-1} \\ med_{\alpha,n-1} & \\ \quad +c_2 \cdot \sigma_{\alpha,n} & x_n > med_{\alpha,n-1} \end{cases},$$

with $\sigma_{\alpha,n}$ analog to $\sigma_{w,n}$ and $c_2 = \frac{1-\alpha}{\pi/2-1}$ chosen so that $med_{\alpha,n}$ and $\mu_{\alpha,n}$ intersect at zero for $x_n = \mathrm{sgn}(n)$ and $\alpha \to 1$. In total, 44 sliding features are calculated.

3. DATA

For evaluating our approach, we use two datasets posing different classification tasks: the Augsburg database of biosignals (AuDB) which contains physiological recordings of four induced emotions [9], and the DRIVAWORK (DRIving under VArying WORKload) database which contains different stress levels [6].

AuDB has been collected by recording ECG, EMG, SC and Resp of one participant while listening to music that was chosen to induce one of the emotions joy, anger, sadness and pleasure. ECG was sampled at 256 Hz, the remaining signals at 32 Hz. The recordings have been taken in 25 separate sessions on different days. For each session and emotion, a 2-minute segment of data is available, totalling to 200 minutes of data. In [9], one feature vector is computed from each segment and classified with different classifiers. The induced emotion is recognized with accuracies around 80 % in a leave-one-session-out cross-validation. With feature selection applied, recognition rates up to 92 % are achieved.

DRIVAWORK contains recordings of ECG, EMG, SC, Temp, BVP and Resp plus audio and video recordings of participants in a simulated car-drive. ECG and EMG are sampled at 2048 Hz, the other signals at 256 Hz; the data amounts to a total of 15 hours from 24 participants. Relaxed and stressed states have been elicited by giving the participant different tasks, partly on top of a driving task. The structured design of the recordings can be used to derive coarse stress labels. In [6] it is shown that using one minute of physiological data as input, it can be recognized with an accuracy of 89 % in a subject-independent cross-validation whether a stressed or relaxed state was intended by the experimental setup for a given point in time. For classification, statistical classifiers were applied to the moving feature set of Section 2.2 computed from multiple resolutions.

A continuous estimate of the stress level that is needed for studying real-time classification is provided in DRIVAWORK by a manual annotation from three labellers. It has been created according to the *feeltrace* [1] approach by tracing the perceived stress level on a continuous scale between 0 for maximally relaxed and 1 for maximally stressed while watching and listening to the video and audio recordings of a session. The annotations of two labellers for one participant have a Pearson correlation coefficient of 0.76 on average. In [5], the mean annotated stress rating of a participant is predicted with linear regression from the moving features computed from one minute of physiological data with a Pearson correlation of 0.69 in a subject-independent cross-validation.

4. CLASSIFICATION

Once the features from each original and derived signal have been computed, they are analysed and combined by the classification system for the final recognition result. Concatenating the feature vectors $\boldsymbol{c}_j$ from all S signals would result in a high-dimensional vector which is disadvantageous for classification. Also, accounting for possible drop-out of a sensor would not be straightforward.

As a solution, we use the concept of *late fusion*, where each signal's feature vector is classified separately and the fusion is implemented by combining the classifiers' outputs. For classification, we apply a statistical approach and use Linear Discriminant Analysis to estimate the conditional probability $p(\boldsymbol{c}_j|k)$ for class k. Taking the (simplifying) assumption of statistical independence, the fusion is carried out by multiplying the class probabilities of the signals:

$$p(\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_S|k) = \prod_{j=1}^{S} p(\boldsymbol{c}_j|k).$$

Note that in this setup, drop-out of signals can trivially be accounted for by omitting the unavailable signals in the multiplication.

More elaborate classification and fusion mechanisms exist, but the goal of our experimental evaluation is to compare different feature sets. For this purpose, it is instructive to choose a fixed, simple and robust setup for classification.

5. EXPERIMENTS AND RESULTS

We compute the moving and sliding features described in Section 2 and the specialized features supplied by AuBT and test their performance for classifying the target states of AuDB and DRIVAWORK. As a baseline comparison, we also compute results with the simplest possible feature, the raw signal value x_n. The class-wise averaged recognition rates (CL) are reported.

For the moving and specialized features, we extract features from one minute of physiological data; for the sliding features, the corresponding α is used according to the rule given in Section 2.3. We simulate on-line processing by only using data from the past, i. e. the analysis window ends at the current point in time. The length of one minute is a compromise between a large context enabling a robust decision and the locality necessary for a quick response to user state changes. The frequency f of feature vector computation and classification is 0.1 Hz, i. e. the distance between two consecutive analysis windows is 10 seconds. Note that a faster rate would not increase the computational effort for the moving or sliding features; however, to limit processing time for the AuBT features, this relatively low frequency was chosen.

It should be noted that the specialized features are at a tiny disadvantage as the derived signals used for the sliding and moving features can implicitly contain some context: ECG-HR, for example, has an estimate of the heart rate at its disposal immediately at

Features	ECG	ECG-HR	EMG	SC	Resp	Resp-rate	all orig	all orig +derived
raw	25.0	48.4	28.4	33.4	23.3	52.3	34.3	72.0
AuBT	63.3	-	69.1	40.9	72.3	-	80.7	-
moving	49.1	60.3	63.6	45.9	66.6	65.1	72.0	78.4
sliding	52.7	59.3	54.4	45.0	61.0	64.3	72.4	79.6
AuBT + moving	62.3	66.0*	71.7	47.0	73.6	73.1*	81.1	83.4*

Table 1: Recognition rates [%] for the 4-emotion AuDB task with different feature sets and physiological inputs. Row "*raw*" refers to current value of the signal. Row "*AuBT*" uses the features computed by the AuBT toolbox, rows "*moving*" and "*sliding*" the features proposed in Sections 2.2 and 2.3, respectively. In the last row, AuBT and moving features are combined. In each column, results are given for specific physiological inputs. Column "all orig" combines features from the originally recorded signals ECG, EMG, SC, and Resp; "all orig + derived" additionally uses the derived signals ECG-HR and Resp-rate. Note that AuBT does not distinguish between original and derived signals. Therefore, row "*AuBT*" contains no results for derived signals (indicated by "-"), and the results in the last row marked with an asterisk (*) were obtained by combining moving features from the derived signals with AuBT features from the original signals.

the start of the analysis window while for the AuBT features, heart rate cannot be estimated until the second R-Peak in the window has been observed.

5.1 Evaluation on AuDB

The experiments on the AuDB corpus were carried out in a leave-one-session-out cross-validation to account for day-to-day variations. Table 1 lists the recognition rates that resulted for this 4-emotion classification task.

When using the trivial feature, the current value of the signal, the resulting recognition rates are close to chance (25 % CL) for the originally recorded signals (cf. row "*raw*", columns "ECG", "EMG", "Sc", "Resp" and "all orig"). For the derived signals (cf. columns "ECG-HR" and "Resp-rate"), results improve; when combining all originally recorded and derived signals (last column), 72.0 % CL results. The specialized features provided by AuDB (row "*AuBT*") reach considerably higher recognition rates for the individual signals. The combination of AuBT features (column "*all orig*") yields 80.7 % CL, which corresponds to about half the error rate of the trivial features computed on all original and derived signals.

Next, the classification performance of the newly proposed moving features (cf. row "*moving*") is compared with that of AuBT. Also the moving features are considerably better than the trivial features, but for the original signals, they do not reach the performance of the AuBT features[4]. For example, the moving features yield 63.6 % CL for EMG while AuBT yields 69.1 % CL. When applied on the derived signals, however, the moving features come close to the performance of AuBT. On ECG-HR, they reach 60.3 % CL which is near the performance of the AuBT ECG features (63.3 % CL). For Resp, this is not so evident; however, when combining the moving features from Resp and Resp-HR, 69.7 % CL are reached (not contained in Table 1), which is again near the performance of the AuBT Resp features (72.3 % CL). Combining the moving features from all original and derived signals, a CL of 78.4 % is reached which is near the performance of all AuBT features (80.7 % CL).

The sliding features are a memory-efficient approximation of the moving features. When comparing the performance of the two variants (cf. rows "*moving*" and "*sliding*"), it turns out that the overall performance is similar. For example, when using all input modalities, the sliding features yield 79.6 % CL, even a little better than the result of the moving features (78.4 % CL).

Finally, the moving features were combined with the AuBT features to see whether complementary information is coded by the feature sets (cf. last row). This seems indeed to be the case: regardless of whether the physiological inputs are used separately or in combination, joining AuBT and moving features always[5] yields better results than any of the feature sets alone. For example, when using all inputs together, the resulting recognition rate of 83.4 % clearly exceeds the performance of the moving features and AuBT features alone (78.4 % and 80.7 %, respectively).

5.2 Evaluation on DRIVAWORK

In the experiments on the DRIVAWORK corpus, the continuous stress level annotation was used to define a binary stress/non-stress classification task: the ratings were mean-variance-normalized per labeller, averaged, and then discretized with threshold zero. The resulting target classes are almost balanced (52 % stress). Due to the complexity of the task, inter-rater agreement is not too high: Cohen's κ is 0.44 for the individual labels; mean class-wise agreement is 74.7 % when comparing one labeller with the remaining labellers in turn.

AuBT currently provides specialized feature sets for ECG, EMG, SC and Resp. Our generic approach, however, can be applied to all signals available in DRIVAWORK. To account for this, experiments on DRIVAWORK were carried out for different input sets: "all-AuBT" comprising ECG, EMG, SC and Resp, and "all", additionally covering BVP and Temp. Table 2 lists the results of the experiments which have been obtained in a leave-one-subject-out cross-validation to estimate subject-independent classification performance.

As in the experiments above, the trivial feature (cf. row "*raw*") yields recognition rates close to chance (50 % CL). The AuBT features (cf. row "*AuBT*") in combination reach 72.5 % CL. Again, the moving features (cf. row "*moving*") on the original signals do considerably better (65.7 % CL when combined, column "all-AuBT orig") than the trivial features, but only when adding the derived features, results comparable to AuBT are achieved. This time however, the moving features (73.6 % CL, column "all-AuBT orig + derived") are slightly better than the AuBT features (72.5 % CL).

Apart from the signals that AuBT currently accounts for (input set "all-AuBT"), DRIVAWORK contains also recordings of BVP and Temp. Adding these inputs (last two columns), the performance of the moving features rises from 65.7 % CL to 71.2 % CL in case of the original signals (cf. column "all orig") and from 73.6 % CL to 74.5 % CL when using derived signals as well (cf. last column).

Next, the sliding features (cf. row "sliding") were studied as an approximation of the moving features. As in the evaluations on AuDB, their performance is similar. For example, when using all original and derived signals (last column), the sliding features (74.3 % CL) perform nearly equal to the moving features (74.5 % CL).

[4]SC is an exception and is utilized better by the new features: they give 45.9 % CL while the AuBT for SC only gains 40.0 % CL.

[5]with the exception of ECG, where the score 62.3 % for the combination is lower then the score 63.3 % CL for AuBT alone

Features	all-AuBT orig	all-AuBT orig + derived	all orig	all orig + derived
raw	51.6	52.9	51.7	57.8
AuBT	72.5	-	-	-
moving	65.7	73.6	71.2	74.5
sliding	65.7	73.3	71.0	74.3
AuBT + moving	71.8	73.6*	74.0*	74.6*

Table 2: Recognition rates [%] for on-line stress/non-stress recognition on the DRIVAWORK corpus with different feature sets and physiological input sets. The features are abbreviated as in Table 1. "all-AuBT orig" refers to ECG, EMG, SC and Resp. "all-AuBT orig + derived" additionally uses the derived signals ECG-HR and Resp-rate. "all orig" comprises ECG, EMG, SC, Resp, Temp and BVP. "all orig + derived" additionally uses the derived signals ECG-HR, Resp-rate, BVP-HR and PTT. The asterisk (*) refers to moving features combined with AuBT features from "all-AuBT orig".

Combining the moving and AuBT features (last row) generally improves results, but not as clearly as was the case in the AuDB evaluations. The combination of all moving features with all AuBT features gives the best observed result (74.6 % CL, last column), minimally better than the moving features alone (74.5 % CL).

6. DISCUSSION AND FUTURE WORK

In the evaluations in the previous Section, the proposed moving features perform very similar to the AuBT features with respect to classification accuracy. When using all physiological inputs, the moving features are slightly worse on the AuDB corpus, and slightly better on the DRIVAWORK database. That means they can be used as an adequate substitute for the purpose of emotion and user state classification using physiological signals. This is intriguing because although they use only a minimum of previous knowledge and specific processing for each signal, they can compete with state-of-the-art features specifically designed for each physiological signal.

The new approach has two advantages. Firstly, splitting feature extraction into a simple, signal-dependent step (computation of derived signals) and a comprehensive, generic module (computation of moving features) simplifies implementation per signal and greatly facilitates the transfer to new signals. The effectiveness of the concept of derived signals is shown by the experiments using the original signals only, where recognition rates drop substantially on both corpora. The adequacy of the generic extraction module, on the other hand, is shown by the overall equivalent classification performance of moving features and AuBT features. Finally, the easy transfer to new signals is illustrated by the experiments on DRIVAWORK where two signals (BVP and Temp) for which no specialized features are available in AuBT, can readily be utilized by our approach, improving the recognition rate.

The second advantage of our approach is that it is ideally suited for on-line classification of physiological signals in terms of efficiency. This is because the large analysis windows required for a robust decision on the one hand and a high update rate required by an on-line classification system on the other hand result in analysis windows with large overlap. For that, the chosen recursive computation is very efficient, as shown in Section 2.2. This efficiency is nice to have when doing research using a full-featured personal computer; in an embedded application, it can be crucial. Due to cost pressure in mass production, it is important to use as few hardware resources as possible in such a context. This is not only true for processing capabilities, but also for memory capacity. As an example may serve the adidas_1 running shoe which features real-time classification of the surface condition using a biomechanical signal (shoe heel compression) in order to dynamically adapt the cushioning setting — in an embedded system with 512 bytes of main memory [2]. For applications where memory is similarly expensive, we provide the sliding features, a highly memory-efficient approximation of the moving features. The experimental evaluation shows that the classification performance of moving and sliding features is very similar. Thus, the sliding features constitute an adequate solution for such application contexts.

In the evaluations on the AuDB corpus, the combination of the proposed moving features with the AuBT features lead to pronouncedly better recognition performance than using features from one of the approaches alone. This indicates that the feature sets code complementary, useful aspects of the physiological signals. In the future, we will study these differences to further improve our approach. This will include, but not be limited to the development of more derived signals for the currently studied signals. Another direction of research will be the transfer of our approach to new signals and modalities, utilizing its generic nature. Lastly, we will evaluate our approach using different target states for classification and using other physiological datasets.

7. SUMMARY

In this paper, we have presented a new approach to physiological feature extraction. The evaluation on two databases shows that the recognition performance of our features, although involving only a minimum of signal-specific processing, is competitive with state-of-the-art approaches where a dedicated feature set has been developed for each physiological modality. We illustrate how the systematic and generic nature of the proposed approach facilitates implementation and the transfer to new signals and modalities. Finally, its algorithmic formulation is shown to be efficient in terms of time and space complexity making it suitable for real-time classification and meeting the requirements of embedded applications.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder. FEELTRACE: An instrument for recording perceived emotion in real time. In *Proc. ISCA ITRW on Speech and Emotion: Developing a Conceptual Framework, Belfast*, pages 19–24, 2000.

[2] B. Eskofier, J. Hornegger, M. Oleson, I. Munson, B. Krabbe, and C. DiBenedetto. Classification of running surface on an embedded system — a digital sports example application. In *Innovationen bei der Erfassung und Analyse bioelektrischer und biomagnetischer Signale*, pages 147–150, Braunschweig und Berlin, 2008.

[3] J. Y. A. Foo, S. J. Wilson, G. R. Williams, M.-A. Harris, and D. M. Cooper. Pulse transit time changes observed with different limb positions. *Physiological Measurement*, 26(6):1093–1102, 2005.

[4] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, 2005.

[5] F. Hönig, A. Batliner, B. Eskofier, and E. Nöth. Predicting continuous stress ratings of multiple labellers from physiological signals. In *Analysis of Biomedical Signals and Images (BIOSIGNAL)*, Brno, 2008.

[6] F. Hönig, A. Batliner, and E. Nöth. Real-time recognition of the affective user state with physiological signals. In *Affective Computing and Intelligent Interaction, Doctoral Consortium*, pages 1–8, 2007.

[7] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.

[8] E. Vyzas and R. W. Picard. Offline and online recognition of emotion expression from physiological data. In *Workshop on Emotion-Based Agent Architectures at the International Conference on Autonomous Agents*, pages 135–142, Seattle, WA, USA, 1999.

[9] J. Wagner, J. Kim, and E. André. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *IEEE Int. Conf. on Multimedia & Expo (ICME 2005)*, pages 940–943, 2005.