

Feature Extraction and Selection for Emotion Recognition from EEG

Robert Jenke, *Student Member, IEEE*, Angelika Peer, *Member, IEEE*, and Martin Buss, *Fellow, IEEE*

Abstract—Emotion recognition from EEG signals allows the direct assessment of the “inner” state of a user, which is considered an important factor in human-machine-interaction. Many methods for feature extraction have been studied and the selection of both appropriate features and electrode locations is usually based on neuro-scientific findings. Their suitability for emotion recognition, however, has been tested using a small amount of distinct feature sets and on different, usually small data sets. A major limitation is that no systematic comparison of features exists. Therefore, we review feature extraction methods for emotion recognition from EEG based on 33 studies. An experiment is conducted comparing these features using machine learning techniques for feature selection on a self recorded data set. Results are presented with respect to performance of different feature selection methods, usage of selected feature types, and selection of electrode locations. Features selected by multivariate methods slightly outperform univariate methods. Advanced feature extraction techniques are found to have advantages over commonly used spectral power bands. Results also suggest preference to locations over parietal and centro-parietal lobes.

Index Terms—Emotion recognition, EEG, feature extraction, feature selection, electrode selection, machine learning

1 INTRODUCTION

CURRENT efforts in human-machine-interaction (HMI) aim at finding ways to better and more appropriately interact with computers/humans. To make HMI more natural, knowledge about the emotional state of a user is considered an important factor. Emotions are important for correct interpretation of actions as well as communication. Interest in emotion recognition from different modalities (e.g. face, posture, motion, voice) has risen in the past decade and recently gained attention in the field of brain-computer-interfaces (BCIs), which has coined the term affective BCI (aBCI) [1].

1.1 Related Work

While the reliability of the ground-truth, duration and intensity of emotions remain ongoing challenges of automatic affect recognition, several attempts have been undertaken to advance the comparatively new field of emotion recognition from electroencephalography (EEG) [2]. Early work on emotion recognition from EEG goes back as far as 1997 [3], and gained more and more attention in recent years. Typically, feature extraction and electrode selection is based on neuro-scientific assumptions. In this, spectral power in various frequency bands is often associated to emotional states, as well as coherence and phase

synchronization of pairs of electrodes [4]. Furthermore, frontal asymmetry in α band power as differentiator of valence levels has received a lot of attention in neurological studies [5].

Beside neuro-scientific assumptions, also advanced signal processing finds application in the field of aBCIs [6], which leads to a vast amount of possible features and makes it necessary to reduce dimensions for recognition tasks in order to avoid over-specification. Computational methods from the field of machine learning are applied to optimize the selection of features and electrodes w.r.t. achieved emotion estimation accuracy [4].

1.2 Problem Statement

In emotion recognition from EEG it is not generally agreed upon which features are most appropriate, and only a few works exist, which compared different features with each other. For example, Schaaff and Schultz compared two sets of features AI and AII (see Table 2) on a self recorded data set of five subjects and three emotions [7]. Different time-frequency techniques and frequency bands were evaluated by Hadjidimitriou and Hadjileontiadis on a data set with nine subjects and different levels of liking [8]. Petrantonakis and Hadjileontiadis [9] studied features in comparison to feature vectors used in [10] and [11] on a self recorded data set of 16 subjects and six emotions.

Other studies looked into finding the best electrode positions, e.g. Wang et al. applied mRMR feature selection (FS) to find electrode positions of TOP-30 features on a self recorded data set with five subjects and four emotions [12]. A channel selection method based on synchronization likelihood was tested on one subject by Ansari-asl et al. [13]. Finally, in our previous study we investigated electrode and feature selection of selected time and frequency-domain features on a self recorded data set of 16 subjects and five emotions [14].

• R. Jenke and M. Buss are with the Institute of Automatic Control Engineering, Technische Universität München, Munich, Germany.
E-mail: {rj, mb}@tum.de.

• A. Peer is with the Institute of Automatic Control Engineering, Technische Universität München, Munich, Germany, and the Institute of Advanced Studies, Technische Universität München, Munich, Germany.
E-mail: angelika.peer@tum.de.

Manuscript received 28 Feb. 2014; revised 2 July 2014; accepted 3 July 2014.
Date of publication 16 July 2014; date of current version 30 Oct. 2014.

Recommended for acceptance by S.-W.Lee.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TAFFC.2014.2339834

However, a major limitation of these approaches is that so far only a handful of features have been compared in each study. Additionally, most studies rely on a different, yet usually small data set. In this work, we aim for a complete review of feature extraction methods used for affective EEG signal analysis and a systematic comparison on one data set that allows for a qualitative evaluation of favorable features and electrodes. In doing so, we first survey techniques for feature extraction from 33 studies and then implement and evaluate them by means of different feature selection methods on a reasonably large database of 16 subjects. Thus, the contributions of this work are threefold: 1) a comprehensive review of EEG features for emotion recognition, 2) a first systematic comparison of features on *one* database using multiple FS methods to arrive at more robust results when 3) comparing them to existing findings from other studies.

The remainder of this paper is organized as follows: Section 2 reviews feature extraction methods used for emotion classification from EEG. Five different feature selection techniques are introduced in Section 3. In Section 4, we present details on the recorded data set and its evaluation. Data processing and experimental setup are described in Section 5. We give results in Section 6 with an interpretation in Section 7 and end with a brief conclusion in Section 8.

1.3 Notation

Throughout this paper the following notation is used: $\xi(t) \in \mathbb{R}^T$ denotes the vector of the time series of a single electrode, T is the number of time-samples in ξ . The time derivative of $\xi(t)$ is written as $\dot{\xi}(t)$. The representation of $\xi(t)$ in the frequency domain is given by $\Xi(f)$.

An instance of a feature of $\xi(t)$ is indicated as x . The matrix of all features is denoted by $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_F]$, where \mathbf{x}_i is the vector of all samples of one feature and F is the number of features. A feature x is considered a random variable. The set of all features is called \mathcal{X} ; a subset of features is called \mathcal{S} .

2 FEATURE EXTRACTION

In this section, we review a wide range of features relevant for emotion recognition from EEG that have been proposed in the past. A good starting point is given in a recent study by Kim et al. [4] which we extended by several important works in the field of emotion recognition from EEG (see Table 2). We generally distinguish features in time domain, frequency domain, and time-frequency domain. Features are typically calculated from the recorded signal of a single electrode, but also a few features combining signals of more than one electrode have been found in literature, which are listed at the end of this section.

2.1 Time Domain Features

Although time-domain features from EEG are not predominant, numerous approaches exist to identify characteristics of time series that vary between different emotional states.

2.1.1 Event Related Potentials (ERP)

Frantzidis et al. used amplitude and latency of ERPs (P100, N100, N200, P200, P300) as features in their study [15]. In an online application, however, it is difficult to detect ERPs related to emotions since the onset is usually unknown (asynchronous BCI).

2.1.2 Statistics of Signal

Several statistical measures have been used to characterize EEG time series [10], [12], [16], [17]. These are:

- Power: $P_\xi = \frac{1}{T} \sum_{t=1}^T |\xi(t)|^2$
- Mean: $\mu_\xi = \frac{1}{T} \sum_{t=1}^T \xi(t)$
- Standard deviation: $\sigma_\xi = \sqrt{\frac{1}{T} \sum_{t=1}^T (\xi(t) - \mu_\xi)^2}$
- 1st difference: $\delta_\xi = \frac{1}{T-1} \sum_{t=1}^{T-1} |\xi(t+1) - \xi(t)|$
- Normalized 1st difference: $\bar{\delta}_\xi = \frac{\delta_\xi}{\sigma_\xi}$
- 2nd difference: $\gamma_\xi = \frac{1}{T-2} \sum_{t=1}^{T-2} |\xi(t+2) - \xi(t)|$
- Normalized 2nd difference: $\bar{\gamma}_\xi = \frac{\gamma_\xi}{\sigma_\xi}$

The normalized first difference $\bar{\delta}_\xi$ is also known as Normalized Length Density, and captures self-similarities of the EEG signal [18].

2.1.3 Hjorth Features

Hjorth [19] developed the following features of a time series, which were used in EEG studies, e.g. [13], [20]:

- Activity: $A_\xi = \frac{\sum_{t=1}^T (\xi(t) - \mu)^2}{T}$
- Mobility: $M_\xi = \sqrt{\frac{\text{var}(\dot{\xi}(t))}{\text{var}(\xi(t))}}$
- Complexity: $C_\xi = \frac{M(\dot{\xi}(t))}{M(\xi(t))}$

Since activity is just the squared standard deviation introduced above, we omit this feature in our implementation.

2.1.4 Non-Stationary Index (NSI)

Kroupi et al. employed the NSI as a measure of complexity by analyzing the variation of the local average over time [18]. The normalized signal ξ is divided into small parts and the average of each segment is computed. The NSI is defined as the standard deviation of all means, where higher index values indicate more inconsistent local averages [21].

2.1.5 Fractal Dimension (FD)

A frequently used measure of complexity is the fractal dimension, which can be computed via several methods. For example, Sevcik's method [13], Fractal Brownian Motion [22], Box-counting [23], or Higuchi algorithm [16] were employed. The latter is known to produce results closer to the theoretical FD values than Box-counting [16] and is implemented here. In order to compute the fractal dimension FD_ξ by the Higuchi algorithm [24], the time series $\xi(t)$, $t = 1, \dots, T$ is rewritten as:

$$\left\{ \xi(m), \xi(m+k), \dots, \xi\left(m + \left\lfloor \frac{T-m}{k} \right\rfloor \cdot k\right) \right\}, \quad (1)$$

where $\lfloor \cdot \rfloor$ denotes the Gauss notation for the floor function, $m = 1, \dots, k$ is the initial time and k is the time interval. Then, k sets are calculated by:

TABLE 1
Frequency Band Ranges and Decomposition Levels of EEG
Signals Recorded at $f_s = 512$ Hz

Bandwidth (Hz)	Frequency Band	Decomposition Level
1-4 Hz	δ	A6
4-8 Hz	θ	D6
8-10 Hz	slow α	D5 (8-16 Hz)
8-12 Hz	α	
12-30 Hz	β	D4 (16-32 Hz)
30-64 Hz	γ	D3 (32-64 Hz)

$$L_m(k) = \frac{T-1}{\left[\frac{T-m}{k}\right]^2} \sum_{i=1}^{\left[\frac{T-m}{k}\right]} |\xi(m+ik) - \xi(m+(i-1)k)|. \quad (2)$$

The average value over k sets of $L_m(k)$, denoted $\langle L(k) \rangle$, has the following relationship with the fractal dimension FD_ξ :

$$\langle L(k) \rangle \propto k^{-\text{FD}_\xi}. \quad (3)$$

We obtain FD_ξ as the negative slope of the log-log plot of $\langle L(k) \rangle$ against k .

2.1.6 Higher Order Crossings (HOC)

Motivated to find an efficient and robust feature extraction method that captures the oscillatory pattern of EEG, Petrantonakis and Hadjileontiadis introduced HOC-based features [9]. Herein, a sequence of high-pass filters is applied to the zero-mean time series $Z(t)$:

$$\mathfrak{S}_k\{Z(t)\} = \nabla^{k-1} Z(t), \quad (4)$$

where ∇^k is the iteratively applied difference operator,¹ and the order $k = 1, \dots, 10$ according to [9]. The HOC sequence D_k , i.e. the resulting k features, comprises the number of zero-crossings of the filtered time series $\mathfrak{S}_k\{Z(t)\}$ by counting its sign changes.

The approach has also been successfully studied in combination with hybrid adaptive filters [25]. Preprocessing for noise reduction is implemented in this work using third level decomposition of Discrete Wavelet Transform (DWT) (see Section 2.3).

2.2 Frequency Domain Features

2.2.1 Band Power

The most popular features in the context of emotion recognition from EEG are power features from different frequency bands. This assumes stationarity of the signal for the duration of a trial. The frequency band ranges of EEG signals are varying slightly between studies. Commonly they are defined as given in the first two columns of Table 1. Alternatively, frequency bands are computed in small equal-sized bins, e.g. $\Delta f = 1$ Hz [26], or $\Delta f = 2$ Hz [7], [27], [28].

The mostly used algorithm to compute discrete fourier transform (DFT) is the fast fourier transform (FFT) applied by [3], [7], [12], [29], [30], [31], [32]. Commonly used alternatives are short-time fourier transform (STFT) [28], [33] or the

estimation of power spectra density (PSD) using Welch's method [15], [18], [20], [26], [27], [34], [35]. When averaging over all windows, STFT is considered more robust against noise. Hence, this method is adopted in the experiment described in Sections 4, 5, and 6, using a Hamming window of length 1,000 ms with no overlap as parameters. The features extracted from the resulting representation of the signal in frequency domain are: avrg. power (mean) of frequency bands and bins ($\Delta f = 2$ Hz), their minimum, maximum, and variance. Additionally, the ratio of mean band powers β/α is calculated for each channel.

2.2.2 Higher Order Spectra (HOS)

The group of frequency domain features. Also belonging to the group of frequency domain features are bispectra and bicoherence magnitudes used by Hosseini et al. [36], which are available in the *HOSA toolbox* [37]:

The bispectrum Bis represents the Fourier Transform of the third order moment of the signal, i.e.:

$$\text{Bis}(f_1, f_2) = \mathbb{E}[\mathfrak{Z}(f_1) \cdot \mathfrak{Z}(f_2) \cdot \mathfrak{Z}^*(f_1 + f_2)], \quad (5)$$

where $\mathfrak{Z}(f)$ is the Fourier Transform of the signal $\xi(t)$, $*$ denotes its complex conjugate, and $\mathbb{E}[\cdot]$ stands for the expectation operation.

Bicoherence Bic is simply the normalized bispectrum:

$$\text{Bic}(f_1, f_2) = \frac{\text{Bis}(f_1, f_2)}{\sqrt{P(f_1) \cdot P(f_2) \cdot P(f_1 + f_2)}}, \quad (6)$$

where $P(f) = \mathbb{E}[\mathfrak{Z}(f) \mathfrak{Z}^*(f)]$ is the power spectrum. Four different features are extracted for each electrode and each of 16 combinations of frequency bands, namely the sum and the sum of squares of both Bis and Bic (for details see [36]).

2.3 Time-Frequency Domain

If the signal is non-stationary, time-frequency methods can bring up additional information by considering dynamical changes.

2.3.1 Hilbert-Huang Spectrum (HHS)

Hadjidimitriou et al. compared three such methods, namely STFT based spectrogram (SPG, also used in [22]), the Zhao-Atlas-Marks (ZAM) distribution, and the Hilbert-Huang Spectrum [8]. Although results were comparable between all methods, the latter, non-linear method, showed to be more resistant against noise than SPG. We hence computed HHS for each signal, which is done via empirical mode decomposition (EMD) to arrive at intrinsic mode functions (IMFs) to represent the original signal:

$$\xi(t) = \sum_{i=1}^K \text{IMF}_i(t) + r_K(t), \quad (7)$$

where $r_K(t)$ denotes the residue which is monotonic or constant. Using the Hilbert transform of each IMF_i , the analytical signal can be described by its amplitude $A_i(t)$ and instantaneous phase $\theta_i(t)$. The derivative of $\theta_i(t)$ can be used to compute the meaningful instantaneous frequency $f_i(t) = \frac{1}{2\pi} \frac{d\theta_i}{dt}$, which yields a time-frequency representation of the amplitude $A_i(t)$. Further, the squared amplitude is

1. The backward difference $\nabla \equiv Z(t) - Z(t-1)$ is used here.

computed, and for both, the average of each frequency band is calculated as features.

2.3.2 Discrete Wavelet Transform

A more recent technique from signal processing is discrete wavelet transform, which decomposes the signal in different approximation and detail levels corresponding to different frequency ranges, while conserving the time information of the signal. The trade-off herein is made by down-sampling the signal for each level. For details, the reader is referred to [38], for example. Correspondence of frequency bands and wavelet decomposition levels depends on the sampling frequency and is given for our case of $f_s = 512$ Hz in the last column of Table 1 [39]. We implemented features from two wavelet functions as described below.

Murugappan et al. introduced feature extraction for emotion recognition from EEG via DWT using “db4” wavelet functions, extracting energy and entropy of fourth level detail coefficients (D4) corresponding to α band frequencies [11]. They extended their approach later to frequency bands β and γ , including root mean square (RMS), recursive energy efficiency (REE) and its $\log(\text{REE})$ and $\text{abs}(\log(\text{REE}))$, as given in the equations below [17], [40]:

$$\text{RMS}(j) = \sqrt{\frac{\sum_{i=1}^j \sum_{n_i} D_i(n)^2}{\sum_{i=1}^j n_i}}, \quad (8)$$

where D_i are the detail coefficients, n_i the number of D_i at the i th decomposition level, and j denotes the number of levels; and

$$\text{REE} = \frac{E_{\text{band}}}{E_{\text{total-3b}}}, \quad (9)$$

where E_{band} is the energy of a subband, and the total energy of subbands $E_{\text{total-3b}} = E_{\alpha} + E_{\beta} + E_{\gamma}$.

A different prototype wavelet was suggested by Frantzidis et al., i.e. third order biorthogonal wavelet functions [15]. Power of δ , β , and α bands are used as features.

2.4 Features Calculated from Combinations of Electrodes

2.4.1 Multichannel Complexity D2

Multi-channel D2 is introduced by Konstantinidis et al. as a non-linear measure representing the signal's complexity [41]. Unfortunately, description and citations in their paper are incomplete making it impossible to re-implement this feature.²

2.4.2 Magnitude Squared Coherence Estimate

This feature (short MSCE) represents the correspondence of two signals ξ_i and ξ_j at each frequency, taking values between 0 and 1 [42]. It is defined as:

$$C_{ij}(f) = \frac{|P_{ij}(f)|^2}{P_i(f)P_j(f)}, \quad (10)$$

where P_{ij} is the cross power spectral density and P_i and P_j are the power spectral density of ξ_i and ξ_j , respectively. In order to reduce the large amount of features resulting from all possible combinations of electrodes, C_{ij} is averaged over the frequency bands. MSCE features are equivalent to cross-correlation coefficients used in earlier studies [3], [7], [20].

Due to suggestions from neuro-scientific findings of hemispherical asymmetry related to emotion (e.g. [43]), some studies implemented features based on the combination of symmetrical pairs of electrodes. These features can be divided into differential asymmetry and rational asymmetry.

2.4.3 Differential Asymmetry

Most frequently used are differences in power bands of corresponding pairs of electrodes, computed as the difference of two features:

$$\Delta x = x_l - x_r, \quad (11)$$

where (l, r) denote the symmetric pairs of electrodes on the left/right hemisphere of the scalp [26], [32], [33], [34], [35]. Additionally, Liu and Sourina studied pairwise hemispherical differences of statistical features and Fractal Dimension [16].

2.4.4 Rational Asymmetry

In few cases, ratios of features from symmetric electrodes are computed. Brown et al. used the spectrogram to compute α -power ratios x_l/x_r for consecutive time windows of $\Delta t = 2$ s [27]. From this, they extracted kurtosis, maximum, and the number of peaks. The latter was defined by a threshold of $\text{mean}(x_l/x_r) + 2\sigma$, where σ is the standard deviation, and normalized to the length of the recording. In analogy to the differential asymmetry features, Lin et al. used the ratio of band powers in their study [33].

3 FEATURE SELECTION

The large amount of possible features makes it necessary to reduce this space in order to avoid over-specification and to make feature computation feasible online. Techniques to achieve this have been used in some of the previous studies (see last column of Table 2), but they are limited to selecting from the set of features included in each study.

Feature Selection methods can generally be divided into filter and wrapper methods (for reviews refer to [47], [48], [49]). While wrapper methods select features based on interaction with a classifier, i.e. an underlying model, filter methods are model-independent. An advantage of filters is that they usually require less computational power than wrappers and are, hence, more suitable for big data sets. We apply several state-of-the-art filter methods introduced below, both uni- and multivariate. The latter are important to find and consider interactions of features. Generally, results obtained from applying multiple FS methods are considered more robust [50].

3.1 ReliefF

This univariate technique is an extension of the *Relief* algorithm, which uses a subsample of all instances to adjust

² Attempted communication with the authors could not resolve this issue, which is why no further attention is paid to this feature.

TABLE 2
Studies on Emotion Recognition from EEG Listed with Used Features and Method to Extract Them, as well as the Number/Position of Electrodes and Methods for Feature Selection, if Applied

Author	Year	EEG Features	Method/Description	Electrodes	Feature Selection
Musha et al. [3]	1997	Cross-correlation form $\theta(5-8 \text{ Hz})$, $\alpha(8-13 \text{ Hz})$, $\beta(13-20 \text{ Hz})$	FFT	Fp1, Fp2, F3, F4, T3, T4, P3, P4, O1, O2	-
Choppin [44]	2000	α , β , θ , asym. diff., HOS, MSCE	PSD	Fp1, Fp2, F3, F4, T3, T4, P3, P4, O1, O2	heuristically
Takahashi [10]	2004	δ , θ , α , β , μ_{ξ} , σ_{ξ} , δ_{ξ} , γ_{ξ} , $\delta_{\xi}/\sigma_{\xi}$, $\gamma_{\xi}/\sigma_{\xi}$	n/a	3 elec. headband	-
Bos [32]	2006	α , β , α/β	FFT	F3, F4, Fpz	-
Ansari et al. [13]	2007	Hjorth features (Activity, Mobility, and Complexity), fractal dimension	Sevcik's method	CP5, CP6, F3, F4, Afz	synchronization likelihood
Chanel et al. [28]	2007	9 bands 4-20 Hz ($\Delta f = 2 \text{ Hz}$)	STFT	64 electrodes (IS)	fast correlation (FCBF)
Murugappan et al. [11]	2007	entropy & energy of 4th level detail coeffs (α)	DWT (db4)	63/24 electrodes (IS)	24 el.: heuristically
Horlings [20]	2008	ERD/ERS, Cross-corr., Hjorth features	PSD (Welch's method)	32 electrodes (IS)	mRMR
Khalili and Moradi [45]	2008	$\theta(4-8 \text{ Hz})$, $\alpha(10-12 \text{ Hz})$, β , γ , δ_{ξ}	FFT	64 electrodes (IS)	GA wrapper
Schaaff and Schultz [7]	2009	AI: 34 bins of freq. components (5-40 Hz) All: α , cross-correlation btw. electrodes	FFT	Fp1, Fp2, F7, F8	correlation coefficient
Li and Lu [31]	2009	different γ bands	FFT	62 electrodes (IS)	CSP, SVM wrapper
Petrantonakis and Hadji-leontiadis [9], [25]	2010	Higher Order Crossings (HOC), DWT	hybrid (adaptive) filtering (EMD & GA)	Fp1, Fp2, F3/F4	-
Khosrowabadi et al. [42]	2010	μ_{ξ} , σ_{ξ} , δ_{ξ} , γ_{ξ} , $\delta_{\xi}/\sigma_{\xi}$, $\gamma_{\xi}/\sigma_{\xi}$	elliptic band-pass filter	P3, P4, T7, T8, C3, C4, F3, F4	-
Khosrowabadi and Rahman [22]	2010	Magnitude Squares Coherence Estimate (1-35 Hz) TF spectrogram with GMM (2-30 Hz), FD	STFT, Higuchi/Box-counting/Brownian	P3, P4, T7, T8, C3, C4, F3, F4	univariate ranking
Lin et al. [33]	2010	δ , θ , $\alpha(8-13 \text{ Hz})$, $\beta(13-30 \text{ Hz})$, $\gamma(31-50 \text{ Hz})$ differential asym., rational asym.	STFT	30 electrodes (IS), 12 symmetric electr.	F-score index
Murugappan et al. [17], [40]	2010	energy, RMS, REE, LREE, ALREE, power, σ_{ξ}	DWT(db4)	64 electrodes (IS)	-
Hosseini et al. [36]	2010	bispectra and bicoherence magnitudes (Higher Order Spectra)	HOSA toolbox	Fp1, Fp2, T3, T4, Pz	GA wrapper (SVM)
Frantzidis et al. [15]	2010	$\delta(5-4 \text{ Hz})$, θ , $\alpha(8-16 \text{ Hz})$ Ampl. and latency of ERPs	DWT (3rd order, biorthogonal wavelet)	Fz, Cz, Pz	SWM wrapper
Nie et al. [30]	2011	δ , θ , $\alpha(8-13 \text{ Hz})$, $\beta(13-30 \text{ Hz})$, $\gamma(36-40 \text{ Hz})$	FFT, log band energy, feat. smoothing	62 electrodes (IS)	correlation coefficient
Kroupi et al. [18]	2011	θ , $\alpha(8-13 \text{ Hz})$, $\beta(14-29 \text{ Hz})$, $\gamma(30-47 \text{ Hz})$, normalized length density (NLD), non-stationarity index (NSI)	PSD (Welch's method)	32 electrodes (IS)	Fisher's meta analysis
Wang et al. [12]	2011	δ , θ , α , β , γ	FFT, Hanning window	128 electrodes (IS)	mRMR
Makeig et al. [46]	2011	μ_{ξ} , σ_{ξ} , δ_{ξ} , γ_{ξ} , $\delta_{\xi}/\sigma_{\xi}$, $\gamma_{\xi}/\sigma_{\xi}$	FFT	128 electrodes (IS)	CSP
Brown et al. [27]	2011	power of 8-200 Hz max, kurtosis, peaks of asym. $\alpha(t)$ ratio	spectrogram ($\Delta t = 2s$) PSD (Welch's method)	Fp1, Fp2, F3, F4, F7, F8, C3, C4	t-test, SFS
Soleymani et al. [35]	2011	avg. power 6-12 Hz ($\Delta f = 2 \text{ Hz}$) θ , slow α , β , γ , differential freq. band asymmetries	PSD (Welch's method)	32 electrodes (IS), 14 symmetric electr.	(lin. ridge reg.)
Sourina and Liu [23]	2011	Fractal Dimension (FD)	Higuchi, Box-counting	Af3, F4, Fc6	-
Liu and Sourina [29]	2012	β/α ratio, β	FFT	Af3, F7, F3, Fc5, Fc6, F4, F8, Af4, P7, P8	-
Konstantinidis et al. [41]	2012	multichannel complexity D2	Correlation Integral	19 electrodes (IS)	-

TABLE 2
(Continued)

Author	Year	EEG Features	Method/Description	Electrodes	Feature Selection
Hadjilimitriou and Hadjileontiadis [8]	2012	TF analysis of β (13-30 Hz) and γ (30-49 Hz) with rest norm.	spectrogram, ZAM, HHS	14 electrodes (IS)	-
Rozgić et al. [34]	2013	θ , slow α , α , β , γ , differential asymmetries	FFT	32 electrodes (IS)	(segment-level feature transformation)
Reuderink et al. [26]	2013	1-64 Hz ($\Delta f = 1$ Hz), diff. asym. of α	PSD (Welch's method)	32 electrodes (IS)	correlation analysis
Liu and Sourina [16]	2013	diff. asym. of statist., FD, δ , θ , α , β , γ ,	Higuchi	32/14 electrodes (IS)	univ. SVM wrapper

Unless noted differently in parenthesis, frequency bands are defined as follows: δ : 1-4 Hz, θ : 4-8 Hz, slow α : 8-10 Hz, α : 8-12 Hz, β : 12-30 Hz, γ : 30-64 Hz.

weights of each feature depending on their ability to discriminate between two classes [51]. Feature selection is done by estimating a quality weight $W(i)$ for each feature x_i , $i = 1, \dots, F$ based on the difference $\text{diff}(\cdot)$ to the nearest hit x_H and nearest miss x_M of m randomly selected instances x_k :

$$W(i) = W(i) - \text{diff}(i, x_k, x_H)/m + \text{diff}(i, x_k, x_M)/m. \quad (12)$$

To overcome the high sensitivity to noise, *ReliefF* uses n nearest hits/misses. Multiple classes are accounted for by searching for the n nearest instances from each class weighted by their prior probabilities. A MATLAB implementation is readily provided in the statistics toolbox.

3.2 Min-Redundancy-Max-Relevance (mRMR)

The most famous method using mutual information to characterize the suitability of a feature subset is called minimal-Redundancy-Maximal-Relevance (mRMR) developed by Ding and Peng [52]. Mutual Information between two random variables x and y is defined as

$$I(x; y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (13)$$

where $p(x)$ and $p(y)$ are the marginal probability density functions of x and y , respectively, and $p(x, y)$ is their joint probability distribution. If $I(x; y)$ equals zero, the two random variables x and y are statistically independent.

The multivariate mRMR method aims to optimize two criteria simultaneously: 1) Maximal-relevance criterion D , which means to maximize average mutual information $I(x_i; y)$ between each feature x_i and the target vector y . 2) Minimum-redundancy criterion R , which means to minimize the average mutual information $I(x_i; x_j)$ between two features. The algorithm finds near-optimal features using forward selection. Given an already chosen set S_k of k features, the next feature is selected by maximizing the combined criterion $D-R$:

$$\max_{x_j \in \mathcal{X} - S_k} \left[I(x_j; y) - \frac{1}{k} \sum_{x_i \in S_k} I(x_j; x_i) \right]. \quad (14)$$

A necessary step to use the toolbox provided by Peng et al. [53] is to discretize the data beforehand (we chose 20 levels).

3.3 Effect-Size (ES)-Based Feature Selection

Several test statistics have been proposed for feature selection [49]. We introduce three effect size measures from analysis of variance (ANOVA) that draw on the difference between within-class and between-class variance of a feature (set).

3.3.1 Univariate

Cohen's effect size f^2 is a generalization to more than two classes of Cohen's $d = |\frac{\mu_x^1 - \mu_x^2}{\sigma}|$ used for the statistical t -test [54]. The spread of the class means μ_x^i , computed from the samples belonging to class i , in the numerator is represented as a standard deviation σ_m . The denominator remains the pooled standard deviation σ of the populations involved. Thus,

$$f = \frac{\sigma_m}{\sigma}, \text{ where } \sigma_m = \sqrt{\frac{\sum_{i=1}^c (\mu_{\mathbf{x}}^i - \mu_{\mathbf{x}})^2}{c}} \quad (15)$$

for equal sample sizes per class. Here, $\mu_{\mathbf{x}}$ is the overall mean of a feature \mathbf{x} . The number of classes is denoted by c .

3.3.2 Multivariate

Multivariate effect size measures from corresponding multivariate ANOVA (MANOVA) in statistics are based on the eigenvalues λ_i of the generalized eigenproblem

$$\mathbf{H}\mathbf{v} = \lambda\mathbf{v}\mathbf{E}, \quad (16)$$

where \mathbf{H} and \mathbf{E} are the between and within sum of square and cross product (SSCP) matrices, respectively [55]. Since solving (16) requires the computation of the inverse of \mathbf{E} , features have to be uncorrelated. We guarantee this by a general preprocessing step explained in Section 5.

The effect size based on Wilk's Λ , which has been applied for feature selection in [56], is defined as

$$\tau^2 = 1 - \Lambda^{1/r}, \text{ where } \Lambda = \prod_{i=1}^r \left(\frac{1}{1 + \lambda_i} \right) \quad (17)$$

and simplifies to f^2 in the univariate case. The variable $r = \min(F, c - 1)$. It adjusts for the degrees of freedom and the number of eigenvalues used to compute the statistic.

Roy's statistic θ considers the largest eigenvalue only, i.e. the explained variance of the first underlying construct and is defined as

$$\theta = \frac{\lambda_1}{1 + \lambda_1}. \quad (18)$$

Since it does not need to be normalized, θ can be used directly as a measure of effect. For large eigenvalues, both statistics converge to one.

For all ES measures, sequential forward selection (SFS) of features is carried out by ranking these measures. In the multivariate cases, the feature that returns the best ES measure when added to the already chosen set of features \mathcal{S} is added.

4 DATABASE

To conduct an experiment comparing features listed in Section 2, we recorded a reasonably large database. Of special importance was to achieve a reasonably high spatial resolution of electrodes as well as to cover a broad range of emotions.

4.1 Experiment and Procedure

In total, 16 subjects (seven female, nine male) aging between 21 and 32 participated in the experiment. For each subject, the recorded data set contains eight trials of 30 s EEG recording for five different emotions (happy, curious, angry, sad, quiet), which were selected to cover a good part of the valence-arousal-dominance (VAD) space. The experiment took place in a closed environment controlled for sound and light conditions. Emotions were induced using IAPS pictures [57]. We selected pictures on the basis of dimensional ratings, that are included in the IAPS database, by

means of an in-house developed tool to find best matches around a user defined mean with minimal variance. Means of above mentioned emotions in VAD space were taken from Russell and Mehrabian [58].³ This way, a total of 32 pictures were selected for each emotion, i.e. a total of 160 emotional pictures.

Subjects were instructed on the experiment procedure⁴ and filled out a questionnaire testing for suitability (e.g. skin allergies), well-being (WHO), and trauma screening (PTSD). No subject had to be excluded. After being equipped with recording sensors, a test-run using only neutral IAPS pictures was conducted prior to the actual experiment to accustom the subjects to the protocol and the concept of the SAM questionnaire (Self-Assessment-Manikin Test [59]), which was completed after each trial. Fig. 3 shows the induction protocol implemented. Each trial started with a black screen, followed by four randomly drawn pictures from an emotion subset, each shown for 5 seconds. Every picture was used only once per subject. To capture late effects of emotional responses, recording of each trial was prolonged by 10 s showing a black screen before SAM ratings on a scale from [1; 9] were taken. Each trial ended with a neutral picture (4 s) to reduce the effect of biased transition probability based on the previous emotion [60]. In total, 40 trials were recorded for each subject, while the order of emotions was randomized.

A 64-channel EEG cap with g.tec gUSBamp and electrodes placed according to the extended 10-20 system was used for recording at 512 Hz. The amplifier was set to include a 50 Hz notch filter and a band-pass filter between 0.1–100 Hz. The effect of artifact removal, which is a common pre-processing step for EEG data, was found to be marginal and thus, was not carried out. This speeds up processing substantially, i.e. favors online recognition capability.

4.2 Evaluation of Induction

Since successful induction of emotion is a major challenge, we tested whether the average of dimensional rating over each combination of four pictures used for induction corresponds to the targeted emotion range. Although the IAPS pictures do not resemble the emotion values to the same extreme as literature suggests, all combinations overlapped with the target emotion region and were well separable. This particularly holds for the hand selected angry pictures.

For validation of induction, we compared the results of SAM tests with the targeted values of the presented picture sets and received an average correlation coefficient of $r = .545$. The second column in Table 3 lists the correlation coefficient for each subject. Its variance is very large with the extreme values of .255 and .848, which suggests a highly individual reaction to emotion stimuli. Particularly high correlation and supposedly successful induction of emotion was reached for subjects 1, 7, 10, 11, and 13. We further discuss our results with respect to the induction success in Section 6.

3. Since emotions angry and disgust overlap in VAD space, some angry pictures were chosen by hand.

4. Ethical approval was obtained from the ethics committee of the medical faculty of TUM.

4.3 Evaluation of EEG Data

Albeit careful preparation and signal checks previous to the experiment, its duration involved the possibility to loose proper contact of shifting electrodes. We therefore analyzed the signal quality of each subject. While for some subjects few electrode signals were permanently noisy over the whole duration of the experiment, for others signal problems occurred in some trials for almost all electrodes. Since these missing trials would reduce the already limited amount of samples per subject, we decided to exclude these subjects from further processing. As depicted in the third column of Table 3 noise corrupted electrodes were found for subject 1, 5, 7, 11, 16.

5 CLASSIFICATION

Pre-studies showed that it takes about 10 s for an emotion to be induced using pictures. The duration of emotions in laboratory environments is typically assumed 0.5–4 seconds [61]. Thus, time intervals between 11–15 s after emotion induction onset, i.e. the appearance of the first picture, were considered in the following analysis. Given a sampling rate of 512 Hz, this resulted in 2,048 time samples in our case.

Fig. 2 illustrates data processing for each subject. A feature matrix was generated from the EEG data of 8 trials and 5 classes, leading to 40 rows or feature vectors. Features were extracted from all 64 electrodes which results in a total of 22,881 features. The resulting number of features is indicated for each feature group. Features like MSCE, which was computed from all combinations of electrodes and frequency-ranges resulted in 12,096 features alone. Except for MSCE and asymmetry features, all features can be associated with one electrode, which allows us to draw conclusions on their importance for emotion recognition in Section 6.3. Features were z -normalized to zero mean and standard deviation equal to one. Further, we eliminated all almost identical features which yield a correlation coefficient higher than .98 in order to avoid problems with singularities, which may occur for some FS methods.

Given the relatively small number of samples for each class, leave-one-out cross validation was chosen. We evaluated each of the five listed FS methods for each subject individually. To avoid overfitting and to limit the computational effort, the maximum number of features was fixed to 200. The data were divided into eight folds with five samples each for cross-validation. Thereof, seven folds were used for feature selection and training of the classifier. In turn, the remaining fold (i.e. one sample from each class) was used for testing, until each fold was tested once. Given five classes, chance level was at 20 percent.

To evaluate and compare the proposed feature selection methods, classification was performed by means of quadratic discriminant analysis (QDA) with diagonal covariance estimates (i.e. Naive Bayes). We chose this classifier, since it is generally known as a robust baseline classifier.

6 RESULTS

As seen in Section 4 and as discussed within the community of affective computing, reliable induction remains an ongoing challenge in the field. However, given the recorded data

set and experimental results, we believe to be able to contribute some insights to the following three questions, which we discuss in the following sections: 1) How do the different feature selection techniques perform and how many features are generally required? 2) Which type of features are most successful? Can we draw conclusions w.r.t. which features are generally better across subjects? 3) Which electrodes are mostly selected? Are there commonalities/differences regarding electrode selection across feature types?

6.1 FS Methods

Table 3 lists achieved accuracies at the optimal number of features in tabular form over subjects and FS methods. The best result for each subject is highlighted in bold font. On average, multivariate methods (mRMR, ES Λ , ES θ) perform slightly better than univariate methods (ReliefF, ES f^2), as can be seen in the bottom row. In particular, the average optimal number of features using ES Λ , i.e. 44.6, is small. Less than 100 features per subject are chosen on average over all FS methods (see last column in Table 3). However, the optimal number of features varies strongly between subjects. It should also be noted that stable results, i.e. accuracy does not vary anymore starting at around 30 features for most subjects and methods.

Classification accuracy ranges subject-dependent from 25.0 to 47.5 percent, which correlates with the evaluation of induction, i.e. the SAM test results ($r = .56$). In this, however, FS using ReliefF shows almost no correlation (.05), while using ES Λ yields $r = .61$.

Confusion Matrices (presented in Table 3 and Fig. 4) summarize class-dependent results across subjects and methods. Noticeably, FS using ES Λ reveals problems recognizing emotion class *sad*, while *happy* and *curious* are mostly identified correctly. Feature selection using ES θ is more balanced across classes, i.e. the diagonal values are high compared to non-diagonal values. But with this method (and also ReliefF) emotions *curious* and *angry* are sometimes confused with each other. Most stable and evenly distributed results are obtained from feature selection using mRMR. Overall, emotion *happy* is generally better recognized than emotion *sad*. Between subjects, we see a high variance of emotions well recognized. Subjects 3, 12, and 13, for example, show good separability of only one or two emotions from the rest (see darker squares in confusion matrices). A more even separability of classes is given for subjects 4, 6, 10, and 15. Emotion *quiet* shows smallest variance in recognition rates across subjects.

6.2 Feature Usage

To investigate which features are superior to others, i.e. have been selected more frequently by FS methods and performed more successfully in emotion classification, we compute the relative frequency of each of the 162 feature types. In this, we create a histogram of feature occurrence given the features selected at the optimal number of features for each subject and FS method. Then, we normalize each bin of the histogram by dividing the occurrence of a feature type by the number of features belonging to this feature type (e.g. 64 in the case of feature type $\max \alpha$ band

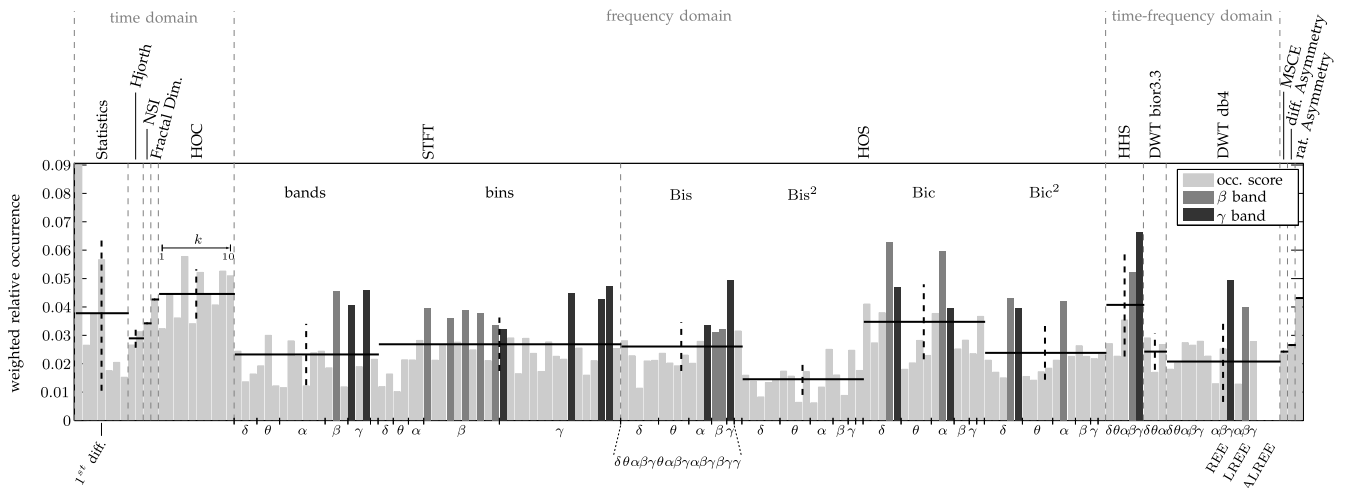


Fig. 1. Feature type usage is measured by weighted relative occurrence scores (gray bars). Mean (solid horizontal lines) and variance (dashed vertical lines) are shown for each subgroup of feature types. Feature types computed from certain frequency bands as well as individual feature types discussed in Section 6.2 are indicated below the figure.

powers, i.e. one feature for every channel) to account for random assignment of features (this is important since combinations of features can not be associated with a single electrode and, thus, are each considered one type of features). We further weighted these relative frequencies to account for the following: 1) when averaging across FS methods, multiply each relative frequency with the accuracy achieved to account for different success rates of FS methods, and 2) when averaging across subjects, multiply each subject's relative frequency with the SAM correlation coefficient to account for general reliability of the data. Given this statistic, we assume that valuable features score higher than features not important for successful emotion recognition. Fig. 1 illustrates the weighted relative occurrence of each feature type together with the average and standard deviation of subgroups of features. In general, commonalities/differences between subjects of different induction and accuracy level are not very apparent. The tendencies visible from the data are discussed in the following:

On the one hand, the most frequently selected feature group is HOC features, including a tendency to favor features with large values for k . Also, features measuring complexity of the time sequence, i.e. fractal dimension, NSI, and Hjorth's Complexity, score well. The most valuable feature type among the combinations of electrodes are the rational asymmetry features. Other (sub)-groups, such as HHS and HOS bicoherence features, yield high mean values on the weighted relative occurrence, too, but show a higher variance than the HOC group. This means that particular feature types of this group are more valuable than others.

Taking a closer look at the feature types reveals that especially features computed from frequency bands β and γ are successfully selected more often than other bands (bars marked darker in Fig. 1). This holds for feature groups STFT, HOS, HHS and DWT db4 features REE and LREE.

On the other hand, results suggest that feature subgroups such as HOS squared bispectra and DWT db4 are less efficient for emotion recognition from EEG, since

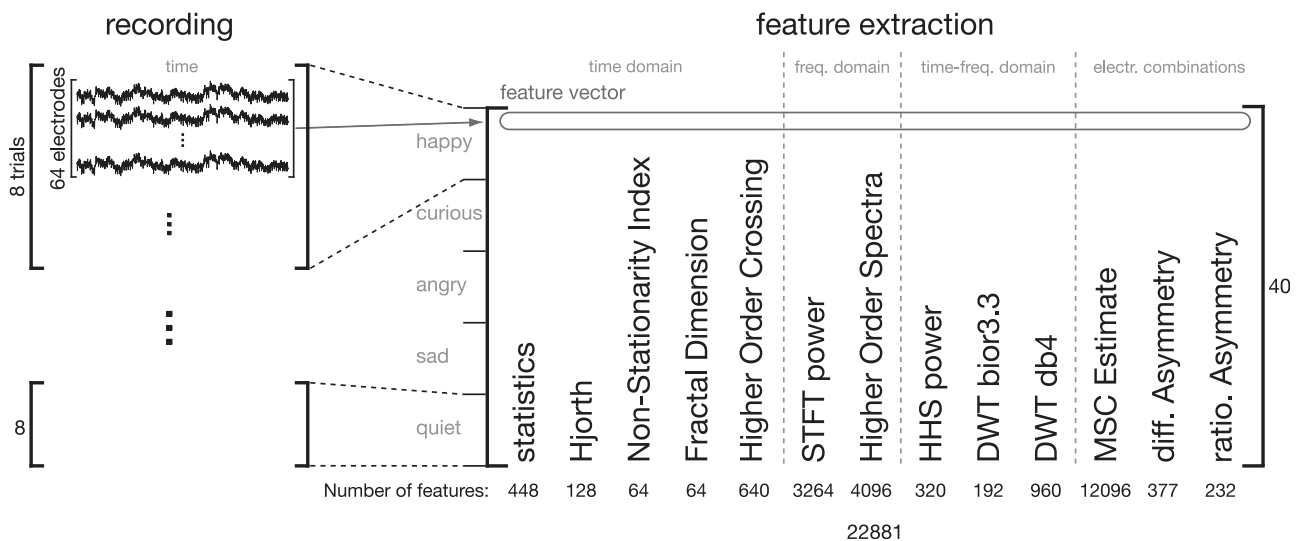


Fig. 2. Schematic characterisation of data processing from trial-wise recording to feature extraction.

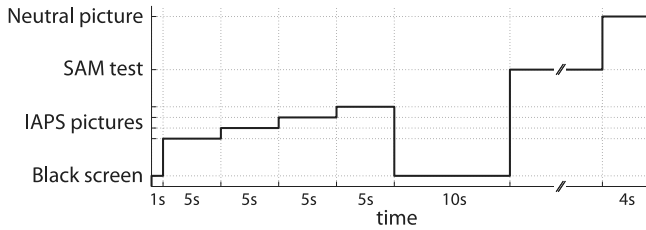


Fig. 3. The emotion induction protocol for each trial consisting of the set of IAPS pictures, prolonged recording time, SAM test, and a neutral picture.

their weighted relative occurrence scores are low. Feature ALREE apparently has no advantage over LREE, since it has never been selected. Both the prevalent frequency band powers and equal sized bins show relatively low scores on average.

The group of statistical features affords a differentiation between feature types, since the proposed statistics are very different from each other. Except for the first difference of the time series, none reveals a very high score on the weighted relative frequencies. A special note regarding the high score of the statistical feature power of signal should be given: This does not mean to be a highly valued feature, but has to be explained by a very high occurrence when using FS method ES θ : the first feature is selected in case of equal ranking scores or no measurable improvement through any feature (saturation of selection criterion).

6.3 Electrode Usage

Based on the findings of Section 6.2, we investigated also electrode usage of feature types and groups that can be associated with a single electrode, i.e. all except MSCE and asymmetry features. In this section, we discuss the electrode usage of the six features identified as most important. These are first difference, complexity features,

the group of HOC features and bicoherence features, and selected (threshold at .03) β , and γ features (i.e. from STFT, bispectrum, absolute and squared bicoherence, and $REE(\gamma)/LREE(\beta)$) as shown in Fig. 5. Here, the darkness of each electrode depicts the number of its occurrences within a feature group, where the scale is normalized by the total number of feature occurrences in the group, so that the scores of all 64 electrodes sum to one (since the first plot only captures electrode selections of one feature type, i.e. 64 features, we expect the results to be more pronounced here).

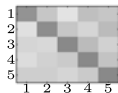
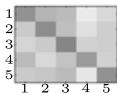
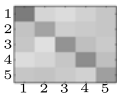
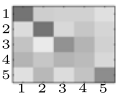
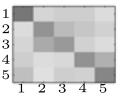
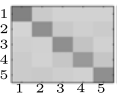
For the feature type first difference δf the prominent electrodes are P9 and P5 with some importance given to P7, P1, and FC2. The three complexity features Hjorth's complexity, NSI, and fractal dimension are mostly drawn upon from electrode locations CP2, CP1, P1, and F5. For the feature group of HOC the important electrodes are CP1 and CP2, but also some right hemisphere locations like T8 and P10 are frequently used. Features for bicoherence are mostly selected from C2, FC2, and Cz. A cluster of electrodes selected for good γ features forms around CP5, CP3, CP1, P5, and P3. Additionally, location CP6 is selected often. Not quite as clear are location of good β features. Prominent electrodes are C2, Cz, P9, F7, CP5, and CP3. Generally, electrodes located over the parietal and central-parietal lobe are favored over occipital and frontal lobes.

7 DISCUSSION

In this section, we interpret and discuss the same underlying questions posed above using results presented in Section 6.

The higher average accuracy of multivariate methods points towards interactions of features that only multivariate FS methods can actively detect and capitalize upon.

TABLE 3
Pair-Wise Correlation Coefficients between IAPS and SAM Ratings, Signal Quality and Accuracy (Optimal Number of Features) for Different FS Methods Listed for Each Subject

Subject	Correlation coefficient	Signal Quality	mRMR	Relieff	ES f^2	ES Λ	ES θ	Average
1	.748	×	-	-	-	-	-	-
2	.588	✓	35.0% (84)	37.5% (56)	32.5% (66)	37.5% (25)	32.5% (131)	35.0% (72.4)
3	.524	✓	27.5% (196)	37.5% (17)	42.5% (4)	32.5% (5)	35.0% (7)	35.0% (45.8)
4	.711	✓	35.0% (82)	35.0% (29)	30.0% (10)	35.0% (31)	27.5% (4)	32.5% (31.2)
5	.447	×	-	-	-	-	-	-
6	.589	✓	45.0% (30)	37.5% (83)	37.5% (31)	35.0% (23)	35.0% (3)	38.0% (34.0)
7	.848	×	-	-	-	-	-	-
8	.573	✓	35.0% (122)	27.5% (12)	40.0% (49)	35.0% (138)	37.5% (37)	35.0% (71.6)
9	.490	✓	32.5% (19)	37.5% (132)	30.0% (149)	30.0% (74)	30.0% (78)	32.0% (90.4)
10	.748	✓	35.0% (192)	42.5% (75)	32.5% (121)	45.0% (92)	40.0% (18)	39.0% (99.6)
11	.794	×	-	-	-	-	-	-
12	.511	✓	37.5% (70)	32.5% (49)	25.0% (149)	37.5% (6)	42.5% (195)	35.0% (93.8)
13	.796	✓	45.0% (4)	37.5% (3)	45.0% (183)	42.5% (40)	45.0% (151)	43.0% (76.2)
14	.255	✓	35.0% (10)	40.0% (40)	30.0% (113)	35.0% (16)	32.5% (180)	34.5% (71.8)
15	.550	✓	40.0% (106)	25.0% (67)	35.0% (133)	35.0% (35)	47.5% (109)	36.5% (90.0)
16	.295	×	-	-	-	-	-	-
Average:			36.5% (87.1)	35.4% (68.8)	34.5% (82.1)	36.3% (44.6)	36.8% (82.6)	35.9% (70.6)
Legend: 1 - happy 2 - curios 3 - angry 4 - sad 5 - quiet								

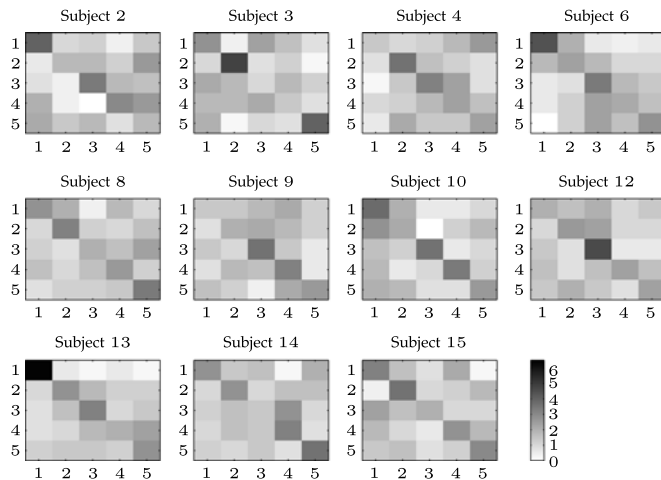


Fig. 4. Confusion matrices summarizing the targeted (y -axis) and predicted (x -axis) emotion class are depicted for each subject individually.

High variance in measured induction success as well as emotion recognition accuracy between subjects, however, reassures the need for ongoing efforts towards a more reliable ground truth.

Spectral power in nearly all frequency bands is used in most studies, which contrasts with its low performance scores in our study. Among these features, Wang et al. found frontal and parietal locations typical under their TOP-30 features [12]. However, features related to β and γ bands seem more valuable if obtained by different (mostly more complex) extraction methods which have found only little attention so far. Moreover, our study revealed that electrode locations over parietal and centro-parietal lobes can be of advantage for these bands. Although the difference between the two STFT feature subgroups is small, the trend of equal sized bins scoring higher than frequency bands is in accordance with earlier argumentation by Reuderink et al. [26], that emotional responses can be visible in small frequency changes. The finding of β and γ bands from HHS being valuable features is in line with findings of Hadjidimitriou and Hadjileontiadis, who found these bands to discriminate best between liking and disliking judgements [8].

The usage of several complexity measures (especially FD) by many of the existing studies is endorsed by our results. The suggested electrode locations over the left anterior scalp as well as central-posterior locations are in line with those suggested in a study by Ansari-asl et al. [13].

A promising, yet little considered feature is HOC. Contrary to the frontal electrode used by Petrantonakis and Hadjileontiadis, centro-parietal locations might enhance this feature type.

To date, frontal regions are of high importance. They are used in almost every study, especially when only a few electrodes are hand-selected. On the one hand, our results opposing this trend might be biased since we did not use artifact removal in this study. On the other hand, recent studies reach consensus that frontal alpha asymmetry primarily reflects approach versus withdrawal motivation rather than the level of valence [62]. This may imply that locations other than the anterior scalp are better to differentiate multiple, discrete emotional states.

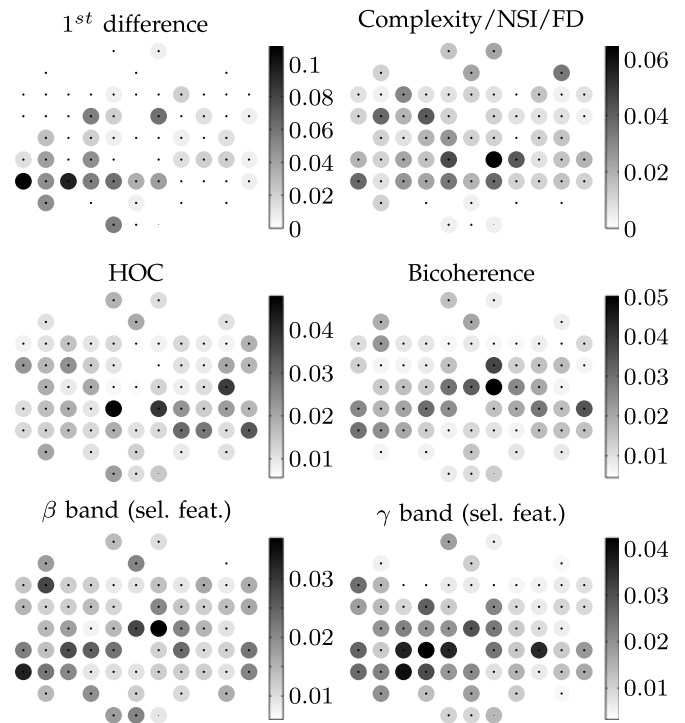


Fig. 5. Electrode usage of six feature (sub-)groups is indicated by the darkness of the circle around each electrode location. Rows correspond to letters, columns to numbers of the extended 10-20 system.

Finally, rational asymmetry seems to have advantages over differential asymmetry, although the latter is used more frequently in the surveyed studies.

When interpreting these results it should be noted that due to the high number of tested features compared to the low number of samples, we could not validate statistical significance of our results.

8 CONCLUSION

Different sets of features and electrodes for emotion recognition from EEG have been suggested by over 30 studies reviewed in this paper. We presented a systematic analysis and first qualitative insights comparing the wide range of available feature extraction methods using machine learning techniques for feature selection. Multivariate feature selection techniques performed slightly better than univariate methods, generally requiring less than 100 features on average. We also investigated which type of features are most promising and which electrodes are mostly selected for them. Advanced feature extraction methods such as HOC, HOS, and HHS were found to outperform commonly used spectral power bands. Results suggest preference to locations over parietal and centro-parietal lobes.

As shown by Mühl et al., different induction modalities result in modality-specific EEG responses [1]. This possibly limits our results to the induction modality of visual stimuli. We expect that repeated experiments of the above procedure on other databases will help to reach a consensus on this topic.

Another compelling question to answer in the future is whether certain types of features work better together than by themselves.

Due to the strong interpersonal variance, multi-individual studies are necessary in order to make emotion recognition from EEG feasible for applications. Therefore, more baseline data sets like the recently published DEAP [63] need to be made available to the community in order to compare approaches.⁵

ACKNOWLEDGMENTS

This work was supported in part by the VERE project within the seventh Framework Programme of the European Union, FET- Human Computer Confluence Initiative, contract number ICT-2010-257695 and the Institute of Advanced Studies of the Technische Universität München.

REFERENCES

- [1] C. Mühl, A.-M. Brouwer, N. C. van Wouwe, E. van den Broek, F. Nijboer, and D. Heylen, "Modality-specific affective responses and their implications for affective BCI," in *Proc. Int. Conf. Brain-Comput. Interfaces*, 2011, pp. 4–7.
- [2] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," *Int. J. Synth. Emot.*, vol. 1, no. 1, pp. 68–99, Jan. 2010.
- [3] T. Musha, Y. Terasaki, H. Haque, and G. Ivamitsky, "Feature extraction from EEGs associated with emotions," *Artif. Life Robot.*, vol. 1, no. 1, pp. 15–19, Mar. 1997.
- [4] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human EEG," *Comput. Math. Methods Med.*, vol. 2013, pp. 1–13, Jan. 2013.
- [5] J. T. Cacioppo, "Feelings and emotions: Roles for electrophysiological markers," *Biol. Psychol.*, vol. 67, no. 1–2, pp. 235–43, Oct. 2004.
- [6] S. Sanei and J. Chambers, *EEG Signal Processing*. New York, NY, USA: Wiley, 2007.
- [7] K. Schaaff and T. Schultz, "Towards emotion recognition from electroencephalographic signals," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Sep. 2009, pp. 175–180.
- [8] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "Toward an EEG-based recognition of music liking using time-frequency analysis," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3498–510, Dec. 2012.
- [9] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.
- [10] K. Takahashi, "Remarks on emotion recognition from multimodal bio-potential signals," in *Proc. Int. Conf. Ind. Technol.*, 2004, pp. 1138–1143.
- [11] M. Murugappan, M. Rizon, S. Yaacob, I. Zunaidi, and D. Hazry, "EEG feature extraction for classifying emotions using FCM and FKM," *Int. J. Comput. Commun.*, vol. 1, no. 2, pp. 21–25, 2007.
- [12] X. Wang, D. Nie, and B. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in *Proc. Int. Conf. Neural Inf. Process.*, 2011, pp. 734–743.
- [13] K. Ansari-asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in *Proc. 15th Eur. Signal Process. Conf.*, 2007, pp. 1241–1245.
- [14] R. Jenke, A. Peer, and M. Buss, "Effect-size-based electrode and feature selection for emotion recognition from EEG," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2013, pp. 1217–1221.
- [15] C. Frantzidis, C. Bratsas, C. Papadelis, E. Konstantinidis, C. Pappas, and P. Bamidis, "Toward emotion aware computing: An integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 589–597, May 2010.
- [16] Y. Liu and O. Sourina, "Real-time fractal-based valence level recognition from EEG," *Trans. Comput. Sci. XVIII*, vol. 7848, pp. 101–120, 2013.
- [17] M. Murugappan, R. Nagarajan, and S. Yaacob, "Classification of human emotion from EEG using discrete wavelet transform," *J. Biomed. Sci. Eng.*, vol. 3, no. 4, pp. 390–396, 2010.
- [18] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2011, pp. 457–466.
- [19] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.
- [20] R. Horlings, D. Datcu, and L. Rothkrantz, "Emotion recognition using brain activity," in *Proc. Int. Conf. Comput. Syst. Technol.*, 2008, pp. II.1–1–6.
- [21] J. Hausdorff, A. Lertratanakul, M. Cudkiewicz, A. Peterson, D. Kaliton, and A. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J. Appl. Physiol.*, vol. 88, pp. 2045–2053, 2000.
- [22] R. Khosrowabadi and A. Rahman, "Classification of EEG correlates on emotion using features from Gaussian mixtures of EEG spectrogram," in *Proc. 3rd Int. Conf. Inf. Commun. Technol. Moslem World*, 2010, pp. E102–E107.
- [23] O. Sourina and Y. Liu, "A fractal-based algorithm of emotion recognition from EEG using arousal-valence model," in *Proc. Biosignals*, 2011, pp. 209–214.
- [24] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.
- [25] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, pp. 81–97, Jul.–Dec. 2010.
- [26] B. Reuderink, C. Mühl, and M. Poel, "Valence, arousal and dominance in the EEG during game play," *Int. J. Auton. Adaptive Commun. Syst.*, vol. 6, no. 1, pp. 45–62, 2013.
- [27] L. Brown, B. Grundelner, and J. Penders, "Towards wireless emotional valence detection from EEG," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, Jan. 2011, pp. 2188–2191.
- [28] G. Chanel, K. Ansari-asl, T. Pun, and K. A. Asl, "Valence-arousal evaluation using physiological signals in an emotion recall paradigm," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2007, pp. 2662–2667.
- [29] Y. Liu and O. Sourina, "EEG-based dominance level recognition for emotion-enabled interaction," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2012, pp. 1039–1044.
- [30] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in *Proc. IEEE Int. Conf. Neural Eng.*, 2011, pp. 667–670.
- [31] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, Sep. 2009, vol. 1, pp. 1323–1326.
- [32] D. O. Bos, "EEG-based emotion recognition," *Emotion*, vol. 57, no. 7, pp. 1798–806, 2006.
- [33] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.
- [34] V. Rozgić, S. Vitaladevuni, and R. Prasad, "Robust EEG emotion classification using segment level decision fusion," in *Proc. Int. Conf. Acoustics, Sound, Signal Process.*, 2013, pp. 1286–1290.
- [35] M. Soleymani, S. Koelstra, I. Patras, and T. Pun, "Continuous emotion detection in response to music videos," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Mar. 2011, pp. 803–808.
- [36] S. Hosseini, M. Khalilzadeh, M. Naghibi-Sistani, and V. Niazmand, "Higher order spectra analysis of EEG signals in emotional stress states," in *Proc. IEEE Int. Conf. Inform. Technol. Comput. Sci.*, Jul. 2010, pp. 60–63.
- [37] A. Swami, J. Mendel, and C. Nikias (2000). Higher-order spectra analysis (HOSA) toolbox, version 2.0.3 [Online]. Available: www.mathworks.com/matlabcentral/fileexchange/3013/
- [38] M. Akin, "Comparison of wavelet transform and FFT methods in the analysis of EEG signals," *J. Med. Syst.*, vol. 26, no. 3, pp. 241–247, Jun. 2002.
- [39] C. Parameswariah and M. Cox, "Frequency characteristics of wavelets," *IEEE Trans. Power Delivery*, vol. 17, no. 3, pp. 800–804, Jul. 2002.
- [40] M. Murugappan, M. Rizon, R. Nagarajan, and S. Yaacob, "Inferring of human emotional states using multichannel EEG," *Eur. J. Sci. Res.*, vol. 48, no. 2, pp. 281–299, 2010.

5. Our database will be made available to other researchers via the authors' webpage at www.lsr.ei.tum.de.

- [41] E. Konstantinidis, C. Frantzidis, C. Pappas, and P. Bamidis, "Real time emotion aware applications: A case study employing emotion evocative pictures and neuro-physiological sensing enhanced by graphic processor units," *Comput. Methods Programs Biomed.*, vol. 107, no. 1, pp. 16–27, Jul. 2012.
- [42] R. Khosrowabadi, H. Quek, A. Wahab, and K. Ang, "EEG-based emotion recognition using self-organizing map for boundary detection," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 4242–4245.
- [43] L. Schmidt and L. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cogn. Emot.*, vol. 15, no. 4, pp. 487–500, Jul. 2001.
- [44] A. Choppin, "EEG-based human interface for disabled individuals: Emotion expression with neural networks," Master's thesis, Dept. Info. Process., Tokyo Inst. Technol., Tokyo, Japan, 2000.
- [45] Z. Khalili and M. Moradi, "Emotion detection using brain and peripheral signals," in *Proc. IEEE Int. Biomedical Eng. Conf.*, 2008, pp. 1–4.
- [46] S. Makeig, G. Leslie, T. Mullen, D. Sarma, N. Bigdely-Shamlo, and C. Kothe, "First demonstration of a musical emotion BCI," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2011, pp. 487–496.
- [47] P. Somol, J. Novovicová, and P. Pudil, "Efficient feature subset selection and subset size optimization," in *Pattern Recognition Recent Advances*, A. Herout, Ed. Rijeka, Croatia: InTech, 2010, pp. 1–24.
- [48] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. London, U.K.: Chapman & Hall, 2008.
- [49] Y. Saeyns, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Oct. 2007.
- [50] C. Lazar, J. Taminiau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, and A. Nowé, "A survey on filter techniques for feature selection in gene expression microarray analysis," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 9, no. 4, pp. 1106–1119, Jul.–Aug. 2012.
- [51] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learn.*, vol. 53, pp. 23–69, 2003.
- [52] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinformatics Comput. Biol.*, vol. 3, no. 2, pp. 185–205, Apr. 2003.
- [53] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [54] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Evanston, IL, USA: Routledge, 1988.
- [55] C. J. Huberty and S. Olejnik, *Applied MANOVA and Discriminant Analysis*, 2nd ed. New York, NY, USA: Wiley, 2006.
- [56] A. El Ouardighi, A. El Akadi, and D. Aboutajdine, "Feature selection on supervised classification using Wilk's Lambda statistic," in *Proc. Int. Symp. Comput. Intell. Informatics*, 2007, pp. 51–55.
- [57] P. Lang, M. Bradley, and B. Cuthbert, "International affective picture system (IAPS): Affective rating of pictures and instruction picture system (IAPS): Affective rating of pictures and instruction manual," Univ. Florida, FL, USA, Tech. Rep. A-8, 2008.
- [58] J. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Res. Personal.*, vol. 11, pp. 273–294, 1977.
- [59] M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment Manikin and the semantic differential," *J. Behav. Theory Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- [60] M. Karg, K. Kühnlenz, and M. Buss, "Towards a system-theoretic model for transition of affect," in *Proc. Symp. AISB Conv.-Mental States, Emotions Their Embodiment*, 2009, pp. 24–31.
- [61] R. W. Levenson, "Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity," in *Social Psychophysiology and Emotion: Theory and Clinical Applications*, H. Wagner, Ed. New York, NY, USA: Wiley, 1988, pp. 17–42.
- [62] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cogn. Emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [63] S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 18–31, Jan.–Mar. 2012.



Robert Jenke received the Diploma Engineer degree in electrical engineering and information technology from Technische Universität München (TUM) in 2010. He is currently working toward the PhD degree at the Institute of Automatic Control Engineering, TUM, on a project aimed to design robotic systems to re-embodiment paralyzed patients, supervised by A. Peer. His research interests include emotion recognition using machine learning methods. He is a student member of the IEEE.



human-object interaction as well as human motor control. She is a member of the IEEE.

Angelika Peer is a senior researcher and lecturer at the Institute of Automatic Control Engineering of the Technische Universität München (TUM), TUM-IAS junior fellow at the Institute of Advanced Studies, and leader of a Haptic Interaction Group. She received the Diploma Engineer degree in electrical engineering and information technology in 2004 and the Doctor of Engineering degree in 2008 from TUM, Munich, Germany. Her research interests include robotics, haptics, teleoperation, human-human and



Martin Buss received the Diploma Engineer degree in electrical engineering in 1990 from the Technical University Darmstadt, Germany, and the Doctor of Engineering degree in electrical engineering from the University of Tokyo, Japan, in 1994. In 2000, he finished his habilitation in the Department of Electrical Engineering and Information Technology, Technische Universität München (TUM), Germany. From 1995 to 2000, he has been senior research assistant and lecturer at the Institute of Automatic Control Engineering, TUM, Germany. He has been appointed as a full professor, the head of the control systems group, and a deputy director of the Institute of Energy and Automation Technology, Technical University Berlin, Germany, from 2000 to 2003. Since 2003, he has been a full professor (chair) at the Institute of Automatic Control Engineering, TUM, Germany. Since 2006, he has been the coordinator of the DFG Excellence Research Cluster Cognition for Technical Systems. His research interests include automatic control, mechatronics, multimodal human-system interfaces, optimization, nonlinear, and hybrid discrete-continuous systems. He is a fellow of the IEEE.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.