

Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources

Hanna Becker, Julien Fleureau, Philippe Guillotel, Fabrice Wendling, Isabelle Merlet,
Laurent Albera *Senior Member, IEEE*

Abstract—Electroencephalography (EEG)-based emotion recognition is currently a hot issue in the affective computing community. Numerous studies have been published on this topic, following generally the same schema: 1) presentation of emotional stimuli to a number of subjects during the recording of their EEG, 2) application of machine learning techniques to classify the subjects' emotions. The proposed approaches vary mainly in the type of features extracted from the EEG and in the employed classifiers, but it is difficult to compare the reported results due to the use of different datasets. In this paper, we present a new database for the analysis of valence (positive or negative emotions), which is made publicly available. The database comprises physiological recordings and 257-channel EEG data, contrary to all previously published datasets, which include at most 62 EEG channels. Furthermore, we reconstruct the brain activity on the cortical surface by applying source localization techniques. We then compare the performances of valence classification that can be achieved with various features extracted from all source regions (source space features) and from all EEG channels (sensor space features), showing that the source reconstruction improves the classification results. Finally, we discuss the influence of several parameters on the classification scores.

Index Terms—EEG, emotion recognition, source localization, functional connectivity.

1 INTRODUCTION

THE automatic recognition of human emotions is of great interest in the context of multimedia applications and brain-computer interfaces [1], [2]. For instance, knowing the emotional state of the user (e.g., during a computer game), the behavior of the computer could be adapted accordingly (to make the game easier or more complicated). Another example concerns user-specific content recommendations, where the knowledge of the user's emotions in response to certain stimuli such as music or videos could be employed to select other content that might be of interest to the user.

Emotions can be characterized based on two different models: the discrete model, which distinguishes a fixed number of basic emotions (anger, fear, surprise, joy, sadness, disgust) [3], and the dimensional model, which describes the emotions in a 2- or sometimes 3-dimensional space with dimensions valence, arousal, and dominance [4]. In this paper, we adopt the two-dimensional emotion model. The first dimension, valence, discriminates emotions on a negative-positive scale. The second dimension, arousal, describes emotions on a scale ranging from calm to excited. To detect emotions, different approaches have been considered in the literature [5], including the analysis of speech or facial expressions on the one hand [6] and the processing of physiological measurements on the other hand [7], [8], [9]. Similar to self-assessment of the felt emotions by the subjects, which may be biased by social expectations, the

facial expressions may also be influenced by the subjects, who can try to mask their feelings. Therefore, we concentrate here on emotion recognition based on physiological recordings, which may yield more objective measures of the emotional state of the subjects. Furthermore, we focus on the analysis of emotions which are evoked by audio-visual stimuli. Previous research in this context has considered galvanic skin response (GSR), heart rate, and electromyogram (EMG) measures [7], [8] as well as electroencephalography (EEG) [9], [10]. While in particular the GSR has been shown to give a good indication of arousal, it does not reveal any information on the valence of the emotion. Therefore, our primary objective is to determine the valence of the evoked emotions. To achieve this, we explore in this article the use of EEG for valence classification in order to complement the GSR arousal classification.

1.1 Previous work

EEG emotion recognition is currently a hot issue in the affective computing community. Starting with a few studies [11], [12] more than a decade ago, the number of papers published on this topic has been exploding in recent years. This progress is well documented in recent review papers [13], [14]. In addition, newer studies can, e.g., be found in [15], [16], [17], [18]. A good introduction to the topic can also be found in [16]. The reported studies usually include two steps: the data acquisition and preprocessing step and the feature extraction and classification step.

1.1.1 Data acquisition and preprocessing

In general, the studies on EEG emotion recognition begin with the collection of EEG recordings from a certain number

• H. Becker, J. Fleureau, and P. Guillotel are with Technicolor R&D France, 975 Avenue des Champs Blancs, 35576 Cesson-Sévigné, France.

• L. Albera, I. Merlet, and F. Wendling are with INSERM, U1099, Rennes, F-35000, France and Université de Rennes 1, LTSI, Rennes, France.

of subjects which are acquired during different emotional states. Most studies are based on EEG recordings from a small number (between 5 and 32) of subjects. The data have been acquired using EEG systems with 3 to 65 sensors, comprising both commercial EEG devices such as the Emotiv EPOC with 14 sensors and clinical EEG systems with 19, 32, or 65 electrodes. There is only one study which is based on data from 110 subjects [19] and one study [20] that used a High Resolution (HR) EEG system with 250 channels, but only 124 sensors have been considered in the analysis. The emotional states are elicited by presenting selected stimuli of various types. The most commonly employed stimuli are pictures or videos, but some researchers have also considered other ways of eliciting emotions such as asking the subjects to recall memories or to imagine situations associated with a particular emotion [20]. While many studies based on picture stimuli make use of the publicly available International Affective Picture System (IAPS) [21], there is no gold standard for the employed audio-visual stimuli, which are selected from varying sources (music clips [9], film excerpts [15], excerpts from films and documentaries [15]) and are often not publicly available. After the acquisition of the EEG data, the measurements are preprocessed in order to reduce artifacts. This is classically achieved by applying bandpass filtering and Independent Component Analysis (ICA), where artifactual components (containing eye blinks) are removed (see, e.g., [9], [22]).

Several of the resulting EEG datasets for emotion analysis are accessible upon demand. Among these databases, one can find the ENTERFACE'06 [23] dataset, comprising 54-channel EEG recordings from 5 subjects presented with IAPS stimuli, the MAHNOB-HCI [10] database, including 32-channel EEG recordings of 27 subjects watching affective video clips, and the DEAP [9] dataset, which contains 32-channel EEG data of 32 subjects watching music video clips. At the time of writing this paper, we also became aware of the recently published SEED database (first described in [24]), comprising EEG data recorded with 62 active electrodes from 15 subjects, each of which participated in several sessions of watching emotional video clips.

1.1.2 Feature extraction and classification

The second step of EEG emotion recognition methods consists in extracting features from the EEG data of each channel. In the literature, different types of features have been considered [14], [25], most of which can be associated with one of the following classes: time domain features, frequency domain features, time-frequency features, and functional connectivity features. In the time domain, features like classical statistical parameters, autocorrelation coefficients, fractal dimension, non-stationarity index, and other parameters characterizing the time signal can be extracted. Frequency domain features include the powers in different frequency bands, power ratios, hemispheric differences or ratios of power (differential or rational asymmetry features), differential entropy features, spectral moments, higher order spectra, and spectral shape descriptors. Time-frequency features classically correspond to wavelet coefficients, but other time-frequency transformations have also been considered. Finally, the connectivity features comprise different measures of functional connectivity such as correlation or

phase synchronization between the time signals of pairs of sensors. The most prevalently used features are the signal powers in the classical α , β , γ , and θ frequency bands.

A classifier is then trained using the features extracted from one part of the available data. Among the classifiers commonly used in the literature, such as Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Linear Discriminant Analysis (LDA), the SVM is the most widespread. Other classification approaches like artificial neural networks (ANN) [17] have also been explored. The classifier is tested on the rest of the data, called the test set, and the classification performance is evaluated employing a cross-validation scheme.

1.2 Contributions

Despite the resemblance of the approaches taken by different studies, the number of proposed features, employed stimuli and considered datasets is almost as vast as the number of studies. This makes it very difficult to evaluate and to compare the results reported in different articles. In this context, we make the following three contributions:

- 1) We provide a new, publicly accessible database¹ of HR-EEG recordings for the analysis of emotions elicited by film excerpts as well as other simultaneously recorded physiological measurements: GSR, Electrocardiogram (ECG), respiration, peripheral oxygen saturation (SpO_2), and pulse rate.
- 2) We compare classification results obtained using a number of features from the literature on our EEG database.
- 3) We consider additional features which are extracted after reconstruction of the activity on the cortical surface and compare the results with those obtained based on features extracted directly from the EEG channels.

To the best of our knowledge, our study is the first to consider high-resolution EEG data from up to 257 electrodes. Even though [20] also recorded 250-channel data, they limited their analysis to 124 channels. While inverse modeling has been sporadically employed in the context of feature extraction for BCI [26], [27], to the best of our knowledge, our study is the first in the field of EEG emotion recognition to consider features extracted from the reconstructed brain signals on the cortical surface, thus taking into account physiological information on the underlying mechanisms of the brain.

This paper is organized as follows: in Section 2, we describe the experiment conducted to construct the HR-EEG database as well as the data processing steps, the types of features, and the classification approach that we consider in this paper. Section 3 presents and discusses the classification results that we have obtained on our database. These results are summarized and conclusions are drawn in Section 4.

1. The database and Matlab scripts to read the data and extract all features considered in this article can be downloaded on the following site: <http://www.technicolor.com/en/innovation/scientific-community/scientific-data-sharing/eeg4emo-dataset>

2 MATERIAL AND METHODS

2.1 Data acquisition and preprocessing

In this section, we describe the experiment we have conducted to collect EEG and physiological recordings for different emotional states elicited by videos and the preprocessing steps, which have been applied to the data prior to their analysis. This experimental study was approved by the local ethical review board of Inserm.

2.1.1 Subjects and stimuli

We recruited 40 adults (age 36.2 ± 13.9 years, 9 females) with normal or corrected to normal vision and hearing to participate in the experiment. For the choice of the stimuli used in our experiment, we considered four publicly available audio-visual databases: MAHNOB-HCI [10], HUMAINE [28], LIRIS-ACCEDE [29], and FilmStim [30]. Among these databases, we chose the FilmStim base as the most suitable for eliciting positive and negative emotions, considering both the nature of the stimuli (film excerpts) and the sufficient stimulus lengths (40 s to 6 min) to successfully elicit emotions. In total, this database comprises 70 excerpts from well-known movies, which have been used by neuropsychologists for studies of emotion. More particularly, the database includes 10 videos from each of the following categories: tenderness, amusement, sadness, anger, disgust, fear, and neutral. For our experiment, we selected 13 of the 70 videos²: 6 videos associated with negative emotions (anger, fear, disgust, and sadness) and 7 videos associated with positive emotions (amusement and tenderness) among which one video is used in the test trial. As the videos in the FilmStim database are in French, all recruited subjects were French-speaking.

2.1.2 Experimental protocol

To study the emotions elicited by the selected videos, we have designed an experiment of about 45 min length, which has been implemented using the presentation software E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). During the experiment, the subject is comfortably installed on a chair in front of a 21" computer screen, which is used for the presentation of the stimuli and which is placed at about 1m distance from the subject. The experiment starts with some general information on the experimental protocol and instructions for the subject to relax and to stay still as much as possible during the experiment. The experiment is composed of one test trial followed by 12 trials. Each trial starts with a black screen shown for 8 s. A white fixation cross then appears for 2 s in order to alert the subject to the beginning of the next video. First, a short neutral video of about 10 s length is presented to disperse lingering emotions. The neutral video is followed by 8 s of black screen and 2 s of fixation cross announcing the start of an emotional video. In order to evoke a positive or a negative emotion, a short video clip with a duration ranging from 40 s to 6 min is then shown to the subject. After the video, the subject evaluates the valence of the emotions he or she felt during the video on a negative (-1) - neutral (0) - positive

2. More particularly, the videos with the following numbers have been selected: 1, 5, 8, 12, 16, 19, 27, 39, 42, 55, 56, 61, 67.

(1) discrete scale. The subject enters the response using the keyboard of the computer on which the experiment is run. There is no time limit for the self-assessment of the valence. The trial ends when a value has been entered in the system. The order in which the emotional and neutral videos are presented over the 12 trials is arbitrary.

The experimental protocol is summarized in Figure 1.

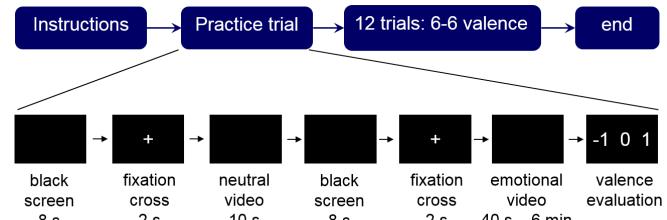


Fig. 1. Experimental protocol.

2.1.3 EEG and physiological recordings

HR-EEG recordings were acquired in a shielded room using the 257-channel EGI system (EGI, Electrical Geodesics Inc., Eugene, USA). The initial sampling rate was 1000 Hz, but to reduce the computational burden of the data analysis, the data were later subsampled to $f_s = 250$ Hz. The impedances of the EEG electrodes were kept below 50 kΩ. The same system was employed to record an ECG signal, the respiration with a piezoelectric effort belt, as well as the peripheral oxygen saturation and pulse rate using a pulse oximeter. Please note though that these peripheral physiological signals are not included in the analysis presented in this paper, but were recorded along with the EEG for potential future use.

We also collected simultaneous GSR recordings using a SenseWear sensor (BodyMedia Inc.) placed on one hand of the subject.



Fig. 2. Setup of the experiment in a shielded room in Pontchaillou hospital, Rennes, France.

2.1.4 Data preprocessing

Among the 40 subjects that participated in the experience, the data of 13 subjects were discarded from the analysis due to bad EEG signal quality (because of bad electrode contact, for example for subjects with long or thick hair) or insufficient elicitation of emotions (only subjects for whom we succeeded in eliciting emotions for more than 50% of

the positive videos and for more than 50% of the negative videos were retained). Our analysis was thus performed for 27 subjects (age 35.0 ± 12.9 years, 5 females). For each of these subjects, we first synchronized the EEG and GSR data with each of the emotional videos. The EEG data of each video were then band-pass filtered between 2 Hz and 80 Hz using a 5-th order Butterworth filter. Bad channels were identified based on visual inspection and were replaced by the interpolated signals from the 4 closest channels. To eliminate eye blinks, we considered applying ICA with Wavelet Denoising (WD) [31]. We also tried to reduce muscle artifacts by rejecting ICA components containing predominantly muscle activity, which were selected by thresholding the sum of several autocorrelation values. However, we abandoned this approach because of its unreliability due to the high subject-dependence of the threshold value.

As we did not expect the subjects to feel strong emotions for the whole duration of each video, we focused our analysis on the 10 s interval of each video which had the highest emotional impact. As emotions are very subjective, these intervals were selected for each subject individually based on a manual annotation of the emotional content of the videos in combination with an analysis of the GSR signals to take into account subject-specific reactions. More precisely, the peaks of the GSR signal are known to indicate high arousal [8] and may therefore also give an indication of strong positive or negative feelings. Nevertheless, the GSR peaks do not systematically coincide with scenes of strong emotional impact. Hence, in order to select the most relevant GSR peak or to select an appropriate interval in the absence of meaningful GSR peaks, the GSR peaks were compared with the manual annotations and the latter were given precedence in the case of disagreement to ensure the reliability of the selected intervals.

2.2 Feature extraction and classification

In order to discriminate between positive and negative emotions, which are elicited by the videos, we need to identify features from the EEG recordings that are characteristic for each of the two conditions. In the literature, the features are generally extracted directly from the signals recorded by the EEG sensors. However, the sensors include a mixture of the signals originating from all over the brain whereas not all brain regions will contribute to the processing of emotions. Therefore, we explore in this article the idea of improving the classification results by reconstructing first the cerebral activity everywhere within the brain (or, more particularly, everywhere on the cortical surface) and by extracting the features in a second step from the reconstructed cortical activity of selected brain regions. Subsequently, we refer to features extracted directly from the EEG channels as features in the *sensor space* whereas the features extracted from the reconstructed cortical activity are referred to as features in the *source space*.

2.2.1 Reconstruction of brain activity

The cerebral activity can be modeled by a large number of current dipoles ($\sim 10,000$) spread over the cortical surface with an orientation perpendicular to this surface. Each of these dipoles can be associated with a specific anatomical

brain region following, for example, the Desikan-Killiany brain atlas [32], which distinguishes 66 brain regions and provides a more macroscopic model of the brain. As some of these regions are quite large, in particular compared to others, we have subdivided these regions into smaller clusters of dipoles, leading to R brain regions of similar size. We are then interested in the temporal dynamics of each of these brain regions during the EEG recordings.

Denoting by $\mathbf{S} \in \mathbb{R}^{R \times T}$ the signal matrix, which characterizes the temporal dynamics of the R brain regions at T time points, the measurements $\mathbf{X} \in \mathbb{R}^{N \times T}$ recorded by N sensors on the scalp can be modeled as

$$\mathbf{X} = \mathbf{GS} \quad (1)$$

where the lead field matrix $\mathbf{G} \in \mathbb{R}^{N \times R}$ describes the diffusion between the signal of each brain region and the signal measured at each sensor. Given a model of the head, the lead field matrix can be computed numerically and can thus be assumed to be known [33]. In this paper, we use the template head model Colin27 provided in Brainstorm [34]. Among the available brain meshes, the white matter surface mesh is used as source space (each vertex corresponding to a source dipole). The 8000 grid dipoles are subdivided into $R = 549$ brain regions using the Brainstorm clustering functions and the lead field matrix is computed using Open-MEEG [35], [36]. To reconstruct the cerebral activity within each brain region, one has to solve an inverse problem, which consists in estimating the signal matrix \mathbf{S} from the data \mathbf{X} . This inverse problem is ill-posed as the number of brain regions is larger than the number of sensors and requires making additional assumptions on the solution. Among the large number of available techniques [37], we employ a commonly used regularized least squares method, the WMNE [38] algorithm, which solves an optimization problem of the form

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{GS}\|_2^2 + \lambda \|\mathbf{WS}\|_2^2 \quad (2)$$

where \mathbf{W} is a diagonal weight matrix and λ a regularization parameter. We have also considered using two other regularized least squares methods, the well-known sLORETA algorithm [39] and SISSY [40], a recently proposed technique for reconstructing brain sources of spatial extent. However, the valence classification results were very similar, but slightly worse than those obtained with WMNE, which is why we do not report them in this article.

While all $R = 549$ brain regions contribute to the measured EEG signals, only part of them will also be involved in processes related with emotions. Indeed, much research has been conducted to identify the brain regions that exhibit a systematic activation during positive or negative affect (see [41], [42] and references therein). For further analysis and feature extraction, we consider only the brain regions which have been observed to be involved in the processing of emotions in functional Magnetic Resonance Imaging (fMRI) studies [41] or intracranial EEG studies [43]. More precisely, features are extracted from 274 of the 549 brain regions (cf. Fig. 3).

2.2.2 Considered features

The features are extracted from several segments of either 1 s, 2 s, or 5 s length of the 10 s interval selected for each

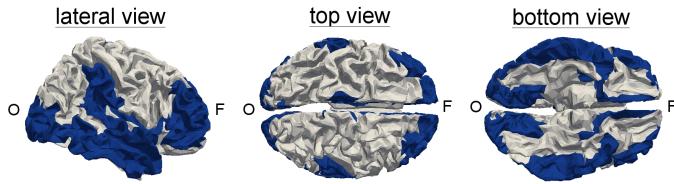


Fig. 3. Selected brain regions (in blue) that are involved in the processing of emotions according to the literature.

video based on the GSR. All features are computed both in the sensor space (for each electrode) and in the source space (for each of the 274 emotional source regions). The features considered in this paper are listed in Table 1 and are described below.

TABLE 1

Features considered in this paper. The variable N corresponds to the number of elements for which the features are extracted (the number of electrodes for features in the sensor space and the number of source regions for features in the source space).

features	number
$\theta, \alpha, \beta, \gamma_e, \gamma_h$ band powers	$5N$
PSI	$N(N-1)/2$
PSI in $\theta, \alpha, \beta, \gamma_e, \gamma_h$ bands	$5N(N-1)/2$
HOC	$50N$
HOC in $\theta, \alpha, \beta, \gamma_e, \gamma_h$ bands	$100N$
statistics	$10N$
spectral moments	$4N$
SCF	$5N$

Band powers The powers in different frequency bands are the most classical and widespread features employed in the context of EEG emotion recognition. Here, we consider the powers in five different frequency bands, the θ (4 Hz - 8 Hz), α (8 Hz - 13 Hz), β (13 Hz - 30 Hz), low γ (30 Hz - 45 Hz), and high γ (55 Hz - 80 Hz) bands.

Connectivity features Functional connectivity characterizes interactions between brain regions, which are quantified based on measures of dependency between their temporal dynamics. By thresholding the connectivity values between all pairs of brain regions, one can identify the functional connectivity networks, showing which brain regions interact with each other. Differences in functional connectivity networks related to positive, negative, or neutral emotions have recently been studied [44] and functional connectivity values have been considered as features to discriminate between different emotions [15]. These analyses have been conducted at the sensor level (between pairs of electrodes) whereas functional connectivity can also be analyzed at the source level [45]. Among the different functional connectivity measures considered in the literature, we employ the Phase Synchronization Index (PSI), which is reported to be more robust than correlation [15]. The PSI is computed based on the instantaneous phases $\phi_m(t)$ and $\phi_n(t)$ of the signals (derived from the Hilbert transform, see [15]) at brain regions m and n :

$$PSI(m, n) = \left| \frac{1}{T} \sum_{t=1}^T \exp \{j(\phi_m(t) - \phi_n(t))\} \right|. \quad (3)$$

The non-redundant elements of the resulting connectivity matrix are stored into a feature vector. We consider connec-

tivity features derived directly from the signals on the one hand and determined from the filtered signals for the θ, α, β , low γ , and high γ bands on the other hand.

Higher Order Crossing (HOC) features The HOC features were originally introduced in [46]. The n -th order HOC value H_n corresponds to the number of zero-crossings of the n -th order derivative of the signal $x(t)$:

$$H_n = NZC\{\nabla^{n-1}x(t)\} \quad (4)$$

where ∇ denotes the backward difference operator such that $\nabla x(t) = x(t) - x(t-1)$ and $NZC\{\cdot\}$ is a function that determines the number of zero-crossings. In this paper, we consider HOC features derived directly from the signals up to order 50 and HOC features for the filtered signals in the θ, α, β , low γ , and high γ frequency bands up to order 20.

Fractal Dimension (FD) features The FD is a measure of signal complexity, characterizing the self-similarity of irregular time series, and has been proposed in [47] for use as a feature for EEG-based emotion recognition. Following [14], we compute it using the Higuchi algorithm [48] based on a set of τ time series derived from the signal $x(t)$:

$$\left\{ x(t), x(t+\tau), \dots, x\left(t + \left\lfloor \frac{T-t}{\tau} \right\rfloor \cdot \tau\right) \right\}$$

with $t = 1, \dots, \tau$. For each of these time series, a length of curve is computed as

$$L_t(\tau) = \frac{T-1}{\left\lfloor \frac{T-t}{\tau} \right\rfloor \tau^2} \sum_{m=1}^{\left\lfloor \frac{T-t}{\tau} \right\rfloor} |x(t+m\tau) - x(t+(m-1)\tau)|. \quad (5)$$

Finally, the FD can be derived from the following relation:

$$\frac{1}{\tau} \sum_{t=1}^{\tau} L_t(\tau) \propto \tau^{-FD}. \quad (6)$$

Statistics We consider a set of classical statistical quantities as features, comprising minimum, maximum, median, standard deviation, mean and maximum of the first two derivatives, skewness, and kurtosis of the signals.

Spectral features In [25], different spectral features were introduced. Among these features, we here consider the spectral moments and the Spectral Crest Factor (SCF). The spectral moments include the spectral centroid S_c , the spectral width S_w , the spectral asymmetry S_a , and the spectral flatness S_f , and are given by

$$S_c = \mu_1 \quad (7)$$

$$S_w = \sqrt{\mu_2 - \mu_1^2} \quad (8)$$

$$S_a = \frac{2\mu_1^3 - 3\mu_1\mu_2 + \mu_3}{S_w^3} \quad (9)$$

$$S_f = \frac{-3\mu_1^4 - 6\mu_1\mu_2 - 4\mu_1\mu_3 + \mu_4}{S_w^4} \quad (10)$$

with $\mu_i = \frac{\sum_{k=0}^{K-1} f_k^i a_k}{\sum_{k=0}^{K-1} a_k}$ where a_k is the magnitude of the k -th element of the Fourier transform associated with the frequency sample $f_k = kf_s/T$. The SCF is computed as

$$SCF(\Delta f) = \frac{\max_{k \in \Delta f} a_k}{\frac{1}{K} \sum_{k \in \Delta f} a_k} \quad (11)$$

where Δf denotes the considered subband. We extract the SCF features for the θ, α, β , low γ , and high γ bands.

2.2.3 Feature selection

To reduce the number of features, which can be very high, especially when features are extracted in the source space, we employ a univariate feature selection method based on the t-test (see also [14]). We have also considered using PCA to reduce the dimension of the features, but the results are not as good as those obtained with feature selection, which is why we do not report them here.

2.2.4 Classification and performance evaluation

In this paper, we employ an SVM classifier with a linear kernel, which relies on the features to be suitably chosen such that they are linearly separable. The advantage of the linear SVM consists in the fact that it is a simple classifier that requires adjusting only one parameter, which controls the trade-off between low misclassification on the training set and good generalizability to other data (by choosing a separating hyperplane with a large margin to both classes). After testing different values for this parameter, which had little impact on the classification results, it was set to a default value. The classification performance achieved for different features is evaluated following a cross-validation scheme. More particularly, we consider a leave-one-video-out (LVO) cross-validation scheme, where the classifier is trained using the data segments of all but one videos, excluding the video whose data segments are used for testing from the training set. The classification performance is assessed in terms of specificity and sensitivity, which indicate the probability of correct classification for each class. As an overall classification score, we employ the average of the specificity and sensitivity values. Contrary to the accuracy, this classification score is independent of the probability of each class. Furthermore, in contrast to the F1-score, it has the advantage that it does not depend on which class is given label 0 and which class is associated with label 1. When averaging the results of different conditions, the illustrated average gap between specificity and sensitivity values corresponds in fact to the average of the absolute value of the difference between specificity and sensitivity for each condition.

3 RESULTS

In this section, we compare the classification performances achieved with different types of features, extracted in the source space or the sensor space. We first analyze the overall performance achieved for different subjects and on average in Section 3.1. Then, in Section 3.2, we evaluate the influence of several parameters on the classification scores. If not stated otherwise, the following default parameters were used: no ICA-WD, 65 sensors, 2s segments. Moreover, to get a complete picture of the performance that can be achieved with the considered features, we conduct several complementary analyses in Section 3.3, where we compare the classification results obtained for different types of training sets, for different frequency bands, and for a combination of all feature types. Finally, to get a better understanding of how a good discrimination of positive and negative emotions is achieved, in Section 3.4, we conduct two tests that shed some light on the contribution of source space features and the spatial distribution of selected features.

3.1 Overall classification results

3.1.1 Individual subject analysis

First of all, we analyze the classification scores that were obtained for each of the considered 27 subjects. These results are shown in Fig. 4. The performance of the emotion recognition obviously varies considerably from one subject to another, with scores below 40% for subject 33 and scores higher than 80% for subject 40 for all types of features. In order to get some insight on the possible causes for the high variability of the classification results for different subjects, we have performed a more detailed analysis of the five datasets (of subjects 2, 3, 17, 26, and 40) for which the best classification results have been achieved and of the five datasets (of subjects 4, 9, 10, 15, and 33) for which the worst classification scores have been observed. This analysis has revealed that in case of the five “best” subjects, the targeted emotions were elicited very successfully. More particularly, for 4 of these subjects (2, 3, 17, and 26), the subjects’ self-assessment indicated that they felt the targeted emotions for all 12 videos. Subject 40 reported having felt positive/negative emotions for only 8 videos, but showed strong reactions to the videos evaluated as positive. For all five subjects, the EEG signal quality was observed to be very good. By contrast, for the five “worst” subjects, the elicitation of the targeted emotions was less successful: for three of these subjects (4, 15, and 33), positive/negative feelings were reported for 9 videos only, and the remaining two subjects felt emotions for 11 videos. Furthermore, for three subjects (9, 15, and 33), the EEG data was quite noisy. These differences in signal quality and success of emotion elicitation might explain the discrepancies in classification performance achieved for individual subjects.

3.1.2 Comparison of results for different feature types

Next, we will compare the classification performances that are achieved with the different types of extracted features, from band powers to SCFs. While different types of features yield different classification results, for most subjects, the values of the classification score tend to be of similar order across features. The type of feature that results in the highest classification score varies from subject to subject. However, from the classification scores averaged over subjects (see, e.g., Fig. 5 or the top of Fig. 6), we note that the best results are generally achieved for connectivity features extracted from the θ , α , β , low γ , and high γ bands, followed by band powers, statistics, and spectral moments. The features that exhibit the lowest performance scores are generally fractal dimension and HOC features. For both connectivity and HOC features, we usually obtain better results when these features are extracted from the five considered frequency bands rather than from the signals covering the whole frequency range.

3.1.3 Interest of source space features

Comparing the results obtained for source space features and for sensor space features, it can be observed that classifying valence with features extracted in the source space generally leads to better results than with features extracted in the sensor space. Even though this improvement of the classification scores is not always systematic for all feature

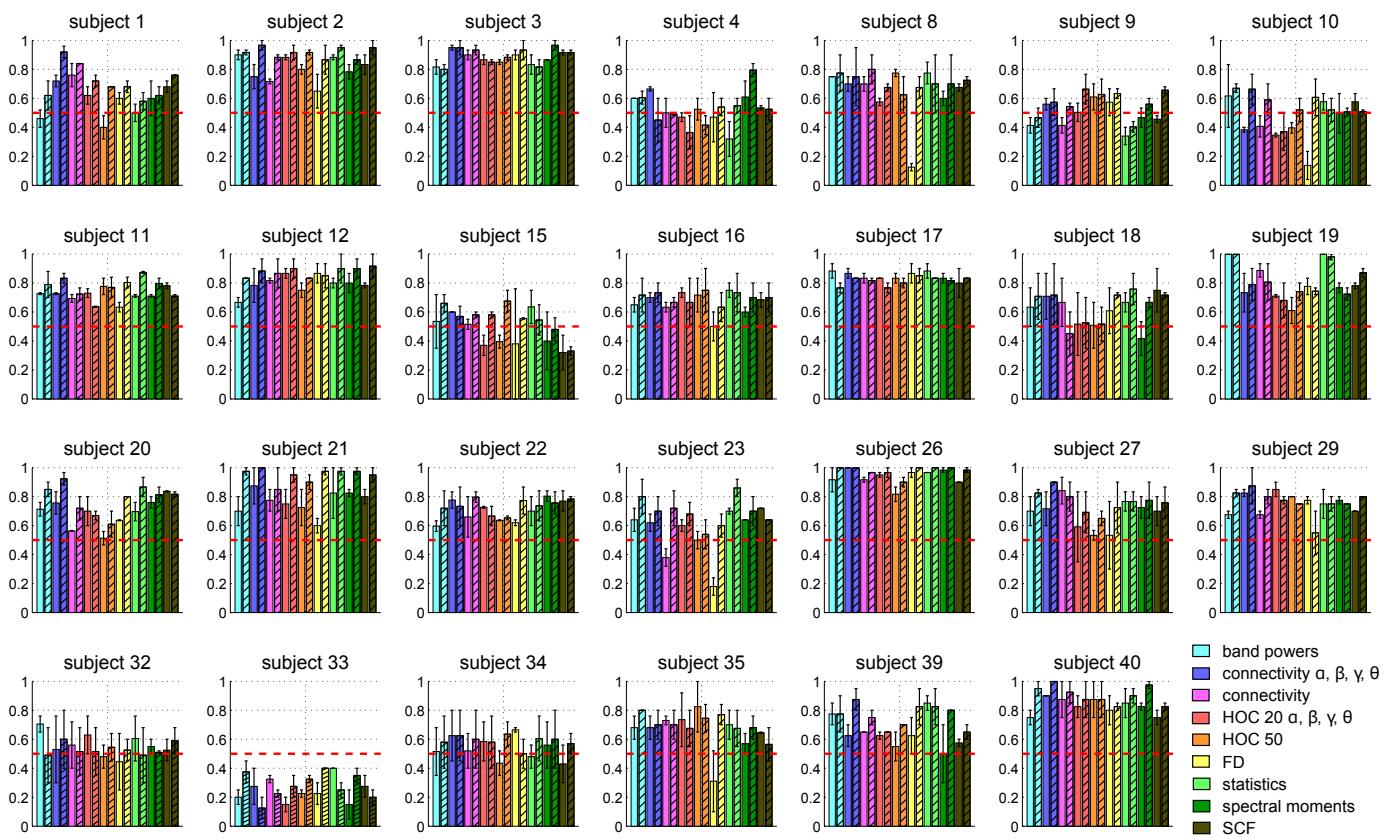


Fig. 4. Performance using different features for each of the 27 analyzed subjects. The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the average range between specificity and sensitivity values. The plain colored bars mark the performance of sensor space features whereas the hatched colored bars indicate the performance of source space features. The red dotted line corresponds to the chance level of classification at 50%.

types, it can be observed for the majority of features for each subject. On average, across subjects, the performance gain achieved by source space features is similar for all types of features and leads to classification scores increased by roughly 5%. This shows that reconstructing the brain activity on the cortical surface prior to feature extraction is clearly of interest for a successful classification of emotional valence.

3.2 Influence of different processing steps and parameters

3.2.1 Influence of artifact removal

In order to assess the importance of removing eye blink artifacts prior to feature extraction and classification, we compare the valence recognition performance for EEG data where ICA-WD was applied in the preprocessing with the results obtained without ICA-WD preprocessing. The classification scores averaged over all subjects for these two conditions are shown in Fig. 5. It can be seen that ICA-WD preprocessing seems to have little impact on the classification results. Therefore, it has not been applied in the remaining analyses reported in this paper.

3.2.2 Influence of the number of sensors

One of the main interests of HR-EEG recordings consists in the possibility to analyze the results achieved with different subsets of electrodes, which may help to draw conclusions

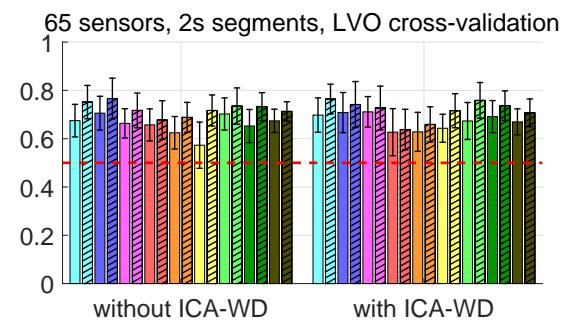


Fig. 5. Classification performance for features extracted from the EEG data with and without ICA-WD preprocessing. The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the average range between specificity and sensitivity. The red dotted line indicates the chance level of classification at 50%.

on an optimal electrode setup. In this section, we consider several standard electrode configurations ranging from 14 sensors (as for the Emotiv EPOC) to 257 channels (cf. Fig. 6, bottom) in order to evaluate the influence of the number of sensors on the classification performance. Fig. 7 shows the classification scores averaged over all types of features and all subjects and the standard deviation of these scores over subjects for each electrode configuration. Furthermore, Fig. 6 shows the results for individual feature types and the variability of the classification scores (median, 25% and 75% quartiles, minimum, and maximum) over subjects. It

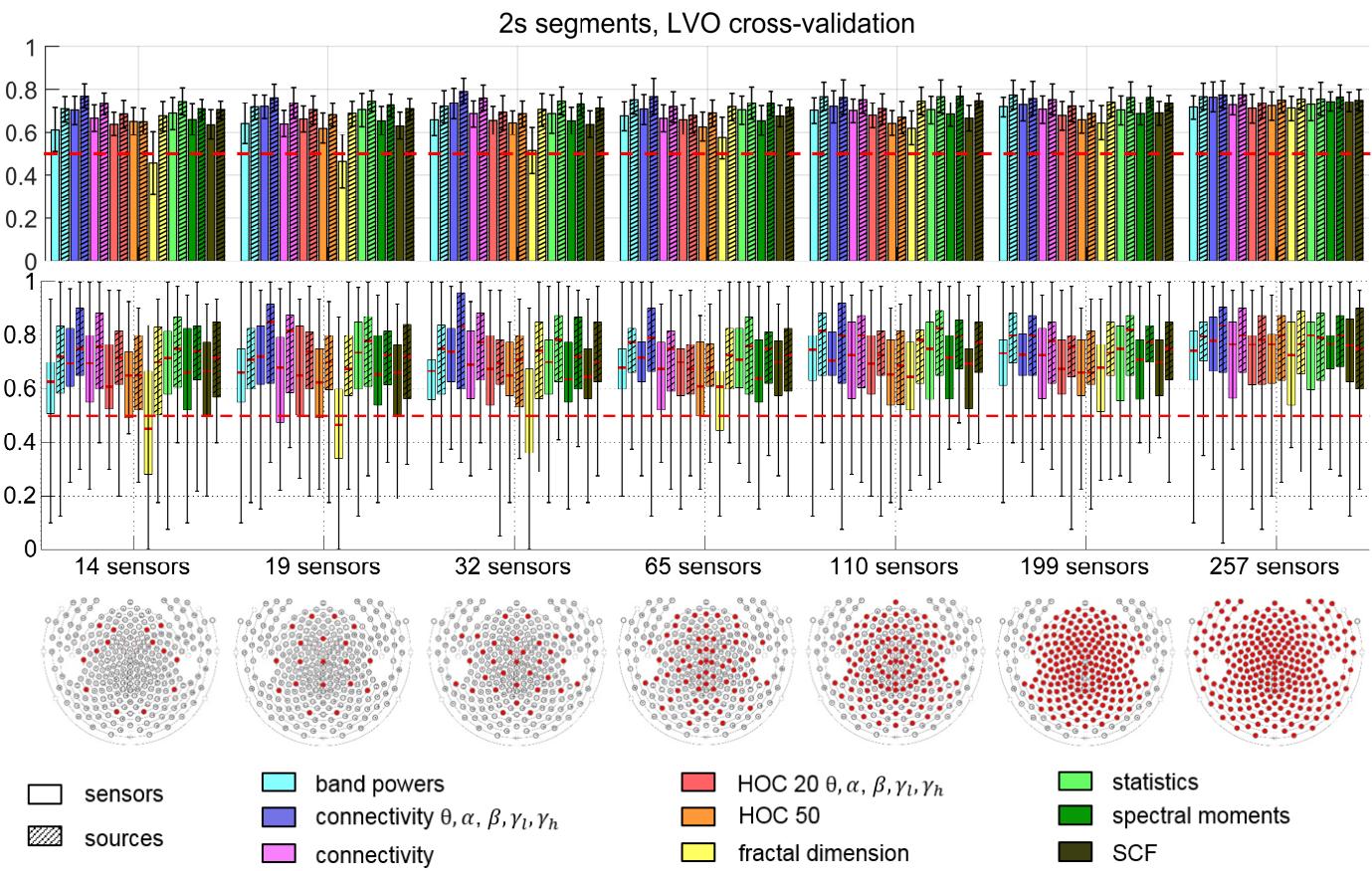


Fig. 6. Classification performance for various electrode configurations. (Top) The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the average range between specificity and sensitivity. The red dotted lines correspond to the chance level of classification at 50%. (Middle) Variability of the classification scores over subjects: The red lines mark the median classification score, the whiskers mark the minimum and maximum and the colored bar extends from the 25% to the 75% quartile. (Bottom) Different electrode configurations: the considered sensors are marked in red.

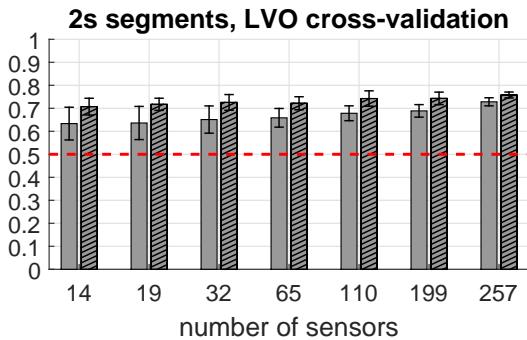


Fig. 7. Average classification performance for sensor space features (plain bars) and source space features (hatched bars) for various electrode configurations. The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the standard deviation. The red dotted line indicates the chance level of classification at 50%.

can be seen that the classification performance generally improves with increasing number of sensors. On average, the performance gain from 14 sensors to 257 sensors amounts to almost 10% for sensor space features and to approximately 5% for source space features. It is also noteworthy that for small numbers of sensors, the classification accuracy is clearly enhanced (by 7% on average) for features extracted

in the source space compared to features extracted in the sensor space. However, for large numbers of sensors, the difference between the classification results achieved with source space features and with sensor space features is rather small (about 3% on average across features). Comparing the results achieved with different types of features, we notice that the best results, in particular for small numbers of electrodes, are achieved with the connectivity features extracted for the θ , α , β , low γ , and high γ frequency bands. The poorest classification performance is obtained for the FD and HOC features. However, while the average classification scores obtained for different features vary considerably for small numbers of sensors (by up to 40%), all features lead to similar results for 257 electrodes (maximal difference of about 6%). On the other hand, the variability of the results from subject to subject is the same for all electrode configurations.

3.2.3 Influence of the segment length

When designing a valence recognition application, one has to decide on the length of the data segments to be analyzed. In order to determine which segment length is required to obtain good classification results, we evaluate the influence of this parameter on the classification performance. To this end, we show in Fig. 8 the classification score averaged over all subjects for the three considered segment lengths. It can

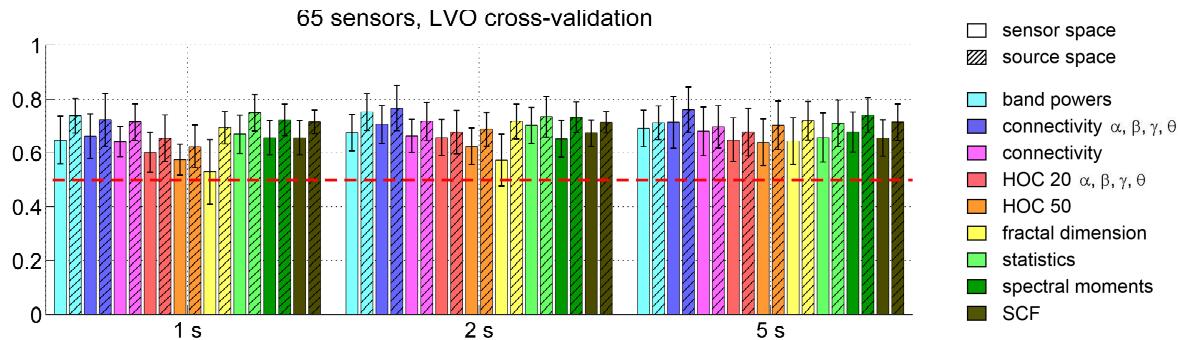


Fig. 8. Classification performance for different segment lengths. The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the average range between specificity and sensitivity. The red dotted line correspond to the chance level of classification.

be seen that the classification results improve only slightly when increasing the segment length from 1 s to 2 s. For 5 s segments, the results are approximately the same as for 2 s segments. Therefore, the segment length does not seem to have a big impact on the classification scores.

3.3 Complementary analyses

3.3.1 Sensitivity to training data

A crucial element in machine learning applications is the dataset that is employed for training the classifier because the suitability of the training set has an important impact on the classification performance. In order to illuminate this issue, we consider two other training sets in addition to the one used by the LVO cross-validation approach. First, we evaluate the classification scores that can be achieved by training the classifier on all data segments except for the one used for testing (including segments from the same video as the test segment), referred to as leave-one-segment-out (LSO) cross-validation. Second, we analyze the performance that can be achieved in a leave-one-subject-and-one-video-out (LSVO) cross-validation, where the training set consists of the features extracted from all other subjects and from all other videos than the one used for testing.

Fig. 9 (left) shows the classification results obtained using LVO cross-validation and using LSO cross-validation. It is obvious that for all features except the band powers in the source space, LSO cross-validation clearly results in much higher classification accuracies. The difference is about 20% on average for sensor space features and slightly less for source space features. Exploiting information originating from other segments of the same video during the training of the classifier thus considerably improves the classification performance. However, in practice it is not possible to reach the performance achieved by LSO cross-validation because the training of the classifier cannot be done “on the fly”. Instead, in an actual application, before being able to make use of emotion recognition, the subject would need to complete a training session where data is acquired solely for the purpose of training the classifier.

This training session could be avoided if the system could be pre-trained beforehand based on data recorded from other subjects. This case corresponds to the LSVO cross-validation scheme in our analysis. Yet it can be seen in Fig. 9 (right) that this approach, tested in a best case scenario (best subject - number 26 - and best features - connectivity in

different frequency bands), leads to very bad classification scores. This suggests that the discrepancies between the EEG signals of different subjects are too important to enable pre-training the classifier on data from various subjects. Instead, individual training of the classifier for each subject seems to be essential for successful valence recognition.

3.3.2 Classification accuracy in different frequency bands

In the literature, several studies have compared the classification results obtained for different frequency bands to identify the frequency bands that contain the most discriminant information for classifying emotions. While most studies report best classification results for features extracted from high frequency bands, e.g., β -band, β - and γ -bands, or γ -band [14], [49], others report the most discriminative information to be present in low frequency bands such as the θ -band [50]. We contribute to this debate by analyzing the classification performances that can be achieved by exploiting only features from individual frequency bands for the band power, PSI, HOC, and SCF features considered in this paper, which have been extracted for the θ , α , β , low γ , and high γ frequency bands.

As can be seen in Fig. 10, the best classification scores are achieved for band power and connectivity features in the low and high γ -bands, amounting to approximately 70% for sensor space features and 75% for source space features, closely followed by the results of the β -band. By contrast, the HOC features perform best in the α - and β -bands. In the α -band, the HOC features outperform all other features and in the β -band, they yield a slightly higher score than band power and connectivity features in the sensor space, but do not reach as good results as the band power and connectivity features in the source space. Finally, for the SCF feature, the classification results obtained in different frequency bands are very similar, but this feature only outperforms the other feature types in the θ -band where band power, connectivity and HOC features do not lead to good results. Overall, our results are in line with the findings of other studies reporting best classification scores for β - and γ -bands.

3.3.3 Combination of all feature types

Similar to the analysis conducted in [14], we also consider combining features from all types and compare the classification performance achieved by this approach to the average performance for separate features. The results, averaged

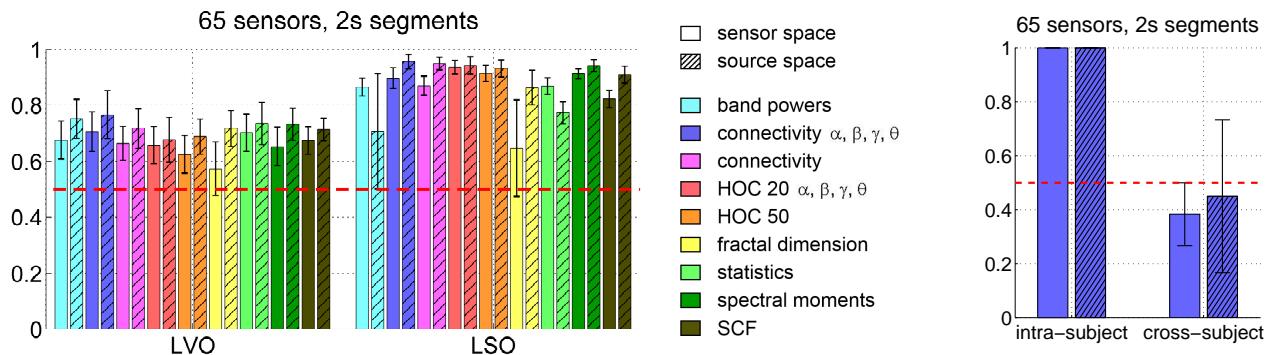


Fig. 9. (Left) Classification performance for LVO cross-validation vs. LSO cross-validation. The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the average range between specificity and sensitivity. The red dotted line indicates the chance level of classification at 50%. (Right) Best case (subject 26, connectivity features) classification results of LVO cross-validation (intra-subject) vs. LSVO cross-validation (cross-subject)

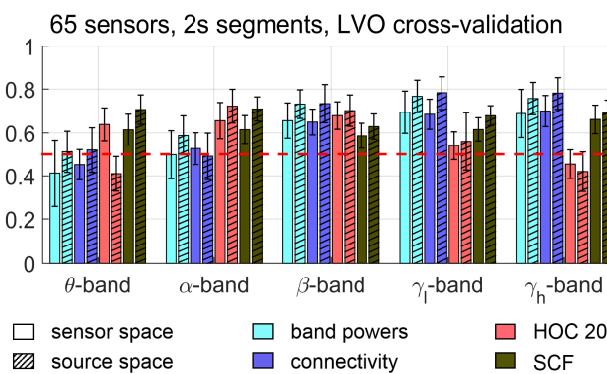


Fig. 10. Classification performance for different frequency bands. The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the average range between specificity and sensitivity. The red dotted lines correspond to the chance level of classification at 50%.

over all subjects, are displayed in Fig. 11 (left), showing that by combining all features, the classification score is about 3% to 4% better than the average score achieved for individual feature types. A closer look at the variability of these results over subjects reveals that the median performance values show a higher improvement of about 5% for the combination of all features. Yet the variability over subjects is also increased, thereby leading to a smaller improvement on average.

To determine which types of features are actually used in the combined feature vector after feature selection, we show in Fig. 11 (right) the percentage of features from each type that are chosen in the feature selection step and the final composition of the feature vector. For example, Fig. 11 (center right) shows that 30% of the band power features originally computed have been retained in the feature selection step. However, these features only make up about 2% of the features in the final feature vector that combines features from all types (see Fig. 11 right). The highest percentages of features are retained from the band powers, statistics, and spectral moments (between 30% and 40%), especially in the source space. Nevertheless, the final feature vector is mainly composed of connectivity features extracted in different frequency bands, followed

by HOC features. Even though the percentage of selected connectivity features is rather small, these features are very numerous (about $5N^2/2$) compared to other feature types (e.g. $5N$ band power features), hence the predominance of connectivity features in the final feature vector. Together with the observation that the connectivity features in different frequency bands lead to the best classification scores when analyzing separate feature types, the selection of a large number of connectivity features from the combined feature vector confirms that these features are well suited to discriminate between positive and negative emotions. Finally, the relatively high percentage of selected band power values, statistics, and spectral moments coincides with the fact that these features also perform well when comparing the classification scores achieved by each feature type.

3.4 Towards a better understanding of the contribution of different features

3.4.1 Arbitrary linear mixture of features

By extracting features in the source space instead of extracting them in the sensor space, the number of features is increased because there are more source regions than sensors. This increase is particularly high for small electrode configurations, for which we also observe the highest gain in performance by reconstructing the brain sources prior to feature extraction. A large number of available features improves the possibility to select the most discriminating ones in the feature selection step. One could therefore wonder whether the improvement of the performance observed for source space features could be explained by the increase in dimension of the features rather than by the exploitation of physiological information.

In order to shed light on this issue, we consider extracting features from an arbitrary linear mixture, described by the matrix \mathbf{Y} , of the channel signals of a 32-electrode configuration. We generate as many new signals as there are emotional brain regions by multiplying the 32-channel data, \mathbf{X}_{32} , by an arbitrary mixing matrix $\mathbf{M} \in \mathbb{R}^{274 \times 32}$ whose elements are drawn from a standard normal distribution: $\mathbf{Y} = \mathbf{MX}_{32}$. The classification results averaged over 10 different randomly drawn mixing matrices and over all subjects are displayed in Fig. 12 in comparison to the results obtained for sensor space and source space features.

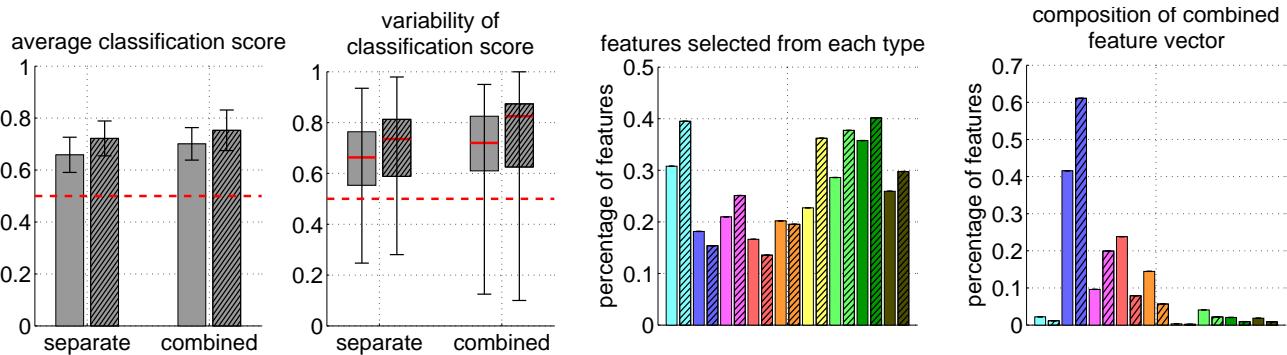


Fig. 11. Classification performance achieved by combining all features in comparison to the average classification scores obtained for separate features. (Left) The height of the bars marks the classification score averaged over all subjects whereas the whiskers mark the average range between specificity and sensitivity. The red dotted line corresponds to the chance level of classification at 50%. (Left center) Variability of the classification scores over subjects: The red lines mark the median classification score, the whiskers mark the minimum and maximum and the colored bar extends from the 25% to the 75% quartile. (Right center) Percentage of features selected from each of the nine considered feature types (averaged over all subjects). (Right) Composition of the final feature vector containing a combination of features from all types (averaged over all subjects).

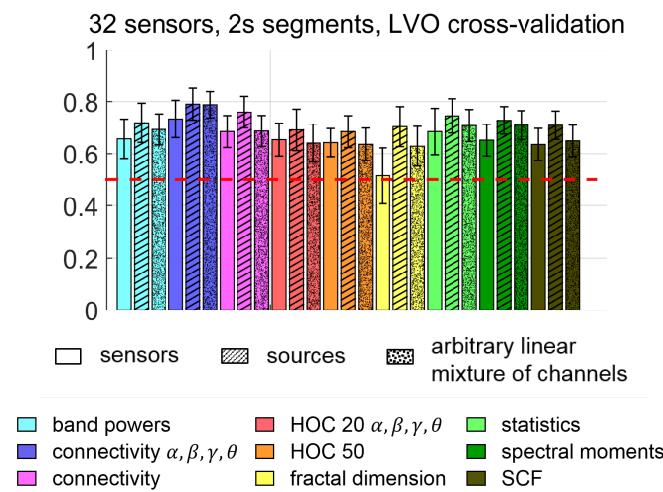


Fig. 12. Classification performance for features extracted from an arbitrary linear mixture of the EEG data in comparison to sensor space features and source space features. The height of the bars marks the classification score averaged over all subjects and over 10 arbitrary mixing matrices (in the case of linear mixing features) whereas the whiskers mark the average range between specificity and sensitivity. The red dotted line corresponds to the chance level of classification.

For the connectivity features extracted in different frequency bands, which lead to the highest classification scores, the classification results obtained from the linear data mixture are as good as those based on the reconstructed sources. However, for connectivity features extracted from the whole frequency range and for HOC features, the linear mixing of the data does not improve the classification results of the sensor space features. For all other feature types, the classification scores achieved using the features extracted from an arbitrary linear mixture of the data lie between those of the sensor space and source space features. Therefore, we conclude that overall, the linear mixing features lead to better results than the sensor space features, but do not achieve as good performances as the source space features. This result suggests that the performance gain observed for the source space features compared to the sensor

space features could be partly explained by the increase in dimension of the features.

3.4.2 Spatial distribution of selected features

To better understand how the discrimination between positive and negative emotions is achieved, we examine which features are selected from each set of computed features. More particularly, we determine the frequency with which features from each channel (for sensor space features), channel pair (for connectivity in the sensor space), brain region (for source space features), or brain region pair (for connectivity in the source space) are selected for all subjects. The results are illustrated in Fig. 13 for band power features, connectivity features in θ , α , β , low γ , and high γ frequency bands, and SCF features. Similar results are obtained for other types of features.

The most frequently selected band power features are associated with the electrodes located on the left and right cheeks of the subject, followed by the temporal electrodes. For the SCF features, we make the same observations. In this case, the predominant selection of cheek and temporal electrodes is even more pronounced than for the band power features. In the source space, the most frequently selected brain regions for band power and SCF features are located on the left and right temporal lobes. But frontal and occipital are also selected frequently, in particular for band power features. For the connectivity features, we note that the predominant connections involve pairs of electrodes on the left and right cheeks, pairs of electrodes on the front, and a pair of temporal electrodes on each side of the head. The most frequently selected connections in the source space consist of various pairs of brain regions located in the frontal lobe and on the temporal pole.

As the electrodes located on the cheeks record mostly EMG signals generated by facial expressions such as smiles and the temporal electrodes are sensitive to muscle activity originating from movements of the jaw (e.g., due to smiles or clenched teeth), the fact that features from these electrodes are most frequently selected implies that the muscle activity present in the EEG undoubtedly plays an important role in the discrimination of positive and negative

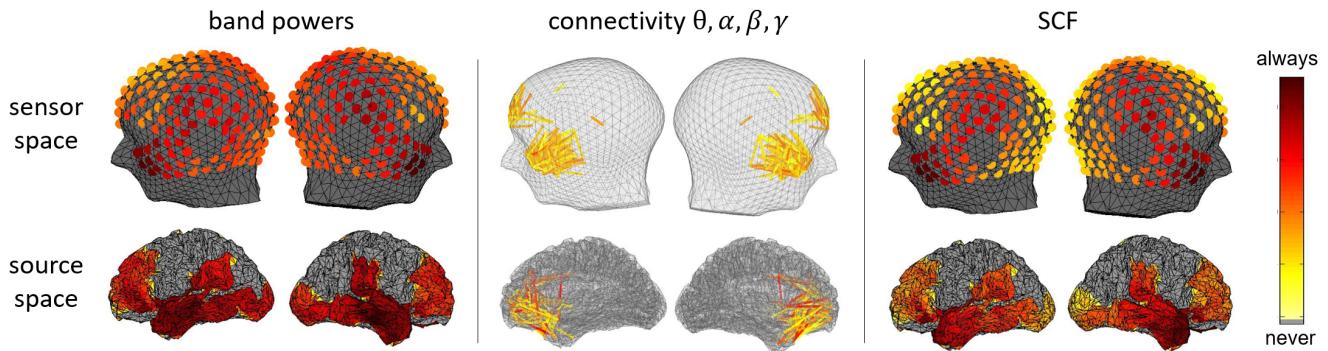


Fig. 13. Frequency of feature selection for each channel/brain region, averaged over all subjects.

emotions. Nevertheless, this does not necessarily mean that the brain activity does not contribute to the classification of valence. Indeed, in the source space we observe a stronger contribution of features extracted from all brain regions, including not only temporal areas, which are the closest to the origin of the muscle activity, but also frontal and occipital areas. This broader selection of features in the source space might also explain the better performance with respect to sensor space features. Unfortunately, we cannot assess the classification performance achieved solely by the neuronal components of the EEG because up-to-date it is not possible to reliably remove the muscle artifacts despite much effort that is dedicated to this subject.

4 CONCLUSIONS

We have presented a new, publicly available database for studies of emotion elicited by audio-visual stimuli, which comprises the HR-EEG and physiological recordings of 40 subjects. We have then compared the classification performance that can be achieved with seven different types of features extracted from the EEG recordings either in the sensor space or in the source space. It can be observed that in general, reconstructing brain sources prior to feature extraction improves the performance of the emotion recognition. Indeed, on average the source localization step leads to an increase of the classification score by 5 %. This confirms our hypothesis that by exploiting physiological information in the source reconstruction and the selection of brain regions, we can extract more discriminative features, thus enhancing classification performance. However, this performance gain is at least partly due to the increase of the number of features in the source space compared to the number of features in the sensor space as has been revealed by analyzing arbitrary linear data mixtures to increase the data dimension.

The fact that features extracted from arbitrary linear data mixtures lead to an improved performance compared to features extracted directly from each EEG channel also suggests that ANN might be very efficient for classifying emotions because they inherently construct large numbers of features by combining and manipulating all input signals in a way that is best suited to a given task. However, ANN usually require huge databases for training, which are currently not available for EEG-based emotion recognition.

Among the tested features, the functional connectivity features extracted from the θ , α , β , low γ , and high γ

frequency bands yield the best results, leading to overall classification scores of up to 70% or 75%. These classification results are in line with performances for EEG-based emotion recognition reported in the literature. For example, they are better than the average classification accuracies of about 58% obtained for valence classification on the DEAP database [9], but not as good as the results reported in [49] where classification accuracies of more than 90% are claimed. However, the classification results show a high variability from subject to subject and seem to be influenced by factors such as the quality of the EEG signals and the success of emotion elicitation for each subject. These factors apparently have a greater impact on the classification score than the choice of parameters such as the considered number of sensors or the length of the analyzed segments. Moreover, the high subject variability probably accounts for the fact that in order to obtain reasonable results, the classifier needs to be trained for each subject individually.

Finally, our analysis has revealed that the presence of muscle activity in the EEG seems to constitute an important discriminative element for distinguishing positive and negative emotions. Similar observations have been made in [18], where emotions have been classified based on EEG recordings and videos of the subjects' facial expressions. More precisely, using an analysis of Granger causality, the authors of [18] have shown that the discriminative EEG features were caused by muscle activity from facial expressions. These results do not rule out the fact that brain activity may also contain discriminative information about positive or negative emotions, but since the muscle activity cannot be completely eliminated from the EEG, it remains an open question how much the actual brain activity contributes to the classification.

Perspectives for future work include the combination of features extracted from EEG with features extracted from other physiological signals of our database. Furthermore, it would be interesting to merge the classification of valence described in this paper with arousal classification based, e.g., on EEG or the GSR, and to study the influence of arousal on valence classification as videos with negative emotional content are often associated with higher arousal than videos eliciting positive emotions. Finally, the performance of different types of classifiers should be analyzed and compared. In particular, the observation that increasing the number of features to choose from improves the classi-

fication performance suggests that the use of ANNs would be a promising approach for EEG-based valence recognition. While deep learning methods are currently not designed for application to EEG recordings, it would be of great interest to specifically adapt the ANN structures to tackle problems related to observations of brain activity in general and to the recognition of emotions in particular.

REFERENCES

- [1] R. W. Picard and R. Picard, *Affective computing*. MIT press Cambridge, 1997, vol. 252.
- [2] F. Nijboer, S. P. Carmien, E. Leon, F. O. Morin, R. A. Koene, and U. Hoffmann, "Affective brain-computer interfaces: psychophysiological markers of emotion in healthy persons and in persons with amyotrophic lateral sceloris," in *Int. Conf. on Affective Computing and Intelligent Interaction and Workshops (ACII)*, Amsterdam, Netherlands, Sep 10–12 2009.
- [3] P. Ekman, *Handbook of Cognition and Emotion*, 1999, ch. Basic Emotions, pp. 45–60.
- [4] P. J. Lang, "The emotion probe: studies of motivation and attention," *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.
- [5] I. B. Mauss and M. D. Robinson, "Measures of emotion: a review," *Cognition and Emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [6] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.
- [7] M. Soleymani, G. Chanel, J. J. M. Kierkels, and T. Pun, "Affective characterization of movie scenes based on content analysis and physiological changes," *International Journal of Semantic Computing*, vol. 3, no. 2, pp. 235–254, June 2009.
- [8] J. Fleureau, P. Guillotel, and Q. Huynh-Thu, "Physiological-based affect event detector for entertainment video applications," *IEEE Trans. on Affective Computing*, vol. 3, no. 3, pp. 379–385, July–September 2012.
- [9] S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: a database for emotion analysis using physiological signals," *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 18–31, January–March 2012.
- [10] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 42–55, January–March 2012.
- [11] T. Mushi, Y. Terasaki, H. Haque, and G. Ivamitsky, "Feature extraction from EEGs associated with emotions," *Artificial Life and Robotics*, vol. 1, no. 1, pp. 15–19, Mar. 1997.
- [12] A. Choppin, "EEG-based human interface for disabled individuals: emotion expression with neural networks," Master's thesis, Department of Information Processing, Tokyo Institute of Technology, Tokyo, Japan, 2000.
- [13] 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," *Computational and Mathematical Methods in Medicine*, 2013.
- [14] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 327–339, July–September 2014.
- [15] Y.-Y. Lee and S. Hsieh, "Classifying different emotional states by means of EEG-based functional connectivity patterns," *PLOS ONE*, vol. 9, no. 4, April 2014.
- [16] Y.-H. Liu, C.-T. Wu, W.-T. Cheng, Y.-T. Hsiao, P.-M. Chen, and J.-T. Teng, "Emotion recognition from single-trial EEG based on kernel Fisher's emotion pattern and imbalanced quasiconformal kernel support vector machine," *Sensors*, vol. 14, pp. 13361–13388, 2014.
- [17] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, 2014, article ID: 627892.
- [18] M. Soleymani, S. Asghari-Esfeden, Y. Fun, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. on Affective Computing*, vol. 7, no. 1, pp. 17–28, January–March 2016.
- [19] M.-S. Park, H.-S. Oh, H. Jeong, and J.-H. Sohn, "EEG-based emotion recognition during emotionally evocative films," in *Int. Winter Workshop on Brain-Computer Interface (BCI)*, 2013.
- [20] C. A. Kothe, S. Makeig, and J. A. Onton, "Emotion recognition from EEG during self-paced emotional imagery," in *Proc. of Human Association Conference on Affective Computing and Intelligent Interaction*, 2013.
- [21] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): affective ratings of pictures and instruction manual," NIMH Center for the Study of Emotion and Attention, Tech. Rep., 1997.
- [22] L. Albera, A. Kachenoura, P. Comon, A. Karfoul, F. Wendling, L. Senhadji, and I. Merlet, "ICA-based EEG denoising: a comparative analysis of fifteen methods," *Special Issue of the Bulletin of the Polish Academy of Sciences - Technical sciences*, vol. 60, no. 3, pp. 407–418, 2012.
- [23] A. Savran, K. Ciftci, G. Chanel, J. Cruz Mota, L. H. Viet, B. Sankur, L. Akarun, A. Caplier, and M. Rombaut, "Emotiondetection in the loop from brain signals and facial images," in *eINTERFACE'06*, Dubrovnik, Croatia, July 17 - August 11 2006, final Project Report.
- [24] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, September 2015.
- [25] A.-C. Conneau and S. Essid, "Assessment of new spectral features for EEG-based emotion recognition," in *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.
- [26] F. Lotte, A. Lecuyer, and B. Arnaldi, "FuRIA: an inverse solution based feature extraction algorithm using fuzzy set theory for brain-computer interfaces," *IEEE Trans. on Signal Processing*, vol. 57, no. 8, pp. 3253–3263, 2009.
- [27] B. J. Edelman, B. Baxter, and B. He, "EEG source imaging enhances the decoding of complex right-hand motor imagery tasks," *IEEE Trans. on Biomedical Engineering*, vol. 63, no. 1, pp. 4–14, 2016.
- [28] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. McRorie, J.-C. Martin, L. Devillers, S. Abrilian, A. Bathliner, N. Amir, and K. Karpouzis, "The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data," *Affective Computing and Intelligent Interaction*, vol. 4738, pp. 488–500, 2007.
- [29] Y. Baveye, E. Dellandréa, C. Chamaret, and L. Chen, "LIRIS-ACCEDE: a video database for affective content analysis," *IEEE Trans. on Affective Computing*, vol. 6, no. 1, pp. 43–55, January–March 2013.
- [30] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers," *Cognition & Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.
- [31] N. P. Castellanos and V. A. Makarov, "Recovering EEG brain signals: artifact suppression with wavelet enhanced independent component analysis," *Journal of Neuroscience Methods*, vol. 158, pp. 300–312, 2006.
- [32] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *NeuroImage*, vol. 31, pp. 968–980, 2006.
- [33] A. Gramfort, "Mapping, timing and tracking cortical activations with MEG and EEG: Methods and application to human vision," Ph.D. dissertation, Telecom ParisTech, 2009.
- [34] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A user-friendly application for MEG/EEG analysis," *Computational Intelligence and Neuroscience*, vol. 2011, no. 8, 2011.
- [35] A. Gramfort, T. Papadopoulou, E. Olivi, and M. Clerc, "OpenMEG: opensource software for quasi static bioelectromagnetics," *BioMedical Engineering OnLine*, vol. 45, no. 9, 2010.
- [36] J. Kybic, M. Clerc, T. Abboud, O. Faugeras, R. Keriven, and T. Papadopoulou, "A common formalism for the integral formulations of the forward EEG problem," *IEEE Trans. on Medical Imaging*, vol. 24, no. 1, pp. 12 – 28, 2005.
- [37] H. Becker, L. Albera, P. Comon, R. Gribonval, F. Wendling, and I. Merlet, "Brain source imaging: from sparse to tensor models," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 100–112, November 2015.
- [38] R. D. Pascual-Marqui, "Review of methods for solving the EEG inverse problem," *International Journal of Bioelectromagnetism*, vol. 1, no. 1, pp. 75 – 86, 1999.
- [39] ———, "Standardized low resolution brain electromagnetic tomography (sLORETA): technical details," *Methods and Findings in Experimental and Clinical Pharmacology*, 2002.

- [40] H. Becker, L. Albera, P. Comon, R. Gribonval, and I. Merlet, "Fast, variation-based methods for the analysis of extended brain sources," in *Proc. of EUSIPCO*, Lisbon, Portugal, 2014.
- [41] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. Feldman Barrett, "The brain basis of emotion: a meta-analytic review," *Behavioral and Brain Sciences*, vol. 35, pp. 121–202, 2012.
- [42] K. A. Lindquist, A. B. Satpute, T. D. Wager, J. Weber, and L. Feldman Barrett, "The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature," *Cerebral Cortex*, 2015, doi: 10.1093/cercor/bhv001.
- [43] S. A. Guillary and K. A. Bujarski, "Exploring emotions using invasive methods: review of 60 years of human intracranial electrophysiology," *Social Cognitive and Affective Neuroscience*, 2014, doi: 10.1093/scan/nsu002.
- [44] N. Martini, D. Menicucci, L. Sebastiani, R. Bedini, A. Pingitore, N. Vanello, M. Milanesi, L. Landini, and A. Gemignani, "The dynamics of EEG gamma responses to unpleasant visual stimuli: from local activity to functional connectivity," *NeuroImage*, vol. 60, pp. 922–932, 2012.
- [45] M. Hassan, O. Dufor, I. Merlet, C. Berrou, and F. Wendling, "EEG source connectivity analysis: from dense array recordings to brain networks," *PLOS ONE*, vol. 9, no. 8, August 2014.
- [46] P. C. Petrantonis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, March 2010.
- [47] O. Sourina and Y. Liu, "A fractal-based algorithm of emotion recognition from EEG using arousal-valence model," in *Proc. of Biosignals*, 2011, pp. 209–214.
- [48] 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.
- [49] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *31st Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*, Minneapolis, Minnesota, USA, September 2-6 2009.
- [50] M. Y. V. Bekkedal, J. Rossi, and J. Panksepp, "Human brain EEG indices of emotions: delineating responses to affective valizations by measuring frontal theta event-related synchronization," *Neuroscience and Biobehavioral Reviews*, vol. 35, pp. 1959–1970, 2011.



Hanna Becker received her B.S. and M.S. degrees in electrical engineering and information technology from the Ilmenau University of Technology, Germany, in 2010 and 2011, respectively, and her Ph.D. degree from the University of Nice-Sophia Antipolis in 2014. In 2010, she received the European Signal Processing Conference (EUSIPCO) Best Student Paper Award. For her Ph.D. thesis, she received in 2015 the research award of the Société Francaise de Génie Biologique et Médical and the French IEEE Engineering in Medicine and Biology Society section. She has worked as a postdoctoral researcher at Technicolor R&D France for two years in the field of biomedical signal processing for human-machine interfaces. In 2017, she has joined Bosch Sensortec, Germany, as a system engineer to develop signal processing algorithms for sensor fusion applications.



Julien Fleureau is Senior Scientist at Technicolor (Rennes, France) in the Media Computing laboratory within the Research & Innovation Division. He received engineering and M.S. degrees in electrical engineering and computer science from the Ecole Centrale de Nantes (France) in 2005, and a Ph.D. degree from the University of Rennes (France) in 2008. His Ph.D. studies in image processing and biomedical modeling were followed by a 2 year post doctoral fellowship at the Laboratoire de Traitement du Signal et de l'Image (INSERM, France). His main research interests are related to signal processing, computer vision and machine learning applied to Virtual / Augmented Reality, Human-Machine Interactions, Biomedical engineering and new technologies from a more general point of view.



Philippe Guillotel received the engineering degree in electrical engineering, telecommunications and computer science from the Ecole Nationale Supérieure des Télécommunications (France) in 1988, the M.S. degree in electrical engineering and information theory from University of Rennes (France) in 1986, and the Ph.D. degree from University of Rennes (France) in 2012. He is Distinguished Scientist at Technicolor Research & Innovation, in charge of video processing & user experiences research topics.

He has worked in the Media & Broadcast Industry for more than 25 years covering video processing, video delivery, coding systems, as well as human perception and video quality assessment. He has been involved in many collaborative projects on those technical areas. More recently he started exploration of new topics to improve user experiences in media consumption, in particular to increase immersion and engagement.



Fabrice Wendling is Director of Research at INSERM. He is heading the team SESAME: "Epileptogenic Systems: Signals and Models" at LTSI, Rennes, France. He got the Biomedical Engineering Diploma (1989) from the University of Technology of Compiègne (France), the Master of Science (1991) from Georgia Tech (Atlanta, USA) and the PhD (1996) from the University of Rennes (France). He has been working on brain signal processing and modeling for more than 20 years, in close collaboration with clinicians. In 2012, he received the award (Prix Michel Montpetit) from the French Academy of Science. He co-authored about 110 peer-reviewed articles.



Isabelle Merlet is a full time research scientist at INSERM. She received the PhD degree in Neurosciences from Lyon 1 University in 1997. She has been working at the Epilepsy department of the Montreal Neurological Institute (1997-2000), at the Neurological Hospital of Lyon (2000-2005) and in the SESAME team of the Signal and Image Processing Laboratory of Rennes (2005 -). Since 1993, her work is devoted to the validation source localization methods and to their application to EEG or MEG signals. She is involved in the transfer of these methods to the clinical ground particularly in the field of epilepsy research.



Laurent Albera was born in Massy, France, in 1976. He received in 2003 the Ph.D. degree in Sciences from the University of Nice Sophia-Antipolis, France. He is now Assistant Professor at the University of Rennes 1 and is affiliated with the INSERM research group LTSI (Laboratoire Traitement du Signal et de l'Image). He was member of the Scientific Committee of the University of Rennes 1 between 2008 and 2010. He received in 2010 the H.D.R. (Habilitation to Lead Researches) degree in Sciences from the University of Rennes 1, France. His research interests include human electrophysiological inverse problems based on high-order statistics, sparsity and multidimensional algebra.