Online Recognition of Facial Actions for natural EEG-based BCI Applications

Dominic Heger, Felix Putze, and Tanja Schultz

Cognitive Systems Lab (CSL)
Karlsruhe Institute of Technology (KIT), Germany
{dominic.heger, felix.putze, tanja.schultz}@kit.edu

Abstract. We present a system for classification of nine voluntary facial actions, i.e. Neutral, Smile, Sad, Surprise, Angry, Speak, Blink, Left, and Right. The data is assessed by an Emotiv EPOC wireless EEG head-set. We derive spectral features and step function features that represent the main signal characteristics of the recorded data in a straightforward manner. With a two stage classification setup using support vector machines we achieve an overall recognition accuracy of 81.8%. Furthermore, we show a qualitative evaluation of an online system for facial action recognition using the EPOC device.

1 Introduction

There has been a great research interest on non-invasive BCI systems over the last couple of years. Numerous systems have been created that are successfully used as communication devices for people with disabilities, for prostheses control, in rehabilitation, or for mind-games. More recently, passive BCIs gain raising attention [14]. Passive BCIs are expected to provide computer systems with information about a user's cognitive and affective mental states. States such as attention, workload, or emotions have been recognized successfully by BCI systems. While BCIs can enhance the efficiency and user satisfaction (e.g. [10]), the integration of passive BCIs into human-machine interfaces causes many challenges which need to be overcome.

First and foremost, those parts of the EEG signal that do not origin from neurological phenomenons are considered as artifacts. Artifacts generated by the user's muscle activity, movements of the eyes, tongue, or electrical potentials caused by body movements may impact the complete frequency range of EEG signals. Therefore, they can interfere with the features used for BCI systems and cause misleading results. To avoid such artifacts, experimental setups are heavily controlled and often force the users to restrict their natural movement (e.g. refrain from muscle activity or eye blinks). However, for intuitive interaction with BCI systems, this is not acceptable and thus, artifacts are unavoidable and have to be dealt with.

Many current BCI systems simply ignore these effects, which causes poor recognition results when artifacts are present. One approach to deal with those influences is to try to clean the signals using automatic rejection, frequency filtering, regression techniques using EOG signals, or blind source separation [7]. However, the impact of such techniques on the signals strongly depends on the user's activity causing them. A blind use of artifact removal techniques may have effects that can be as harmful as the original artifacts themselves (e.g. due to overestimation of artifacts or removal of relevant brain activity). In order to avoid deteriorating effects by blind artifact removal we propose to use explicit classification of non-brain activity to enable specialized artifact reduction methods.

From a user's perspective, one may unconsciously learn to control a system that is susceptible to artifacts by mimicking EEG activity, for example by raising the eyebrows or producing eye blinks [13]. Furthermore, there is a strong relationship between facial actions and natural communication signals, such as social cues and emotions [2]. In addition to that, interesting behavioral information can be assessed from eye movements or the eye blinking rate. Both, facial activity and eye movement can be recognized by a passive BCI system from systematic artifacts to get additional information about the user. Therefore, for BCIs, recognition of non-brain activity is important to be removed when unwanted, but to be interpreted when it contains relevant information.

Facial activity can also be assessed using video or EMG electrodes. However, deriving it directly from the EEG signal comes with the benefit of avoiding additional sensors. Furthermore, it yields the most direct connection between the facial activity and the influence on the EEG signal which is important when artifact removal is of interest. Consequently, we present in this paper the explicit recognition of facial actions and thus to provide a means to directly reduce artifacts of facial mimics and eye movements in a passive BCI system. For this purpose we describe a new approach for recognition of 9 different facial actions including facial mimics and eye activity from data recorded by the Emotiv EPOC, a low-cost EEG head-set designed for the consumer market.

2 Related Work

Recognition of facial expressions using visual information is widely researched. Ekman and Friesen developed the Facial Action Coding System (FACS) [6], a taxonomy to characterize nearly any facial activity. The fundamental atomic facial muscles or small muscle groups involved in facial expressions are called Action Units (AUs).

Some articles (e.g. [2], [11]) reviewed findings on the neural basis of spontaneous and voluntary expression of facial actions. The reported activation patterns strongly depend on which facial action is performed and involve several parts of the brain, such as motor cortex, frontal cortex, and areas involved in emotional processing. EEG has only rarely been used to investigate brain activity associated with the expression of facial actions because of the artifact problems [11].

Only few works have addressed classification of facial actions by EEG. The first research paper that proposed a recognizer for voluntary facial expressions in the context of EEG based BCIs was Chin et al. [4]. They presented an extension of the Filter Bank Common Spatial Pattern algorithm to multiclass classification and showed very good classification performance of 86%. They evaluated the system on 6 types of facial expressions, i.e. smile, straight, wince, agape, stern, and frown. Data was recorded using a standard EEG system with 34 electrodes. For classification they used a Naive Bayes Parzen Window classifier which was extended by a decision threshold based classification mechanism, which increased the recognition results for classes with lower accuracies.

Boot [3] presented a system for facial expressions recognition from EEG. They used an EEG cap with 32 electrodes to discriminate between four expression classes corresponding to neutral, angry, smile, and angry-pout. They applied Common Spatial Patterns and used Linear Discriminant Analysis for classification. They found that predominantly muscle activity was classified and that frontal EEG electrodes are most important for classification performance.

The Emotiv EPOC neuro head-set is a low cost EEG acquisition system for the consumer market [1]. In contrast to traditional EEG caps it allows for rather unobtrusive EEG recordings. The wireless device can be attached by the user himself or herself within a very short amount of time. It is comfortable to wear and uses saline electrodes which do not require electrode gel in the user's hair. However, the data acquired by such a consumer device may be more challenging for automatic processing due to the lower signal quality. Emotiv provides software that aims to recognize facial expression of the user. The system can discriminate eye blinks, winks, clenching of teeth, movement of eye brows, smiling, smirking, and laughing [1]. The growing user community of the EPOC has built numerous applications with the device including the control of a wheelchair using mimics. However, the Emotiv software is only available as a black-box and no published information are available which describe the functionality and performance of the algorithms.

In this work, we extend on the existing systems by including classes which describe facial actions such as speaking and different types of eye movement typically appearing in less controlled experimental setups. We also use a much more comfortable recording device with less electrodes than a standard EEG cap to foster more natural setups.

3 Data Acquisition

3.1 Facial Actions

In our experiment participants executed the 9 different facial action classes listed in Table 1. We selected these facial actions because they are relevant for human-machine interaction and occur frequently as social signals in natural interaction (e.g. [2]). Furthermore, they have a significant influence on the EEG signal, which makes artifact handling necessary to obtain the brain signal.

Table 1. Facial expression classes and corresponding Action Units

Class name	Action Units involved
NEUTRAL	AU0 (Neutral)
SMILE	AU6 (Cheek raiser), AU12 (Lip corner puller)
SAD	AU15 (Lip corner depressor), AU17 (Chin raiser)
SURPRISE	AU4 (Eyebrow lowerer)
ANGRY	AU1 (Inner brow raiser) AU2 (Outer brow raiser)
SPEAK	AU50 (Speaking)
BLINK	AU45 (Eye blink with both eyes)
LEFT	AU61 (Eye movement to the left)
RIGHT	AU62 (Eye movement to the right)

The NEUTRAL facial action class is characterized by relaxed muscles and eyes focused on a point on the screen. SMILE corresponds to a broad smile that mainly involves muscles around the mouth and the cheek. It may be related to happy social signaling, however it is well known that the relationship between smiling and affective states is much more complex. A prototypical SAD expression moves the mouth corners downwards and the lower lip upwards. This pout expression can be related to a depressed or offended state. The ANGRY facial mimic moves the inner eyebrows into direction of the face center and gives an evil or angry impression. For the SURPRISE class subjects were instructed to tear up the eyebrows so that it creates wrinkles on the forehead. The expression gives the notion of being skeptical, surprised or puzzled. The SPEAK class contains speech produced by counting numbers aloud. BLINK corresponds to one eye blink with both eyes. LEFT contains horizontal eye movement to the left and RIGHT horizontal eye movement to the right.

Multiple of these classes share the same face region and muscle groups, which is expected to make the classification task more challenging and to give insights on the possible limitations of the assessment of facial actions by EEG technology. As we see from the involved AUs (see table 1), SMILE and SAD mainly involve muscles around the mouth, ANGRY and SURPRISE movement of the eyebrows, and BLINK, LEFT and RIGHT contain eye movements. SPEAK also involves complex activity of the facial muscles and additionally movement of the tongue.

3.2 Experiment Design

For development and evaluation of the system, data from five subjects have been recorded at the Karlsruhe Institute of Technology (KIT). Before the experiment all subjects were instructed shortly and practiced to perform the different facial actions according to the descriptions in section 3.1. To reduce 'artifacts' in the data, subjects were asked to avoid unrelated muscle and eye activity during the recording parts of the trials. The execution of the facial actions was very intuitive and natural for the subjects. However, small deviations from the FACS and variability in the execution of the facial actions might occur as the data has not been filtered for correct execution using validated EMG or video recordings.

For each trial an icon and the class name were presented on the computer screen. Subjects started the expression phase on their own by a key press. After

2 seconds of recording, a gray bar was shown for 4 seconds to avoid influences of the previous trial. Subsequently, the next trial started.

This procedure was repeated for 190 trials, i.e. 20 trials for each class, except for NEUTRAL which had 30 trials. To avoid temporal effects in the classification the facial action classes were randomly ordered.

3.3 Spectral Data Analysis

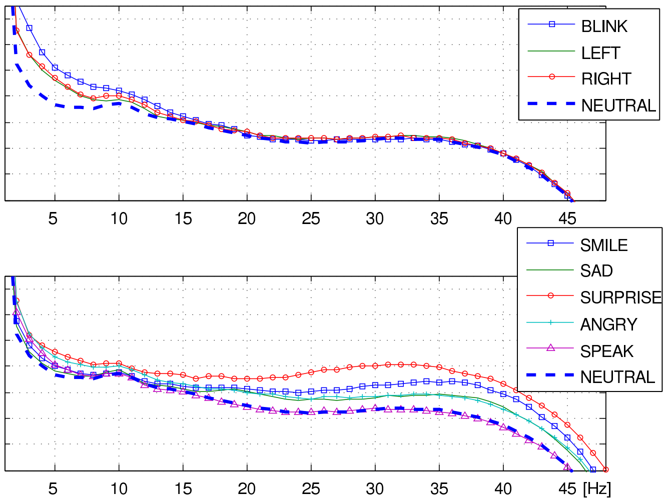

Fig. 1. Log power spectra of the facial action classes. Top: BLINK, LEFT, and RIGHT in contrast to NEUTRAL. Bottom: SMILE, SAD, SURPRISE, ANGRY, and SPEAK in contrast to NEUTRAL.

Biological artifacts, such as potentials caused by muscle activity, movements of eye or tongue, cause significant distortions on the electric fields over the scalp, which can easily be recognized in the measured signals by visual inspection. Figure 1 shows log spectrograms for the different classes averaged over all channels and all subjects. The first plot shows the eye activity related classes BLINK, LEFT, and RIGHT in contrast to the NEUTRAL class (dashed curve). A strong increase of power for frequencies below 15 Hz due to the eye activity can be observed. These potentials are caused by the movement of the retinal or cornea-retinal dipole and the eyelids [5]. They produce high-amplitude patterns in time-domain predominantly at the frontal electrodes. The second plot shows the mimics activity related classes SMILE, SAD, SURPRISE, ANGRY, and SPEAK in contrast to the NEUTRAL class (dashed curve). Muscle activity has a strong influence on a wide frequency range, with greatest amplitude between 20 and 30 Hz [8]. For the SPEAK class, we expected influences at low frequencies by movement of the dipole at the tongue (glossokinetic potentials) [12], as well as muscle activity from articulation at higher frequency bands. However, the spectrogram of Speak matches the one of Neutral very closely at frequencies above 10 Hz. This indicates that muscle activity due to articulation had only a small impact in our experiment. The frequency characteristics of all activity classes differ significantly from the Neutral class. Due to the characteristic impact of muscle, eye, and tongue movements on the signals, we expect that brain signals play a minor role in this classification task.

4 Recognition System

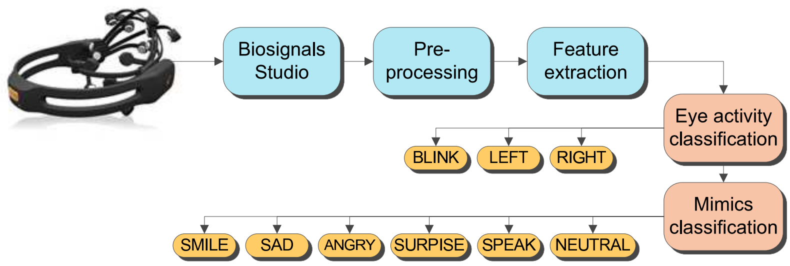

Fig. 2. Block diagram of the recognition process.

Figure 2 shows a block diagram of the system components involved in the recognition process. First, data is acquired from the Emotiv EPOC device using our recording software BiosignalsStudio [9]. The EPOC headset has 16 saline electrodes at positions AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 referenced to P3 and P4. The raw sensor data from the device has a sampling rate of 128 Hz and is notch filtered at 50 Hz and 60 Hz. We use the raw signal data of each trial, re-reference it to common average montage and remove baseline offsets and linear trends from each channel. Then 3 different types of features are extracted that model the main characteristics of the classes as described in section 3.3:

1. We estimate the spectral density in the rage 0-45 Hz using Welch's method. This results in a feature vector of 34 elements for each channel when using a window length of 0.75 seconds. Including higher frequencies showed no increase of recognition accuracy, which appears to be caused by a strong attenuation of spectral power above 45 Hz due to the notch filters of the EPOC device at 50 Hz and 60 Hz (see Figure 1).
2. Additionally we calculate the ratio between the frequency bands 0-10 Hz and 25-45 Hz as feature.
3. To calculate features that are able to describe horizontal eye movement activity, we use the potentials from electrodes left of the left eye and subtract them from the potentials assessed right of the right eye:

$$X_{HEOG} = (AF3 + F7) - (AF4 + F8)$$

To this time series, we optimally match two simple step functions, each consisting of two piecewise constant segments, by finding the largest local increase and decrease h_{up} and h_{down} of the level of the time series:

$$h_{up} = \arg\max_k \frac{1}{l}\sum_{i=1}^{l} X_{HEOG}[k-i] - \frac{1}{l}\sum_{i=1}^{l} X_{HEOG}[k+i],$$

$$h_{down} = \arg\min_k \frac{1}{l}\sum_{i=1}^{l} X_{HEOG}[k-i] - \frac{1}{l}\sum_{i=1}^{l} X_{HEOG}[k+i],$$

where l is the length of the interval before and after the step approximated by a constant function. For the experiments in this paper we chose $l = 20$ samples.

We apply a two stage classification scheme to recognize the facial actions. In the first stage, a linear support vector machine (SVM) discriminates the eye activity classes Blink, Left, and Right from a class consisting of the remaining facial actions. If the latter class is classified, the second stage uses a linear SVM to discriminate Neutral, Smile, Sad, Angry, Surprise, and Speak. Within both stages a one-vs-one approach is used for multiclass classification. SVMs with radial basis function kernels gave slightly worse classification results, which can be accounted to a higher robustness of the linear models when training with a small amount of data.

The two stage classification scheme allows to calculate specialized features for each of the two stages. In the first stage we use a feature vector composed of the spectral density features for each channel (1), h_{up}, and h_{down} (3). In the second stage, we use a feature vector composed of the spectral density features for each channel (1) and the power density ratio (2).

5 Evaluation and Results

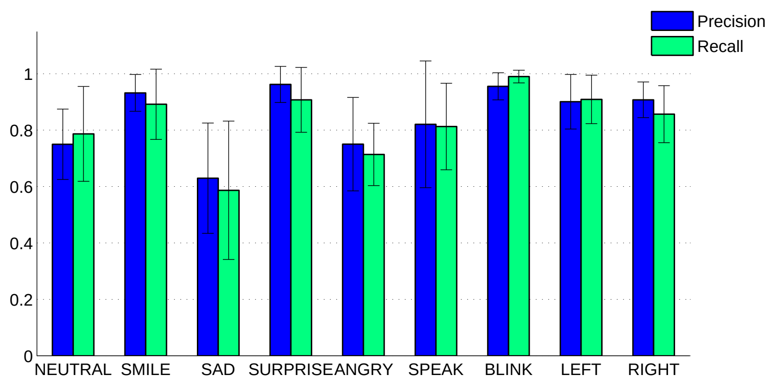

Fig. 3. Precision and recall for each facial action class averaged over the five subjects.

We evaluated the two stage classification system using 10-fold cross-validations. This resulted in the following recognition accuracies for the five subjects:

Table 2. Confusion matrix of the five subjects using the two stage classifier. Facial action classes are Neutral (N), Smile (Sm), Sad (Sa), Surprise (Su), Angry (A), Speak (Sp), Blink (B), Left (L), and Right (R).

	true **N**	true **Sm**	true **Sa**	true **Su**	true **A**	true **Sp**	true **B**	true **L**	true **R**
pred. **N**	119	3	23	1	6	10	0	2	6
pred. **Sm**	0	89	5	1	0	1	0	0	0
pred. **Sa**	10	5	56	2	12	4	0	3	0
pred. **Su**	1	0	0	90	2	0	0	0	0
pred. **A**	7	1	8	3	71	1	2	1	2
pred. **Sp**	11	0	6	1	1	80	0	0	3
pred. **B**	0	0	2	0	3	1	98	0	0
pred. **L**	1	0	0	1	3	1	0	90	5
pred. **R**	1	2	0	1	2	2	0	4	84

S1 86.8% (SD=5.1%), S2 71.1% (SD=7.6%), S3 81.1% (SD=11.2%), S4 89.5% (SD=10.2%), S5 80.5% (SD=6.6%) (means and standard deviations over the 10 iterations). The overall recognition accuracy was 81.8% (SD=7.1%).

Figure 3 shows precision and recall for each of the facial action classes averaged over the five subjects. Precision is the ratio between the number of times the class was correctly classified and the number of times the class was predicted by the classifier. Recall is the ratio of the number of times the class was correctly classified and the number of times the class should be classified according to the ground truth.

Table 2 shows the classification results in form of a confusion matrix summed over all iterations of the cross-validation of all subjects. The most frequent confusions occur between Neutral and Sad (19% off all misclassifications). For most facial action classes confusions occurred rarely in the experiment or are rather equally distributed across subjects. However, the confusions of Sad and Speak are predominantly caused by single subjects: The results of S2 account for 16 of 21 confusions between Neutral and Speak, and S5 causes 18 of 33 confusions between Neutral and Sad. This results in a high inter-subject variance for these classes, which can also be observed in the error bars of Figure 3.

An evaluation using only the frontal electrodes AF3, F7, F3, F4, F8, AF4 (to investigate the possibility of electrode reduction) shows considerably more confusions among the classes Neutral, Smile, and Sad, which results in a drop of average recognition accuracy from 81.8% to 67.7%. In contrast to Chin et al. [4], who associated the lower recognition results with missing information on activity of the motor cortex, we assume that in our case this is caused by muscle and eye movement activity measured at the non-frontal locations with the full electrode setup. This is indicated by the low coverage of the motor cortex by the EPOC device. Additionally, visual inspection of the temporal channels T7 and T8 of Smile shows strong muscle activity which dominates any activity emitted by the motor cortex.

To test the online capabilities of the facial action recognition system described in this paper, subject S1 performed each of the facial mimics for about 10 seconds. After that, 3 eye blinks were recorded, followed by 2 times moving

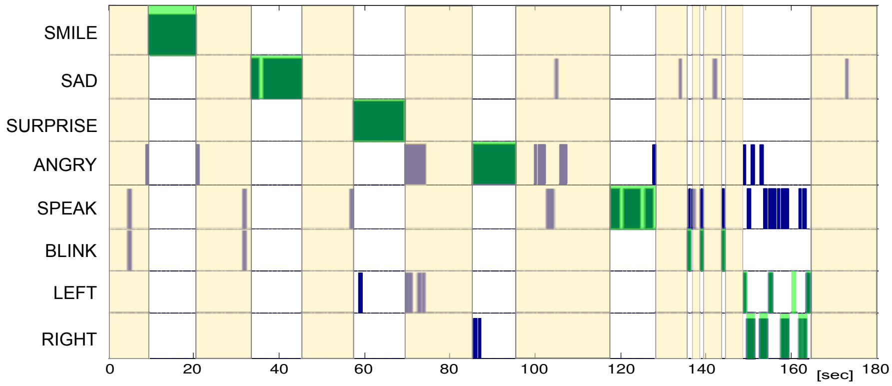

Fig. 4. Online classification results of a sequence of facial actions. Binary classification results are shown as green and blue bars for each second. Green overlays indicate time periods where a particular facial action should have been recognized, yellow overlays indicate time periods where Neutral should have been recognized. Therefore, dark-green bars indicate correct classifications, light-green regions indicate false negatives, light-blue and dark-blue bars indicate false positives.

the eyes to the left and back to center and 2 times to the right and back to the center. We modified the two stage classifier to output the recognition result of the second stage (for mimics classification) in addition to the first stage result (for eye activity recognition), when eye activity is recognized. This enables the system to recognize eye activity and facial mimics at the same time.

Figure 4 shows the online recognition results. Green overlays indicate the periods of time when a particular facial action should have been recognized, yellow overlays indicate periods of time when Neutral should have been recognized. A high precision of all facial action classes can be observed. However, recognition of Blink always implied a simultaneous recognition of Speak by the second classification stage. A similar effect can be found for Left and Angry. This indicates that the second stage classifier is influenced by the eye activity, which has not occurred in the training data. Removal of eye activity, e.g. by Independent Component Analysis, before classification by the second stage could mitigate this effect. Most recognition errors occur at start and end of an expression. This can be associated with the missing alignment in online recognition (i.e. windows can partly contain facial activity). Furthermore, the onset and offset of a facial actions can cause strong low frequent potentials. To cope with such effects, sequential classifiers, such as Hidden Markov Models, might be used to better model the dynamic character facial expressions.

6 Conclusion

In this paper, we showed that it is possible to use an EEG-based system for the effective classification of a large number of different facial actions. We showed that high recognition rates can be achieved using straightforward spectral features and step function features in a two stage classification scheme. Furthermore, we extended the system for online application and the more challenging recognition of parallel occurring facial actions. Such a recognition system can give additional insights on the user's behavior, as well as on the acquired signals that may be useful for artifact handling.

Further experiments are needed to investigate how the findings transfer to spontaneous facial actions in natural situations. In such a setup the ground truth of the performed facial actions could be assessed by EMG or video recordings.

References

1. Emotiv Software Development Kit User Manual for Release 1.0.0.4
2. Blair, R.: Facial expressions, their communicatory functions and neuro–cognitive substrates. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 358(1431), 561 (2003)
3. Boot, L.: Facial expressions in EEG/EMG recordings. Master's thesis, University of Twente (2009)
4. Chin, Z., Ang, K., Guan, C.: Multiclass voluntary facial expression classification based on filter bank common spatial pattern. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. pp. 1005–1008. IEEE (2008)
5. Croft, R., Barry, R.: Removal of ocular artifact from the eeg: a review. Neurophysiologie Clinique/Clinical Neurophysiology 30(1), 5–19 (2000)
6. Ekman, P., Friesen, W.: Facial action coding system. Consulting Psychologists Press, Stanford University, Palo Alto (1977)
7. Fatourechi, M., Bashashati, A., Ward, R., Birch, G.: Emg and eog artifacts in brain computer interface systems: a survey. Clinical Neurophysiology 118(3), 480–494 (2007)
8. Goncharova, I., McFarland, D., Vaughan, T., Wolpaw, J.: Emg contamination of eeg: spectral and topographical characteristics. Clinical Neurophysiology 114(9), 1580–1593 (2003)
9. Heger, D., Putze, F., Amma, C., Wand, M., Plotkin, I., Wielatt, T., Schultz, T.: Biosignalsstudio: a flexible framework for biosignal capturing and processing. In: KI 2010: Advances in Artificial Intelligence. pp. 33–39 (2010)
10. Heger, D., Putze, F., Schultz, T.: An adaptive information system for an empathic robot using eeg data. In: Social Robotics. pp. 151–160 (2010)
11. Korb, S., Grandjean, D., Scherer, K.: Investigating the production of emotional facial expressions: a combined electroencephalographic (eeg) and electromyographic (emg) approach. In: Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on. pp. 1 –6 (sept 2008)
12. Vanhatalo, S., Voipio, J., Dewaraja, A., Holmes, M., Miller, J.: Topography and elimination of slow eeg responses related to tongue movements. Neuroimage 20(2), 1419–1423 (2003)
13. Wolpaw, J., Birbaumer, N., McFarland, D., Pfurtscheller, G., Vaughan, T.: Brain-computer interfaces for communication and control. Clinical neurophysiology 113(6), 767–791 (2002)
14. Zander, T., Kothe, C., Jatzev, S., Gaertner, M.: Enhancing human-computer interaction with input from active and passive brain-computer interfaces. Brain-Computer Interfaces pp. 181–199 (2010)