



Switching between manual control and brain-computer interface using long term and short term quality measures

Alex Kreilinger¹, Vera Kaiser¹, Christian Breitwieser¹, John Williamson², Christa Neuper^{1,3} and Gernot R. Müller-Putz^{1*}

¹ Laboratory of Brain-Computer Interfaces, Institute for Knowledge Discovery, Graz University of Technology, Graz, Austria

² Department of Computing Science, University of Glasgow, Glasgow, UK

³ Department of Psychology, University of Graz, Graz, Austria

Edited by:

José Del R. Millán, Ecole Polytechnique Fédérale de Lausanne, Switzerland

Reviewed by:

Robert Leeb, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Thorsten O. Zander, Max Planck Institute for Intelligent Systems, Germany

*Correspondence:

Gernot R. Müller-Putz, Laboratory of Brain-Computer Interfaces, Institute for Knowledge Discovery, Graz University of Technology, Krenngasse 37, Graz 8010, Austria.
e-mail: gernot.mueller@tugraz.at

Assistive devices for persons with limited motor control translate or amplify remaining functions to allow otherwise impossible actions. These assistive devices usually rely on just one type of input signal which can be derived from residual muscle functions or any other kind of biosignal. When only one signal is used, the functionality of the assistive device can be reduced as soon as the quality of the provided signal is impaired. The quality can decrease in case of fatigue, lack of concentration, high noise, spasms, tremors, depending on the type of signal. To overcome this dependency on one input signal, a combination of more inputs should be feasible. This work presents a hybrid Brain-Computer Interface (hBCI) approach where two different input signals (joystick and BCI) were monitored and only one of them was chosen as a control signal at a time. Users could move a car in a game-like feedback application to collect coins and avoid obstacles via either joystick or BCI control. Both control types were constantly monitored with four different long term quality measures to evaluate the current state of the signals. As soon as the quality dropped below a certain threshold, a monitoring system would switch to the other control mode and vice versa. Additionally, short term quality measures were applied to check for strong artifacts that could render voluntary control impossible. These measures were used to prohibit actions carried out during times when highly uncertain signals were recorded. The switching possibility allowed more functionality for the users. Moving the car was still possible even after one control mode was not working any more. The proposed system serves as a basis that shows how BCI can be used as an assistive device, especially in combination with other assistive technology.

Keywords: brain-computer interface, BCI, hybrid BCI, assistive technology, electroencephalography, EEG

1. INTRODUCTION

Brain-computer interfaces (BCIs; Wolpaw et al., 2002) provide a means of communication for patients who have lost most of their residual muscle functions and are therefore incapable to interact with their environment. Examples of these kinds of severe impairments are people suffering from symptoms of amyotrophic lateral sclerosis (ALS), people in a locked-in state, and people who have a spinal cord injury close to the brain.

A BCI makes use of brain signals which can be derived from various sources with different methods. In this study we used a non-invasive method to record electrical brain signals, the electroencephalogram (EEG; Mason et al., 2007). EEG-based BCIs can be subdivided into three categories according to the used signal types: first, dynamics of brain oscillations such as event-related (de)synchronization (ERD/ERS; Pfurtscheller and Lopes da Silva, 1999) which establish the basis for motor imagery (MI) BCI (Pfurtscheller and Neuper, 2001; Neuper et al., 2006); second, steady-state evoked potentials (SSEPs; Middendorf et al., 2000; Müller-Putz et al., 2006); and third, evoked potentials (Regan, 1989) with the well-known example, the P300 (Farwell and Donchin, 1988).

The benefit of BCI is the independence from any remaining muscular functions, which means that muscle fatigue is irrelevant. However, one major drawback with BCIs is that the performance for most users is still far from perfect. BCIs are often afflicted with low bit rates, low accuracy, and bioelectrical signals are generally prone to be corrupted with artifacts. Because it is difficult to improve BCI technology itself, applications could be developed that make better use of BCIs, acknowledging the advantages and disadvantages and deal with them in the most appropriate way. For example, a BCI can be used to provide additional communication channels on top of other assistive devices that are used by people who still have some residual motor functions (Rupp and Gerner, 2004).

To increase the attractiveness of BCI technology for patients it is essential to find practical applications that provide maximum control at all times, depending on the current physical and/or mental condition of the patient. Thus, providing the best means of communication at any time would be reasonable. As long as residual motor functions are still working, they offer a more reliable and natural communication channel. However, due to fatigue and/or additional interferences like tremors or spasms, a signal

based on motor functions may lose its control capability after a long time of usage. At this moment, a switch to a control mode which is not based on muscular activity might become a lot more attractive and could be used to restore control over the assistive system. This approach can be realized by using a multimodal interface (Blattner and Glinert, 1996; Jaimes and Sebe, 2007) which is able to deal with at least two different control signals. A particular multimodal interface which incorporates BCI is called hybrid BCI (hBCI; Scherer et al., 2007; Allison et al., 2010; Millán et al., 2010; Pfurtsceller et al., 2010). Here, a BCI is combined with any other user-driven signal. This signal can be a biosignal like electromyogram (EMG), electrocardiogram (ECG), electrooculogram (EOG), or EEG not used for BCI, but also sensor signals and other control signals generated from assistive devices like shoulder joysticks, mouses, buttons, and eye trackers (Zander et al., 2011). Moreover, the use of hybrid BCIs may be an interesting tool for healthy users in special working environments where common control mechanisms are unreliable or not enough, e.g., operating an additional EEG-based command in a spacesuit, or also in the field of gaming (Zander and Kothe, 2011).

According to (Pfurtsceller et al., 2010), an hBCI must fulfill following four conditions: “(i) the device must rely on signals recorded directly from the brain; (ii) there must be at least one recordable brain signal that the user can intentionally modulate to effect goal-directed behavior; (iii) real time processing; and (iv) the user must obtain feedback.” The hBCI introduced in this paper will follow these definitions except one small deviation: the BCI provided is purely optional, just like any other input into the system; users are not forced to use BCI when there is a better alternative. This approach concurs with the concept developed and described in (Millán et al., 2010). A more detailed description of the hBCI platform can be found in (Müller-Putz et al., 2011).

The combination of multiple inputs can be handled in a few different ways: (i) each input can be linked to a single application; (ii) all the inputs are fused and weighted to generate a single output which controls an application (Leeb et al., 2011); (iii) a monitoring module monitors inputs and decides which is best suited to be used as a control signal.

The goal of this work was to evaluate a practical combination of multimodal inputs with the sole purpose of making an application more usable for patients. This means, on the one hand, that a system should be easy to use and functional all the time by providing different options to communicate with it, but also, on the other hand, that an application can be controlled for a longer time than usual. Interaction with the assistive device should still be possible after the primary control strategy would no longer be possible due to fatigue and/or a growing lack of concentration. Therefore, the hBCI system presented in this paper is relying on the approach (iii): a monitoring module monitors inputs and decides which is best suited to be used as a control signal. A joystick (JS) signal to simulate assistive devices and a control signal derived from an MI-based BCI were constantly monitored and weighted to achieve a solution with long functionality for the user. The weighting was based on four individual quality measures per control mode. These measures were designed to detect signal specific artifacts and malfunctions, e.g., noisy EEG or a joystick signal made unusable due to strong tremors.

The proposed combination of inputs was used in a car game. A constantly moving car could be controlled with either one of the two inputs to collect coins and avoid obstacles. We investigated how well the selection of quality measures could detect a low performance, caused by a poor signal quality. To speed up the simulation an artificial deterioration was used for the joystick signal to simulate signal impairments that can be expected from patients. BCI was not deteriorated, as artifacts were expected to occur all the same.

Additionally, we investigated how the switching capability increased the maximum score when compared to a simulation without switching.

After running the experiments and evaluating the data we could show increased scores and a trend that links good performance during the car game with the quality rating determined by the quality measures.

2. MATERIALS AND METHODS

In the feedback application subjects could move a car on a vertically scrolling street, see **Figure 1**. On the sides of the street coins and obstacles (barriers) appeared randomly. Subjects were asked to collect as many coins as possible with the car while avoiding obstacles. The car was controlled either manually with a joystick, or mentally with BCI. The joystick represented any kind of assistive device, relying on muscular functions. This device could stop working permanently, after a long usage due to fatigue, or temporarily, during periods of tremor and spasm.

BCI on the other hand is prone to noise, distraction, and fatigue as well. Considering the drawbacks of both input modalities, the system offered switching between inputs to increase flexibility and

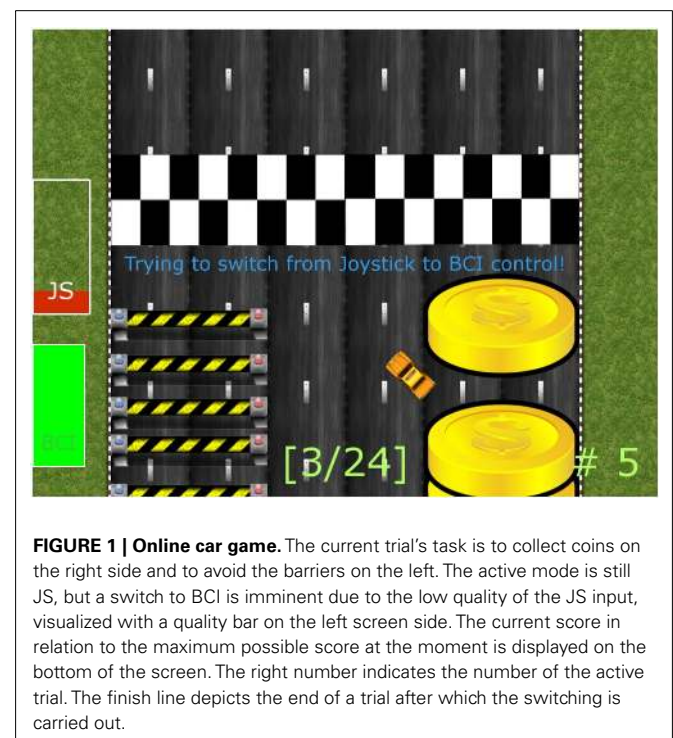


FIGURE 1 | Online car game. The current trial's task is to collect coins on the right side and to avoid the barriers on the left. The active mode is still JS, but a switch to BCI is imminent due to the low quality of the JS input, visualized with a quality bar on the left screen side. The current score in relation to the maximum possible score at the moment is displayed on the bottom of the screen. The right number indicates the number of the active trial. The finish line depicts the end of a trial after which the switching is carried out.

functionality for the users. The switching was carried out whenever the quality rating (QR) of one signal was considerably worse than the other's. An overview of the proposed system is demonstrated in **Figure 2**.

2.1. INPUTS

Both inputs, BCI and joystick, provided control signals from -1 to $+1$ where -1 would move the car to the left side of the street and $+1$ to the right. The joystick was limited mechanically so it could not generate values out of this range. BCI, which used an LDA classifier to discriminate between two MI classes, was saturated at -1 and $+1$. The joystick signal was further processed with artificial artifacts.

2.2. ARTIFICIAL ARTIFACTS FOR THE JOYSTICK INPUT

To simulate the system on healthy subjects instead of patients the joystick signal was deteriorated with artificial noise. This deterioration can be expected from patients with a spinal cord injury at C4/C5 which causes loss of hand control and heavily limited shoulder function. The artificially induced deterioration included tremors (Anouti and Koller, 1995), spasms (Kawamura et al., 1989), and an increasing weakness over time. To speed up the simulation, unrealistically high values were chosen: maximum fatigue was reached within minutes and tremors and spasms occurred frequently as long as fatigue was still low.

2.2.1. Tremors

During periods of a tremor a heavily shaken JS can be expected which renders control completely unreliable. We simulated this effect by adding a normally distributed random signal, band-pass filtered between 2 and 10 Hz, to the recorded JS signal. The tremor signal's amplitude and probability of occurrence was inversely

proportional to the current weakness. Every 20 s, with a probability of $p = 100 - \text{weakness level in \%}$, either a tremor or a spasm was triggered at random. The amplitude of the tremor signal was affected directly by the current weakness as the whole JS signal was decreased.

2.2.2. Spasms

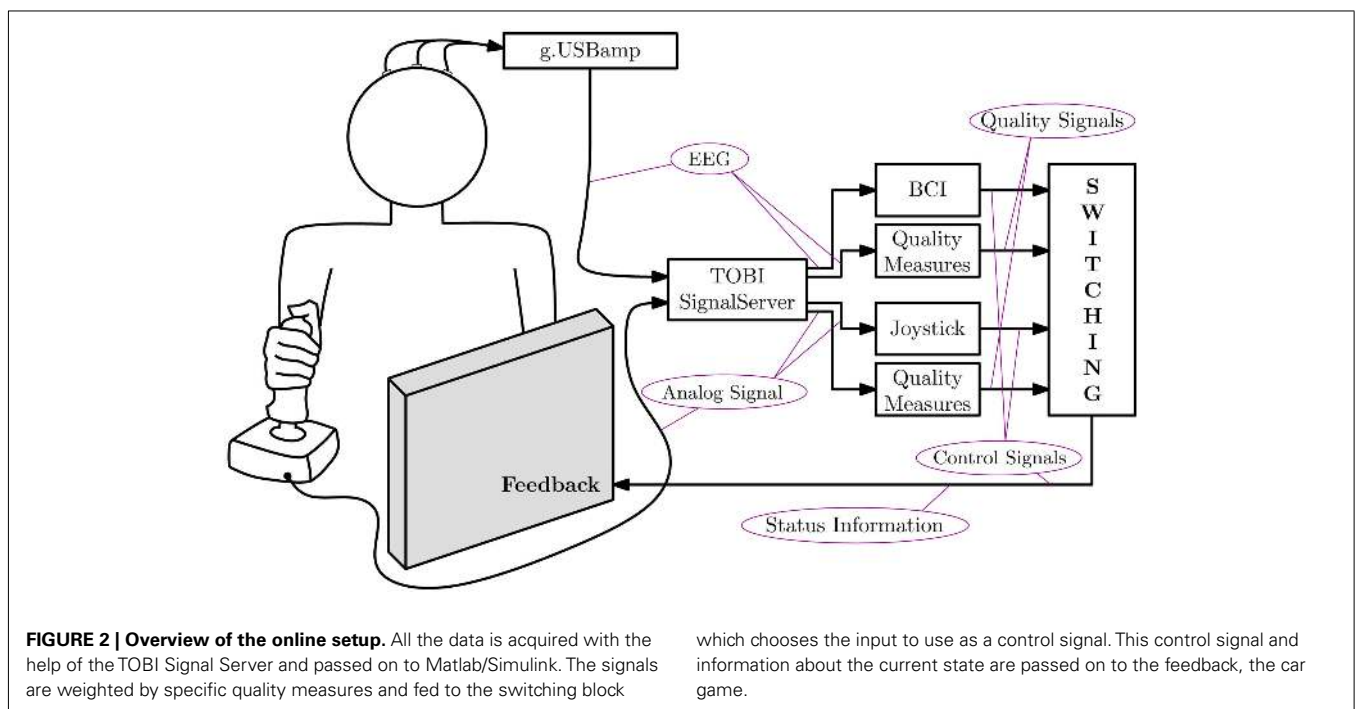
These involuntary muscle contractions can also have a strong and negative effect. We simulated spasms by applying a heavy bias to either the left or the right. The same rules were applied here as for the tremor activation. The added bias was also reduced with the weakness level.

2.2.3. Weakness

The most important factor was the weakness as it was used to simulate fatigue. The parameters were set to allow a stepwise increase of weakness after each trial. A weakness level of 0% indicated no impairment, whereas 100% were reached as soon as subjects were no longer able to collect coins due to the strong reduction of amplitude. How fast the maximum weakness was reached depended on the stage of the experiment, either at the 10th or the 30th trial. Weakness could recover, with the same rate it was increased before, during times of no active joystick usage.

2.3. QUALITY MEASURES

The currently active control signal was evaluated with four specific long term quality measures. These measures were customized to check specifically for indicators of a bad quality. These indicators could be a high noise level or unreliable behavior like an unstable classifier output. Both signals were measured individually. On top of these long term measures, both input types had one short term measure. Basically, short term measures were an additional effect when the worst indicators, used for long term



measures, were detected, e.g., noisy EEG and a heavily shaking joystick. When short term measures were active, control of the car was no longer possible because the car was fixed to the middle lane.

2.3.1. Joystick measures

The quality score depended on following factors: shaking caused by tremors, low range of movement as a result of weakness, invariability (a total lack of movement possibly indicating hardware defects), and bias induced by an imbalanced preference of one side which can be caused by spasms.

2.3.2. BCI measures

These measures monitored the noise of the EEG signal caused by EMG artifacts, instability due to unreliable classification, invariability (indicating hardware or software errors), and also bias for a one-sided classifier output.

The problem about setting the weights for the measures was to find the right values in order to rate negative effects correctly. That is, very strong impairments should cause a fast decrease of quality and minor impairments only a slow decrease. These weights were initially set to arbitrary values. However, we tried to give strong negative effects like strong noise and shaking heavier weights. In contrast, not so acute effects were weighted lower, e.g., effects like fatigue whose influences built up over a longer time. All weights were adjusted empirically while conducting preliminary tests before running the final experiments.

The differently strong impacts on the quality for the final measures were as follows: BCI noise and JS shaking would decrease the quality by 10(%/s), BCI instability by 5(%/s), low JS amplitude by 2(%/s), BCI and JS invariability by 1(%/s), whereas the bias was proportional to the bias itself; a strong bias over a long time would increase the weight steadily. All those measures were able to recover individually whenever they were not currently detected. Additionally, all individual quality measures for one control mode recovered when the other mode was currently active; i.e., BCI measures recovered with 1(%/s) during joystick mode and vice versa. The currently inactive signal was never monitored, i.e., the quality of an inactive signal was only allowed to increase, not decrease. The described measures are also demonstrated in **Table 1**.

$$QR_{BCI/JS} = 100 + \sum_{i=1}^N \int \mathbf{w}_i(x(t)) dt$$

$$\text{with } x(t) = \begin{cases} 1 & \text{current mode, noise active} \\ 2 & \text{current mode, noise inactive} \\ 3 & \text{other mode, recovery} \end{cases} \quad (1)$$

e.g., $\mathbf{w}_{1,BCI} = \begin{bmatrix} -10 \\ +3 \\ +1 \end{bmatrix}$; $\mathbf{w}_{2,JS} = \begin{bmatrix} -2 \\ +4 \\ +1 \end{bmatrix}$

Equation 1 shows a simplified formula of how the qualities for both control modes were calculated during the online experiment. In the equation QR depicts the quality rating of one of the two control signals which always ranges between 0 and 100%. \mathbf{w}_i

Table 1 | BCI and JS quality measures.

Measures	BCI		Measures	JS	
	QR↑↓ (%/s)			QR↑↓ (%/s)	
	↓	↑		↓	↑
EMG noise	-10	+3	Shaking	-10	+2
Instability	-5	+1	Low amplitude	-2	+4
Invariability	-1	+4	Invariability	-1	+4
Bias	α bias	α bias	Bias	α bias	α bias

Four measures for both control modes, BCI and joystick, are shown. These measures can either decrease the quality (100% + numbers in the second and fifth column) when they are currently detected but also recover over time otherwise (third and sixth column). The bias' measure, as an exception, is depending on the magnitude of the bias itself. The decrease rate of the QR is higher than the recovery rate to allow for a quick response in case of bad input signals. All quality measures of one mode recover with 1(%/s) when the other mode is active at the moment.

describes the $N=4$ different weight vectors, one for each quality measure. The indexes of these vectors depend on the current state of detected criteria and on the actual active control mode. The weights either increase or decrease the whole quality of one mode. The equation is simplified inasmuch as it does not include context-sensitive factors that were considered additionally online. These factors are represented in the following list which explains the four long term quality measures per control mode, the relation with short term measures, and how exactly measures were combined in the online model:

- BCI EMG noise: all EEG channels were filtered between 20 and 100 Hz, squared, averaged with a moving rectangle window of 1 s, and logarithmized. A threshold was set before online measurements after subjects were instructed to produce EMG noise and clean EEG. EMG noise was only detected within active trials. When detected, an integrator would start to increase from 0 to 100% with 10% every second. Otherwise the integrator could recover toward the minimum value of 0% with $-3(\%/s)$. These and following values are given in **Table 1**, however with inversed signs. EMG noise, when detected, also triggered the short term measure of the BCI signal. This effect was shown in the car game by fixing the car to the middle lane and a swiveling animation of the car.
- BCI instability: this measure was based on the number of zero crossings within active trials. As soon as the middle line of the street was crossed more than three times within one trial, a second integrator increased from 0 to 100% until the trial was over. Anytime else the value could recover.
- BCI invariability: a total lack of LDA variance after 1 s started a very slow increase of a third integrator, also limited between 0 and 100%, which would decrease four times faster in case of any movement. This measure was active at all times.
- BCI bias: as soon as the system switched to BCI, the bias was measured constantly. A one-sided classifier output resulted in a continuously increasing weight. In detail, when the absolute

value of the LDA classifier exceeded 0.2, an integrator was growing either toward +100 or -100%, depending on the sign of the LDA output. This value was multiplied by 1/20. The absolute value was then subtracted with 1 and the final weight passed on to the fourth integrator. This way, the bias weight could lie between 1 and 5(%/s) and recovered with 1(%/s), due to the subtraction.

- BCI combination: the four BCI integrators were either increased or decreased with the specific weights or each decreased with -1(%/s) when JS was currently active. The outputs of all four integrators were added but the total sum was limited to 100%. This total value was finally subtracted from the current BCI QR which started at 100%.
- JS shaking: the absolute value of the derivation of the JS input was smoothed with a moving rectangle window of 1 s length. A predefined threshold was found in foregoing measurements. If the threshold was exceeded, the integrator increased and decreased otherwise. This measure was active all the time. JS shaking was the equivalent to BCI EMG noise in terms of the short term measure, a detection of shaking also rendered control of the car impossible.
- JS low amplitude: the absolute value of the JS input was compared to the threshold value 1/3. Between $\frac{-1}{3} - \frac{1}{3}$ no object collection was possible. Only during active trials, a second JS integrator started to increase or decrease, depending on whether the JS input was below the threshold or not.
- JS invariability: to detect hardware errors, this measure was applied exactly the same way as BCI invariability.
- JS bias: the bias measure was similar to the BCI bias measure. The only difference being the multiplication of 1/10 and a subtraction of 0.5. The resulting weights could therefore range from 0.5 to 10(%/s) with a recovery rate of 0.5(%/s).
- JS combination: JS weights were combined the same way as BCI measures, however, only when the active control mode was JS.

2.4. EXPERIMENT SETUP

The experiment was designed to allow completion within one session, not longer than 3 h. It consisted of three steps: (i) two runs of offline BCI training to set up a classifier for the MI BCI; (ii) two runs with a car game controlled only with a joystick to collect data of runs without the switching system; (iii) six runs with combined BCI and manual control to analyze how well the designed switching approach worked online.

EEG was recorded with a g.USBamp amplifier (g.tec medical engineering GmbH, Austria, Graz). Six Ag/AgCl electrodes were placed anterior and posterior to C3, Cz, and C4 to obtain three bipolar channels. Data was recorded with a sample rate of 512 Hz and filtered between 0.5 and 30 Hz and an activated notch filter at 50 Hz. After the BCI training session, which only needed pure EEG, a joystick was added that provided an analog signal between -1 and +1. This analog signal was later used to control the car; -1 would move the car to the leftmost side of the street; +1 to the right. Both input types were acquired with the TOBI Signal Server (<http://www.tobi-project.org/download>; Breitwieser et al., 2011), a software that is able to combine multiple inputs and provide data in a standardized and synchronized way for various clients via network protocols.

2.4.1. BCI training

In the beginning, two short BCI training runs were carried out, each with 40 randomized trials of movement imagination, one half both feet and the other half right hand. Subjects performed the standard Graz-BCI training paradigm (Pfurtscheller and Neuper, 2001) to allow selection of features and calculation of a classifier for MI. Trials contaminated with artifacts were removed manually before searching for relevant features. The features consisted of frequency bands recorded over the three bipolar channels. They were selected manually after evaluating ERD/S maps (Grimann et al., 2002). Here, the frequency bands with the most significant differences between hand and feet MI were selected by plotting difference maps of both classes. ERD/S maps showed only significant changes ($\alpha = 0.05$) of frequency band-powers after the cue compared to a reference period between 1.5 and 0.5 s before the cue. The difference maps only showed significant differences between two classes in the time after the cue with the same significance level.

The band-powers of the best found frequency bands were used to generate an LDA classifier. A time window, covering the time between cue appearance and end of trial (5 s), was processed in 100 ms steps. At each step, the features corresponding to one time step were used to calculate a temporary classifier with which the data was analyzed by a 10×10 cross-validation. As soon as the whole time window was tested, the best point in time was used to set up the final classifier with the whole data set for online measurements. Additionally, the 10×10 cross-validation was nested within a 10×5 outer cross-validation that split data into an outer training set and a validation set. Here, classifiers generated at the best points in time, which were found via an inner cross-validation, were applied on unseen data to make sure that these points were really stable and to evaluate potential overfitting.

An online LDA classifier generates two outputs: the class label (-1 or +1) and the distance (an analog value). The distance was used to control the car in the later online runs. Scale and bias of the classifier were adjusted to achieve a distance between -1 and +1 on average, similar to the joystick range, with an average of zero for both classes combined. A possible transgression of -1 or +1 resulted in a saturation during later online experiments. A classification of foot MI would result in a negative distance value and move the car to the left; a classification of hand MI in a positive value and a movement to the right.

Additionally, the offline performance of this chosen classifier was evaluated by testing the classifier's accuracy on all 100 ms steps between 1 s after cue appearance and end of trial.

2.4.2. JS only

The second part of the session simulated a system without BCI to have a comparison of data with only joystick control and data with joystick and BCI combined. Subjects were asked to perform two runs controlling the car game, see **Figure 1**, with just the joystick. The participants were asked to collect coins and to avoid obstacles with a moving car on the screen in front of them. The car was constantly driving with a fixed speed toward the top edge of the screen. One single trial included a sequence of coins and barriers which appeared at the top of the screen, always six coins in a row with six barriers on the opposite side of the street. The interval between

coin/barrier and next coin/barrier was 0.5 s. These objects could be reached by the car exactly 4 s after they appeared. The joystick signal was being deteriorated over time in a way that it reached the maximum weakness at 30 trials, out of 40 trials per run; the subjects were not supposed to be able to collect anything during the final trials, because switching to BCI mode was not yet possible.

2.4.3. JS + BCI

The final part of the experiment combined manual with BCI control. Six runs with 40 trials each were carried out. Before starting to record the online runs, subjects were asked to perform a few trials in BCI mode to let the supervisors adjust the bias and the scale of the classifiers, if necessary.

The setup was the same as before with two differences. First, the participants could control the car with the joystick or with the BCI by performing the previously trained MI tasks. Second, the JS weakness reached its maximum already at 10 trials.

An overview of the setup is demonstrated in **Figure 2**. The runs always started with active JS control. The JS signal deteriorated continuously in order to simulate weakness and to force a switch to BCI. A switch was only permitted to happen when the active quality was below 20% and the other one above 50%. Additionally, switches were never triggered within active trials, instead, the system waited for a break to switch to the other mode. The length of this break was automatically increased by 5 s to allow accommodation to the other control mode.

To facilitate switching back after some time, the inactive signal's quality could recover by 1(%/s) per criterion. Subjects were asked to avoid switching as long as possible, i.e., to avert quality reducing factors. The measures that affected the quality were called long term measures.

Additionally, so-called short term measures were used to inhibit control during times of severe noise impairment by forcing the car to the middle of the street and giving a visual alert (swiveling of the car). In BCI mode this could have happened during a detection of noise; in JS mode the inhibition was caused by a detection of strong shaking. The long term quality measures were not influenced by this inhibition: the noise/shaking measure and eventual other measures could still decrease the QR of the current control mode. **Figure 1** shows an excerpt from the ongoing feedback during an online run. Here, the subject was currently collecting coins but the system decided to switch from JS to BCI mode since the JS quality had fallen below the threshold of 20%.

2.5. EVALUATION

After all the runs were conducted, the recorded data was evaluated with following methods.

2.5.1. Score, collection rate, performance measure

We analyzed how well the subjects performed in terms of collected points with "JS only" compared to "JS + BCI" control. This outcome was rated in three different ways:

- (i) online scoring was based on adding or subtracting points. Subjects could increase the score +1 by collecting a coin and decrease it with -1 in case of a collision with a barrier. To avoid frustration, the score could never fall below zero. The maximum score in one run was 240 (6 × 40 coins);

- (ii) offline, the rate of positive: negative collection was analyzed. Only the relation between collected coins and barriers was of interest, not the percentage of collected objects out of the maximum possible number. Left out objects on the street were not taken into account (e.g., a missed coin or barrier);
- (iii) also offline, a performance measure was introduced, depending on collected coins, barriers, and left out objects. This trial-based performance measure ranged between 0 and 100%. One hundred percent indicated that all possible coins within a single trial were picked up, 50% that either no object at all or the same number of coins and barriers were collected, and 0% were achieved when only barriers were hit. This specific performance measure could be directly compared to the mode-specific QR over time. Later mentioned performance refers to this kind of performance measure. Equation 2 shows how the performance measure per trial (PM_{trial}) was calculated. The $Score_{\text{trial}}$ could range between -6 and 6 points, the max ($Score_{\text{trial}}$) was 6 points.

$$PM_{\text{trial}} = Score_{\text{trial}} + \frac{50\%}{\max(Score_{\text{trial}})} + 50\% \quad (2)$$

2.5.2. Correlation of time and performance

We evaluated the correlation of BCI performance with time in BCI mode. Because the quality monitoring was purely based on characteristics of the inputs and not on the online performance itself, it was not guaranteed that these two values would show a correlation. However, it would be a good sign of a working switching approach if it was found to be true.

2.6. SUBJECTS

Ten healthy subjects took part in the study, all of them had experience in BCI to permit a short training session of just two runs. Based on results from previous experiments we selected BCI performers with two class accuracies above 60%. They were aged between 21 and 30 years (25.4 ± 3.1 years), half of them were female, and all of them right handed.

3. RESULTS

3.1. BCI TRAINING

The first two runs of BCI training provided good classifiers for all the subjects. **Table 2** shows the mean accuracies in the time period 1–5 s after the cue when applying the classifier that was generated after the best point in time was found by the search with the 10 × 10 cross-validation. Additionally, the accuracies at the best points in time are shown for each subject. Furthermore, the best points in time found via inner cross-validation were used to create classifiers that were applied on validation data sets within a 10 × 5 outer cross-validation to check for overfitting. The results, after averaging the achieved accuracies, are also listed in **Table 2**. Therefore, overfitting was shown to not be a large problem. This was done by comparing accuracies achieved by only using the best points in time, without a nested cross-validation, to accuracies that were found with the best points in time, found via inner cross-validation, and tested with a validation set in each loop of the outer cross-validation. The averaged outer cross-validation accuracies were only 2.9% lower than the accuracies achieved without nested cross-validation.

3.2. ONLINE CAR GAME

Figure 3 shows the point collection rates, averaged over all subjects, and all conducted runs. The collected points for both parts are compared: “JS only” on the left and “JS + BCI” on the right. For both types the first increase in points is caused by the JS control which is beginning to stagnate as soon as the artificially appended weakness has reached its maximum at trial number 10 in the combined “JS + BCI” runs, and at 30 trials for “JS only.”

The important noticeable difference is that the score starts to increase after the quality of the joystick signal decreased enough to trigger the first switch to BCI in the combined paradigm. In “JS only” mode there was no way to further increase the score. After the time of the first switch, subjects remained in BCI mode for different amounts of time but also had the possibility to go back to JS in case of a bad EEG input. This is also indicated with the increasing SD in the plot. The maximum number of points per run was 240 (40 trials, each with six coins).

Figure 4 illustrates how the monitoring system worked online with the example of the current performance, the signal qualities,

the actual control modes, and the occurrence of switches from BCI to JS. Only BCI → JS switches are highlighted to maintain clarity of the figures. To demonstrate the quality evaluation effects two subjects were chosen to represent a good (BCI performance of $77 \pm 29\%$, subject number 1) and a bad (subject number 6 with $57 \pm 25\%$) BCI performer. The subjects were picked according to the values shown in **Table 3**, taking into account the BCI performance and the time in BCI mode. For each of them the left plot shows the course of actions over the whole time (data from all 6 runs), whereas the right one shows one exemplary run to show performance and BCI quality in more detail. The figures consist of five features: (i) the performance, as mentioned in 2.5, visualizing the general performance which is relying on collected coins, barriers, and left out objects; (ii) the QRs of BCI and JS mode, as obtained online by the four specific long term quality measures. The JS quality is only shown in the examples with the single run; (iii) the occurrence of actual system-induced switches from BCI to JS; (iv–v) indication of the current control mode, either BCI or JS.

The correlation between BCI performance and time in BCI mode is addressed in more detail in **Table 3** and **Figure 5**. The table lists the collection rates and performances in BCI mode and the according times actually spent in this mode. The figure demonstrates the relation between these two values. The plot shows two linear fits, one of them using all subjects and one with subject S8 removed as an outlier. The correlation coefficients were $r=0.34$ ($p=0.33$) and $r=0.6$ ($p=0.09$), respectively. Apparently, S8’s BCI quality was not recognized as a poor one. The measure weighting the bias was not strong enough to decrease it sufficiently to cause a switch but the bias was strong enough to cause a bad performance. Also, the other measures were not triggered very often in order to have an effect on the QR.

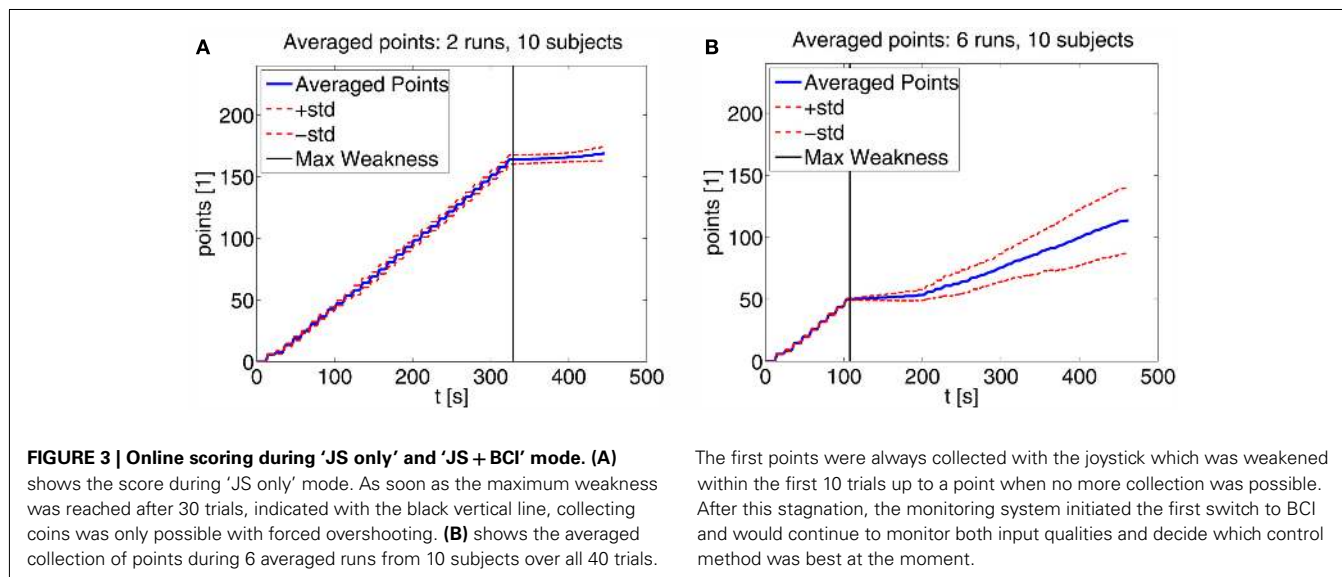
The maximum time in BCI mode (100%) was only reachable if there was no switch back to joystick mode at all. That said, 100% means the whole period of all runs minus the first time of joystick mode which was always initiated at the start of a run.

Another important outcome to evaluate was whether the monitoring system actually made sense for the users. Did switches occur more frequently during bad performance? How good was

Table 2 | Offline classification rates after applying the classifier for online use on all time steps between 1 and 5 s after the cue, accuracies at the best points in time, and test results after using validation sets in an outer cross-validation (oCV) routine.

S	Accuracy [%]			S	Accuracy [%]		
	1–5 s	t _{best}	oCV		1–5 s	t _{best}	oCV
1	80.1	92.4	90.0	6	74.9	84.7	85.2
2	62.2	79.6	79.3	7	79.4	86.7	81.6
3	79.1	87.8	81.2	8	82.8	95.7	95.5
4	70.9	87.8	87.5	9	74.4	84.2	83.3
5	73.9	91.3	83.3	10	56.1	71.4	66.4
	Average				73.4 ± 8	86.2 ± 7	83.3 ± 8

For each subject, one to four features were chosen individually. These were band-powers in frequency bands recorded on the three bipolar channels.



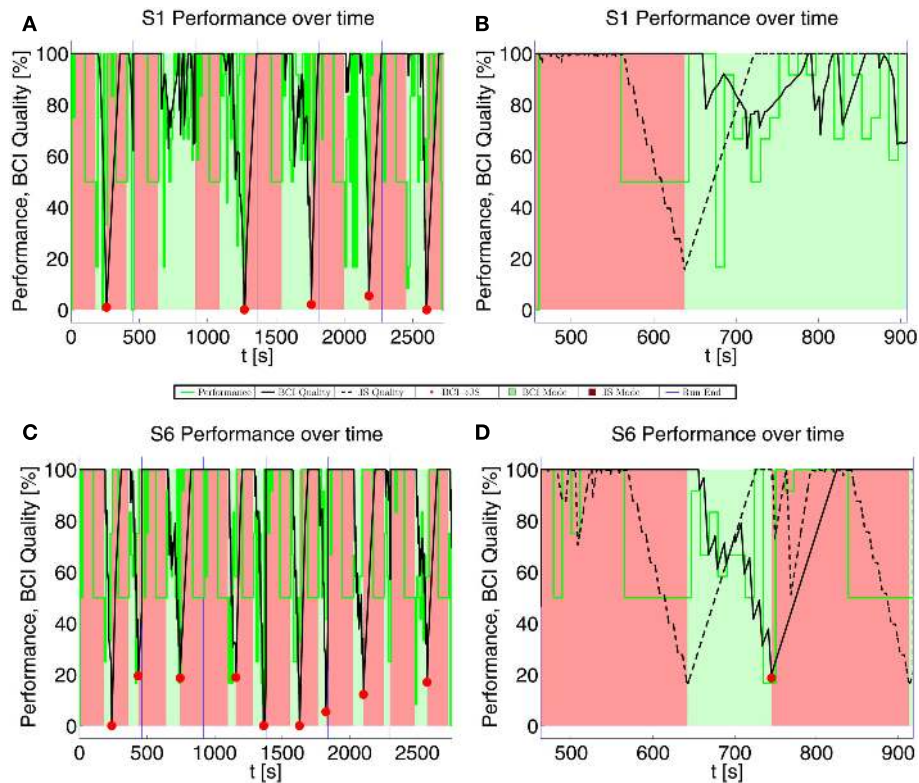


FIGURE 4 | Performance and BCI quality during progress of the whole experiment (A,C) and for one exemplary single run (B,D) for subject S1 who represents a good BCI performer compared to subject S6 as a bad example. The bad performer triggers BCI → JS switches more frequently. Performance depends on currently picked

up coins and barriers, whereas the quality decreases and increases due to active quality measures and recovery rates, respectively. The start of a new quality measures and recovery rates, respectively. The start of a new run is indicated with blue vertical lines. A new run always started in JS mode, no matter in which mode the foregoing run concluded.

Table 3 | Time in BCI mode in relation to BCI performance and object collection rate (pos:neg).

S	pos:neg [%]	BCI perf. [%]	BCI time [%]
1	80:20	77 ± 29	69
2	60:40	59 ± 37	49
3	82:18	78 ± 24	77
4	84:16	81 ± 24	93
5	66:34	65 ± 30	80
6	59:41	57 ± 25	45
7	78:22	74 ± 23	66
8	55:45	54 ± 35	93
9	72:28	69 ± 26	92
10	51:49	50 ± 32	66
Average	69:31 ± 12	66 ± 11	73 ± 17

The performance measures are slightly lower than the collection rates due to the inclusion of missed objects into the calculation. Subjects with a BCI performance above 70% are highlighted in bold black, whereas a performance below 60% is indicated with bold gray.

the performance directly at the time of switching or some time before? These questions are partly answered with the information in **Table 4**. The good BCI performers are highlighted in bold black,

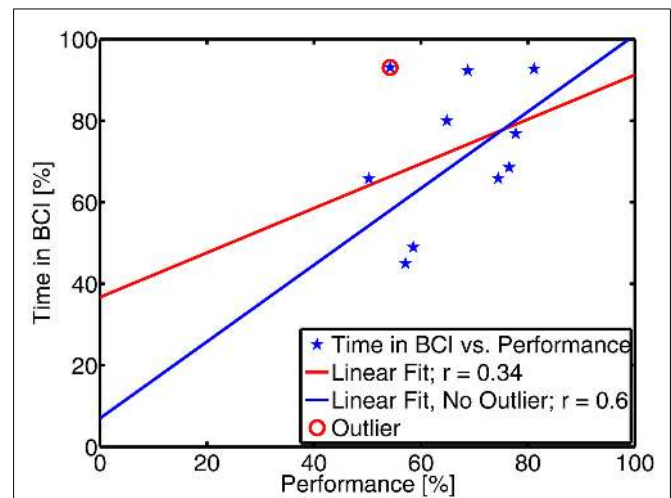


FIGURE 5 | Correlation between time in BCI mode and BCI performance on the basis of all subjects and all the subjects minus one outlier (highlighted star).

whereas subjects with a performance below 60% are grayed out. Apparently, good performance led to a relatively low number of

Table 4 | Number of BCI → JS switches and performances at respective points in time.

S	Switch #	Preceding trials				
		1 [%]	2 [%]	3 [%]	4 [%]	Mean(1–4) [%]
1	5	47	85	68	73	68
2	6	44	67	81	65	64
3	3	64	86	58	94	76
4	4	77	92	77	90	84
5	4	71	75	63	73	70
6	9	52	57	49	66	56
7	4	65	90	81	67	76
8	3	50	67	42	31	47
9	1	83	100	58	83	81
10	6	42	58	69	42	53
Average		59 ± 14	78 ± 14	65 ± 13	68 ± 19	68 ± 12

The table shows how many switches there were for each subject and how good their performance was at that time. The averaged performances, including one to four foregoing trials, are shown over all subjects.

switches (from BCI to JS), as opposed to higher numbers for bad performers with the exception of subject S8. Additionally, the value of the current BCI performance preceding a BCI → JS switch was low: 59% on average for immediately preceding trials and 68%, averaging the performance of the four preceding trials.

The impact of the short term measures on the online car game can be seen in **Table 5**. The number of possible collections was reduced in case of detected short term measures. The resulting missed objects were considered as left out objects for the evaluation methods. On average, the short term measures were active $0.4 \pm 0.5\%$ of the time in BCI mode and $1.7 \pm 0.5\%$ of the time in JS mode. The maximum possible score for BCI mode was reduced by 4.2 ± 5 points and for JS mode 17.6 ± 8 on average.

4. DISCUSSION AND CONCLUSION

We developed a monitoring system which allowed the combination of two different control signal. The system is based on quality measures that monitor signals and generate quality ratings. The evaluation of the experiment included basically two main points. First, we analyzed how well subjects were able to control the car game in general. Second, the functionality of the switching system was evaluated.

As expected after selecting average and good BCI performer, the scores during the car game, especially when in BCI mode, had a large variance. Also, the BCI accuracy during the online car game was worse due to the fact that subjects had to maintain MI and a good LDA classifier output for a longer time, as opposed to the offline runs where we selected classifier based on the best time of separability. Since we were also interested in how the switching would work for a mediocre BCI signal, this was not disappointing. In fact, the outcomes allowed us to better examine the functionality of the switching system. The system was expected to always use the best control strategy at the moment, in terms of quality rating. We hypothesized that the signal with the best quality would also be the one to achieve the best performance.

After evaluating the relationship between time in BCI mode and BCI performance of the subjects, we found positive correlation coefficients. However, only the calculation with the one outlier S8 removed showed a statistical trend with $r = 0.6$ ($p = 0.09$). However, with the low number of samples (nine subjects), statistical significance was not expected. Analyzing all 10 subjects reduced the correlation coefficient from 0.6 to 0.34 with $p = 0.33$. The outlier can be explained by the relatively low weighting of the classifier bias and that other measures were not affecting the quality rating heavily enough to induce earlier switches. The bias could affect the performance more negatively than it was accounted for in the beginning. Also, subject 8 had difficulties maintaining the classifier output for the time needed to collect all the coins. This outcome points out that measures have to be individually adjusted to each patient and to the used application before it can be used in real life situations. However, we only wanted to show a relatively large number of measures, all combined in one setup. This combination should serve as a basis for further experiments where we can use the findings from this switching system and alter the way measures are used and add or remove individual measures and rules for combination.

The functionality of the system can be best observed in **Figure 4**. The most significant detail is the relation between time in JS and BCI mode. JS mode was active longer for subjects with a low BCI performance, because BCI quality dropped faster and switches from BCI → JS were triggered more frequently. As a result there were not only more BCI → JS switches but also switches back from JS → BCI, because the quality of the JS signal did not have enough time to recover. The reason why switches did not occur exactly at the alleged 20% was that switching was only allowed between trials; therefore, the quality often had time to change for the worse or the better for a few seconds before the switching was actually carried out.

The positive effect of the switching capability is demonstrated in **Figure 3**. Increasing the score was possible, even after JS control was no longer working. The weakness was deteriorating faster for the combination of JS and BCI, 10 trials compared to 30 for “JS only.” The period of score stagnation which ranged approximately from 100 to 200 s was purely depending on the choice of weighting for the JS quality measures. Weights of the measures which monitored the small range of motion could be increased to force a faster quality drop in case of weakness and therefore induce an earlier switch to BCI mode.

On top of the individual long term quality measures to determine quality ratings, the short term measures also had a positive effect. These two measures were strictly speaking a byproduct of detected quality measures for BCI EMG noise and JS shaking. When these two measures were detected, the car was forced to the middle of the street, thereby prohibiting possible false but also correct collections, which were in any case not reliable. For BCI there was a higher chance that fewer or no short term measures at all were triggered. On average, the reduction of the maximum possible score and the activated time in BCI mode was lower than in JS mode: 4.2 ± 5 versus 17.6 ± 8 points and 0.4 ± 0.5 versus $1.7 \pm 0.5\%$, respectively. This was based on the fact that subjects had the chance to produce noise-free EEG but could not avoid the artificially induced tremor artifacts in JS mode. For possible

Table 5 | Activation of short term measures and their consequences.

S	BCI time [%]	JS time [%]	BCI max	JS max	−ΔBCI	−ΔJS
1	1.4	2.2	594 → 584	846 → 817	10	29
2	0.0	1.0	420 → 420	1020 → 1013	0	7
3	0.0	1.9	666 → 666	774 → 752	0	22
4	0.0	2.3	810 → 810	630 → 609	0	21
5	0.6	1.9	696 → 690	744 → 726	6	18
6	0.9	1.6	384 → 379	1056 → 1041	5	15
7	0.0	1.7	570 → 570	870 → 850	0	20
8	0.0	0.7	810 → 810	630 → 620	0	10
9	0.7	1.2	798 → 783	642 → 636	15	6
10	0.3	2.0	570 → 564	870 → 842	6	28
Average	0.4 ± 0.5	1.7 ± 0.5			4.2 ± 5	17.6 ± 8

Columns 2 and 3 demonstrate the percentage of time when activated short term measures inhibited the control of the car in both modes. The resulting reduction of possible collections is shown in columns 4–7.

applications with real patients, these short term measures can work as a kind of safety mechanism that can be applied for assistive devices to permit control only with noise-free input signals.

The main concern found in this study was the difficulty to adjust measures, as many parameters have to be adapted to the users' needs in detail. Measures and weights have to be very flexible. Caregivers should be able to add and/or remove measures and to change the weights according to different factors which are very specific. Nevertheless, with some beforehand knowledge, technicians can set up a basic selection of measures and weight ranges to facilitate adjustment for individual usage.

Another problem of the simulation was, in fact, that it was just a simulation. We could only assume how the control would be affected by factors like spasms, tremors, and fatigue. Therefore, the study should also be tested with actual patients who really have to deal with assistive devices that might become unusable over the time of usage as a result of real influences. Here, combining more than one control signal should be really useful for daily activities.

To sum up, the switching approach proved to be promising for future use in experiments with real patients. For these professional users, fatigue and other deteriorating factors concerning assistive

devices are highly anticipated and the possibility to additionally use BCI can improve the functionality significantly. The setup also lends itself to be expanded. First, more signals could be combined instead of just two. For example, sensors could be used to give more information about the current state; EMG signals could serve as an additional control mode. A second possible enhancement would be to combine quality measures with a kind of fusion described in (Leeb et al., 2011). Here, we could let the quality ratings determine how much importance each of the used signals gets when they are fused instead of using a discrete switches. Third, the weighting rules can be improved to permit more feasible solutions for users. Finally, the study should also encourage researchers to find measures based on other factors that can also serve online to detect a bad performance, e.g., the loss of controllability (LoC; Jatzev et al., 2008).

ACKNOWLEDGMENTS

This work is supported by the European ICT Programme Project FP7-224631. This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

REFERENCES

- Allison, B., Brunner, C., Kaiser, V., Müller-Putz, G., Neuper, C., and Pfurtscheller, G. (2010). Toward a hybrid brain-computer interface based on imagined movement and visual attention. *J. Neural Eng.* 7, 026007.
- Anouti, A., and Koller, W. C. (1995). Tremor disorders. Diagnosis and management. *West. J. Med.* 162, 510–513.
- Blattner, M. M., and Glinert, E. P. (1996). Multimodal integration. *Multimed. IEEE* 3, 14–24.
- Breitwieser, C., Daly, I., Neuper, C., and Müller-Putz, G. R. (2011). Proposing a standardized protocol for raw biosignal transmission. *IEEE Trans. Biomed. Eng.* 1. doi: 10.1109/TBME.2011.2174637
- Farwell, L. A., and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clin. Neurophysiol.* 70, 510–523.
- Graimann, B., Huggins, J. E., Levine, S. P., and Pfurtscheller, G. (2002). Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. *Clin. Neurophysiol.* 113, 43–47.
- Jaimes, A., and Sebe, N. (2007). Multimodal human-computer interaction: a survey. *Comput. Vis. Image Underst.* 108, 116–134.
- Jatzev, S., Zander, T. O., DeFlilippis, M., Kothe, C., Welke, S., and Roetting, M. (2008). "Examining causes for non-stationarities: the loss of controllability is a factor which induces non-stationarities," in *Proceedings 4th International BCI Workshop and Training Course* (Graz: Graz University of Technology Publishing House).
- Kawamura, J., Ise, M., and Tagami, M. (1989). The clinical features of spasms in patients with a cervical cord injury. *Paraplegia* 27, 222–226.
- Leeb, R., Sagha, H., Chavarriaga, R., and Millán, J. (2011). A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities. *J. Neural Eng.* 8, 025011.
- Mason, S. G., Bashashati, A., Fatourech, M., Navarro, K. F., and Birch, G. E. (2007). A comprehensive survey of brain interface technology designs. *Ann. Biomed. Eng.* 35, 137–169.
- Middendorf, M., McMillan, G., Calhoun, G., and Jones, K. S. (2000). Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Trans. Rehabil. Eng.* 8, 211–214.
- Millán, J., Rupp, R., Müller-Putz, G., R. Murray-Smith, C. G., Tangermann, M., Vidaurre, C., Cincotti, F., Kübler, A. R., Leeb, C. N., Müller, K., and Mattia, D. (2010). Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges. *Front. Neurosci.* 4:161. doi:10.3389/fnins.2010.00161

- Müller-Putz, G. R., Breitwieser, C., Cincotti, F., Leeb, R., Schreuder, M., Leotta, F., Tavella, M., Bianchi, L., Kreilinger, A., Ramsay, A., Rohm, M., Sagebaum, M., Tonin, L., Neuper, C., and Millán, J. (2011). Tools for brain-computer interaction: a general concept for a hybrid BCI (hBCI). *Front. Neuroinform.* 5:30. doi: 10.3389/fninf.2011.00030
- Müller-Putz, G. R., Scherer, R., Neuper, C., and Pfurtscheller, G. (2006). Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces? *IEEE Trans. Neural Syst. Rehabil. Eng.* 14, 30–37.
- Neuper, C., Müller-Putz, G. R., Scherer, R., and Pfurtscheller, G. (2006). “Motor imagery and EEG-based control of spelling devices and neuroprostheses,” in *Event-Related Dynamics of Brain Oscillations*, eds C. Neuper and W. Klimesch (Amsterdam: Elsevier), 393–409.
- Pfurtscheller, G., Allison, B. Z., Brunner, C., Bauernfeind, G., Solis-Escalante, T., Scherer, R., Zander, T. O., Müller-Putz, G. R., Neuper, C., and Birbaumer, N. (2010). The hybrid BCI. *Front. Neurosci.* 4:30. doi:10.3389/fnpro.2010.00003
- Pfurtscheller, G., and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857.
- Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* 89, 1123–1134.
- Regan, D. (1989). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. Amsterdam: Elsevier.
- Rupp, R., and Gerner, H. J. (2004). Neuroprosthetics of the upper extremity-clinical application in spinal cord injury and future perspectives. *Biomed. Tech. (Berl.)* 49, 93–98.
- Scherer, R., Müller-Putz, G., and Pfurtscheller, G. (2007). Self-initiation of EEG-based brain-computer communication using the heart rate response. *J. Neural Eng.* 4:L23–L29.
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791.
- Zander, T. O., Gaertner, M., Kothe, C., and Vilimak, R. (2011). Combining eye gaze input with a brain-computer interface for touchless human-computer interaction. *Int. J. Hum. Comput. Interact.* 27, 38–51.
- Zander, T. O., and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural Eng.* 8, 025005.
- could be construed as a potential conflict of interest.

Received: 31 May 2011; paper pending published: 18 July 2011; accepted: 19 December 2011; published online: 18 January 2012.

Citation: Kreilinger A, Kaiser V, Breitwieser C, Williamson J, Neuper C and Müller-Putz GR (2012) Switching between manual control and brain-computer interface using long term and short term quality measures. *Front. Neurosci.* 5:147. doi: 10.3389/fnins.2011.00147

This article was submitted to *Frontiers in Neuroprosthetics*, a specialty of *Frontiers in Neuroscience*.

Copyright © 2012 Kreilinger, Kaiser, Breitwieser, Williamson, Neuper and Müller-Putz. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that