Artificial Intelligence in Medicine 59 (2013) 99–110

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Brain Painting: Usability testing according to the user-centered design in end users with severe motor paralysis

Claudia Zickler[a], Sebastian Halder[b], Sonja C. Kleih[b], Cornelia Herbert[b], Andrea Kübler[a,b,*]

[a] Institute of Medical Psychology and Behavioural Neurobiology, University of Tübingen, Silcherstr. 5, D-72076 Tübingen, Germany
[b] Institute of Psychology, University of Würzburg, Marcusstr. 9-11, D-97070 Würzburg, Germany

ARTICLE INFO

Article history:
Received 10 November 2012
Received in revised form 1 August 2013
Accepted 3 August 2013

Keywords:
Brain–computer interface
Entertainment
P300 event-related potential
Assistive technology
Usability testing
User-centered design
Users with disabilities

ABSTRACT

Background: For many years the reestablishment of communication for people with severe motor paralysis has been in the focus of brain–computer interface (BCI) research. Recently applications for entertainment have also been developed. Brain Painting allows the user creative expression through painting pictures.
Objective: The second, revised prototype of the BCI Brain Painting application was evaluated in its target function – free painting – and compared to the P300 spelling application by four end users with severe disabilities.
Methods: According to the International Organization for Standardization (ISO), usability was evaluated in terms of effectiveness (accuracy), efficiency (information transfer rate (ITR)), utility metric, subjective workload (National Aeronautics and Space Administration Task Load Index (NASA TLX)) and user satisfaction (Quebec User Evaluation of Satisfaction with assistive Technology (QUEST) 2.0 and Assistive Technology Device Predisposition Assessment (ATD PA), Device Form).
Results: The results revealed high performance levels ($M \geq 80\%$ accuracy) in the free painting and the copy painting conditions, ITRs (4.47–6.65 bits/min) comparable to other P300 applications and only low to moderate workload levels (5–49 of 100), thereby proving that the complex task of free painting did neither impair performance nor impose insurmountable workload. Users were satisfied with the BCI Brain Painting application. Main obstacles for use in daily life were the system operability and the EEG cap, particularly the need of extensive support for adjustment.
Conclusion: The P300 Brain Painting application can be operated with high effectiveness and efficiency. End users with severe motor paralysis would like to use the application in daily life. User-friendliness, specifically ease of use, is a mandatory necessity when bringing BCI to end users. Early and active involvement of users and iterative user-centered evaluation enable developers to work toward this goal.

1. Introduction

A brain–computer interface (BCI) provides a connection between the brain and a machine that requires no voluntary muscular control. Therefore, BCIs are useful tools for people with severe disabilities due to motor degeneration or brain damage. The re-establishment of communication, environmental control and mobility for this group of patients has been the primary motivation of BCI research and it was shown that it is possible to use BCI for spelling [1–3], browsing the internet [4], controlling lights or DVD player [5], and driving a wheelchair [6]. While these “functional” activities have been in the focus of research, in recent years new applications for entertainment, such as games [7–9] have been developed. A further application is Brain Painting [10,11]. It allows users to actively participate in the leisure activities of painting pictures and thereby expressing their creativity and possibly their emotional state in a non-verbal manner.

However, even though many authors mention motor impaired individuals as target group, less than 10% of the published studies deal with people with severe disabilities [12] clearly showing the translational gap between active research and bringing the application of BCIs to potential end users in their home environment.

* Corresponding author at: Institute of Psychology, University of Würzburg, Marcusstr. 9-11, D-97070 Würzburg, Germany. Tel.: +49 931 31 80179; fax: +49 931 31 87059.
E-mail address: andrea.kuebler@uni-wuerzburg.de (A. Kübler).

0933-3657/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.artmed.2013.08.003

Only recently studies about long-term home use [13–15] and systematic evaluation of BCI end users rendered more the focus of research [1,4,10,16–18]. The aim of the current study was the systematic user-centered evaluation of the Brain Painting prototype according to the standards of the International Organization for Standardization (ISO).

1.1. P300 Brain Painting application

The Brain Painting application was designed by the German artist Adi Hösle in cooperation with the Institute of Medical Psychology and Behavioural Neurobiology at the University of Tübingen [11]. It is based on the P300 BCI.

The P300 is a positive deflection in the electroencephalogram (EEG) that occurs 200–700 ms after stimulus onset. It is evoked when participants attend to a rare stimulus (target) in a random series of stimulus events [3,19]. The first P300 BCI application [20] was a 6 × 6 matrix containing the letters of the alphabet and the numerals from 0 to 9. Rows and columns were flashing randomly and the participant had to focus attention on the letter to be selected while ignoring the others. The desired item constituted the rare stimulus. The P300 elicited by the rare stimulus was detected, the corresponding letter of the matrix was chosen and presented on the computer screen. The P300 BCI is non-invasive, can be parameterized for a new user in a few minutes [21], requires no training and it has been shown repeatedly that people with severe motor disabilities can control it successfully [1,3,22]. Most recently even performance of patients who did not achieve sufficient accuracy for controlling the classic spelling matrix, could be increased to above 80% and up to 100% correct responses by overlaying the matrix with flashing well-known or famous faces [23]. We are well aware that the oddball paradigm realized in the P300 speller can elicit several event-related potentials (ERP) such as the N170, N200 or N400 and thus, the term ERP speller would be more appropriate [24,25]. However, the term P300 speller is often used and in this paper we stick to this terminology.

In the first Brain Painting prototype a colored 6 × 8 matrix was used that contained symbols, numerals or letters to indicate colors, color intensities, object shape, object size, etc. In a pilot-study the artist and a user with amyotrophic lateral sclerosis (ALS) proved the general feasibility of the application [11].

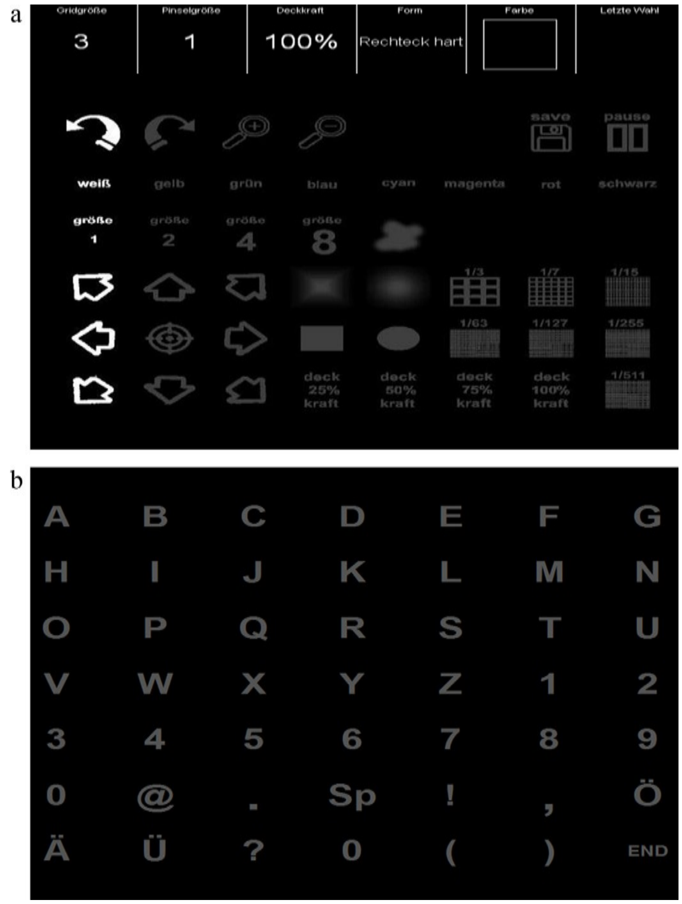

Fig. 1. (a) P300 Brain Painting matrix. For painting an object its properties had to be defined. Only after the selection of "color" the object was transferred to the "canvas". In the toolbox at the top of the screen past selections were shown (from left to right in this figure): grid size (3), object size (1), transparency of color (100%), object shape (rectangle), color (black). In the last square of the toolbox the final selection was shown, which in this example was "black". (b) P300 speller matrix.

1.2. Users' evaluation

1.2.1. Principles and stages of the user-centered design process

The usability of an application depends on how well it suits its purpose and meets target users' needs. User-centered design is an approach for realizing usability and was standardized in the ISO 9241-210 (Ergonomics of human–system interaction – Part 210: Human-centred design for interactive systems) [26]. It defines the following principles (P) for a human-centered design process: (a) understand the user, the task and environmental requirements, (b) encourage early and active involvement of users, (c) be driven and refined by user-centered evaluation, (d) include iteration of design solutions, (e) address the whole user experience, (f) encourage multi-disciplinary design. The user-centered developmental process can be subdivided into four stages (S): 1, understand and specify the context of use; 2, specify the user requirements; 3, produce design solutions to meet user requirements; and 4, evaluate the designs against requirements. These four stages are repeated until a user-adapted product can be released. In the following the principles and stages are referred to as P-a-f and S1–4.

Brain Painting is the only BCI application up to now that allows creative expression and was therefore considered to have an impact on the lives of people with severe disabilities (S1). The artist, a BCI research team and a team of assistive technology (AT) experts [27] closely worked together (P-f) on the development of Brain Painting. Potential end users were involved from the very beginning (P-b). Needs and requirements of potential BCI users were assessed with a questionnaire (P-a, P-b, P-e, S1, S2) [16]. Users indicated "functionality" as the most important aspect, followed by "possibility of independent use" and "ease of use".

The first prototype was then evaluated by ten healthy participants and three participants with ALS [10] (P-c). Effectiveness and ease of use of the prototype was in the focus of this study (S4). The comparison of a copy spelling task [28] with the black and white P300 speller matrix [1] and a copy painting task with the colored Brain Painting matrix revealed in healthy subjects significantly lower accuracies (92% vs. 81%) and P300 amplitudes in the Brain Painting application. Two of the three ALS patients achieved high accuracies in both tasks (89%/100% and 95%/89%). In the group of healthy participants the results of a custom made questionnaire showed a significant difference in clarity and ease of use between the matrices of the P300 spelling and the P300 Brain Painting application. As effectiveness (accuracy) and ease of use of the Brain Painting matrix were not satisfying, the matrix of the first prototype was revised and a new black and white matrix (Fig. 1a) was developed (S3, P-d). Symbols in the new matrix were more self-explanatory, organized in groups (colors, shapes, etc.) and had supplementary descriptions. Tested with healthy subjects (S4), no drop was found in accuracy and P300 amplitudes as compared to the P300 speller (Fig. 1b) [10].

Table 1
Age and disease-related data of end users.

	User A	User B	User C	User D
Age	55	54	45	39
Diagnosis	Amyotrophic lateral sclerosis	Amyotrophic lateral sclerosis	Stroke (pontine infarction due to basilar artery thrombosis)	Duchenne muscular dystrophy
Artificial ventilation	No	Yes, non-invasive some hours per day	No	Yes, non-invasive 24 hours
Artificial nutrition (PEG)	No	No	No	Yes
Wheelchair	Yes	Yes	Yes	Yes
Residual muscular control	Eye movement, mimic movement of the head intact speech	Eye movement, mimic, movement of the head and shoulders restricted speech	Eye movement, mimic movement of head and shoulders residual movement of fingers restricted speech due to restricted respiration	Eye movement, mimic restricted speech due to reduced muscular strength
Living situation	At home	At home	Home for people with disabilities	At home
In daily life supported by	Family (wife, son)	Caregivers	Caregivers	Personal assistants
Computer input device	Chin joystick/switch and screen keyboard	Head tracker	Standard mouse and screen keyboard	Chin joystick and Wergen keyboard
Regular use of computer	Yes	Yes	Yes	Yes
Experience with Brain Painting	Yes	No	No	No
Previous participation in BCI studies	Yes	Yes	No	No

Note: PEG = percutaneous endoscopic gastrostomy.

The goal of the current study was to compile further user requirements that will support and guide development and enhancement of the P300 Brain Painting application (P-d, S-4).

1.2.2. Usability testing

The ISO 9241-201 [26] defines usability as the “extent to which a . . . product . . . can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [26, p. 3]. Effectiveness refers to how accurate and complete the users accomplish the task. Efficiency relates the invested costs, i.e. users’ effort and time, to effectiveness. User satisfaction refers to the perceived comfort and acceptability while using the product. Context of “use refers to users, tasks, equipment (hardware, software and materials) and the physical and social environments in which a product is used” [26, p. 2].

Participants in a user-centered design process should be chosen to match the expected user population as close as possible [29]. Therefore, we included people with severe motor disabilities due to neurodegenerative disease or brain damage who could speak or use AT to give clear and non-ambiguous feedback. Three of the participants (users A, B, and D) were potential users of BCI controlled AT, because they may loose their abilities of voluntary muscular control of AT due to their diseases (ALS or Duchenne muscular dystrophy, see Table 1). User C has been without speech, paralyzed and only able to give small yes and no-signals with vertical eye movement for 2.5 years after a brain stem stroke before recovering and regaining some of his abilities (Table 1). As indicated by Dix and Colleagues [30], all participants should have a comparable amount of experience with computers in general. Thus, a further inclusion criterion was that participants used a personal computer on a regular basis. Furthermore, participants needed to be interested in creative expression, playful experimentation with colors, or “painting”.

One of the most important user requirements is the “possibility of independent use” [16]. Family members and professional caregivers also rated “possibility of independent use” highly important [31]. However, currently, BCI-systems that can be used without the support of others are not available. Therefore, for people with severe motor disabilities the environment in which the product is used, involves explicitly the support that is available in daily life situations. Especially the living situation and support systems of the user may have an impact on his or her needs and satisfaction with the AT product. Therefore, we included users in two different living situations and with three different support systems (see Table 1).

In product evaluation, the chosen tasks should be representative with respect to the functions desired by most users [30]. Thus, the focus was on free painting where the users could paint, playful experiment with colors and express themselves in any way they wanted for several sessions (see Section 2). In addition, the new Brain Painting application was compared to the well studied P300 spelling application.

2. Methods

Four end users evaluated the prototype with the black and white Brain Painting matrix according to the ISO 9241 – in terms of effectiveness, efficiency and user satisfaction. In a copy painting and a copy spelling task the Brain Painting matrix was compared to the P300 spelling matrix. In addition, participants used the brain painting device during five free painting sessions. Effectiveness was assessed for all sessions. Efficiency was compared for the copy spelling, the copy painting and the last free painting session. Satisfaction with the Brain Painting application was evaluated after the final free painting session.

2.1. Participants

Four end users participated in the study (Table 1). Although none of the users was in the locked-in state, in which only few muscles are available for interaction, they were severely disabled and fully dependent on AT input devices for computer access (for further details see Table 1). Three participants (users A, B, and D) lived at home and were supported by either their family, professional caregivers or personal assistants. User C lived in a home for people with disabilities. Participants gave informed consent to the study, which was approved by the Ethical Review Board of the Medical Faculty, University of Tübingen.

2.2. Data acquisition

EEG was measured with a 16-electrode cap (F3, Fz, F4, T7, C3, Cz, C4, T8, Cp3, Cp4, P3, Pz, P4, PO7, PO8, and Oz) based on the 10–20 international electrode system [32]. All channels were referenced to the right mastoid and grounded to the left mastoid. Impedance was maintained below 5 kΩ. The EEG signals were recorded with a 16-channel amplifier (g.USBamp, g.tec, Graz, Austria) and the BCI2000 software [33] on a Hewlett-Packard (HP) 6730b notebook (Intel Core 2 Duo P8600 2.4 GHz, 2 GB RAM, Windows Vista). The EEG signals were band-pass filtered from 0.1 to 30 Hz and sampled at 256 Hz. An additional notch filter (48–52 Hz) was applied to remove the power-line noise.

2.3. Calibration and signal classification

For calibration, users had to select the letters BRAIN, PAINTING and POWER using the copy spelling [28] function of the P300 spelling application [1], where the letter to be written was indicated in the upper left corner of the screen. With the data acquired during calibration, individual parameters for each subject were calculated using the P300-GUI (part of the BCI2000 software package), an application developed in Matlab (The MathWorks™, MA, USA).

In a first step the EEG data was smoothed with a moving average filter of 25 samples and then every 25th sample was used as a feature. Stepwise Linear Discriminant Analysis (SWLDA, the standard method of the BCI2000) was applied for signal classification [34]. The SWLDA adds and removes variables from a multi-linear model following their statistical significance in the discriminant function. It starts with an initial model and then performs a series of forward and backward regressions in discrete steps. During the forward regression the variable, not yet included in the model, showing the lowest p-value below the tolerance threshold (0.10) is added to the model. During the backward regression the variable, already included in the model, showing the highest p-value above the tolerance threshold (0.15) is removed from the model. The backward regression procedure stabilizes the addition of new variables which modify the model, potentially making an already included variable no longer suitable, which is then removed. This selection process was repeated until no further predictor variables were found which met the inclusion/exclusion criteria better than the ones already included in the model or until the maximum number of features (60) was reached [34]. Then the classifier weights were applied to all EEG segments, the outputs summed and then the symbol with the maximum value was selected for feedback.

The calibration was performed in the first session. In the following six sessions (see Section 2.6) users selected the letters BRAIN to verify that the parameters still allowed for an accuracy of at least 80%. In case the accuracy was below 80% a new calibration was performed. For user A the parameter of the first calibration worked in all except the fifth session where a new calibration had to be performed. However, user performance was low (see Section 3) even with the new parameter. The reasons might have been a lack of sleep reported by the user, additional medication we were not aware of or environmental noise.

User B could use the parameter of the first calibration in all except the last free painting session. User C needed a new calibration in the second session. User D used the parameter of the first calibration in sessions one to three. In session four, five and six new calibrations had to be performed.

2.4. Offline analysis of EEG data

EEG-data was analyzed with Brain Vision Analyzer Version 2 (BrainProducts GmbH, Munich, Germany). Figures were created with Matlab R2012a. As the copy spelling and copy painting tasks comprised 20 selections, only the first 20 selections of the last free painting session were included in the analysis. While the items that had to be selected in the copy spelling and painting tasks were predefined, users individually chose items in the free painting session. Intended and "wrong" selections were documented (see Section 2.6). Data segments were extracted from the onset markers of each "flashing" period until 800 ms after the end marker of each period. An interval of 100 ms before the onset marker was used for baseline correction. The EEG was averaged across target and non-target segments separately for each of the above mentioned tasks and participants. Pearson product-moment correlation coefficients (between the EEG samples and target/non-target labels) were calculated from which the determination coefficient was derived (signed r^2). The r^2 is a measure for the amount of variance that can be attributed to the task.

It is well known that the scalp distribution of the P300 is centro-parietal and has its highest amplitudes over midline scalp sites [35]. Therefore, the P300 amplitudes per task and participant were calculated for the Cz-electrode. A P300 was defined as the most pronounced positive peak between 200 and 700 ms post target stimulus.

EEG-data was not corrected for eye-blinks and other artifacts thereby providing a measurement of accuracy under conditions matching the "use in daily life" as close as possible (in which an environment free of artifacts is unlikely).

2.5. Matrices, toolbox and canvas

A 6 × 8 matrix was used for the P300 spelling and the P300 Brain Painting application. For the P300 spelling application, the letters of the German alphabet, the numerals 0–9 and some additional punctuation marks were displayed in white on a black background (Fig. 1b).

For the P300 Brain Painting application, the items of the spelling matrix were replaced by symbols necessary to "paint" a picture. The symbols were arranged in groups (cursor movement, grid size, object shape, object size, color, transparency of color, zoom, undo/redo) and displayed in white on a black background (Fig. 1a). In the painting conditions a separate screen was connected to the BCI, placed next to the HP notebook and served as "canvas". A toolbox placed at the top of the matrix displayed the last 5 selected items (Fig. 1a). For painting, an object, its position, and its properties (size, shape, transparency of color, color) had to be defined. Only after the selection of "color" the object was transferred to the "canvas".

2.6. Design and procedure

Users evaluated the BCI prototype either at home (users A and B) or at the "Beratungsstelle für Unterstützte Kommunikation" (Information Center for Supported Communication), Bad Kreuznach, Germany (users C and D), with caregivers, family members or AT experts present. Daily life distractions, such as family members entering the room or telephone ringing were deliberately not avoided to guarantee for highest possible external validity. During the measurement, users sat in their wheelchair approximately 1 m away from the computer screen on which the spelling or the Brain Painting matrix was displayed.

Users performed a copy spelling, a copy painting and five free painting tasks in seven daily sessions. In the copy spelling session users wrote the sentence "(UWE,27.BÜRO_LINKS!)". This sentence comprises items from each row and column of the matrix. Instead of using the copy function in the BCI2000 application the sentence was displayed on a piece of paper fixed to the top of the screen to allow for equal conditions between copy spelling and copy painting in which a copy function is not available. In the copy painting task, users had to paint a pre-set picture by selecting 20 items from the Brain Painting matrix. Errors in both tasks were not to be corrected. To help subjects maintaining concentration, the selection of the 20 items was split into three runs allowing for short breaks after the 5th and the 13th selection in both copy tasks. In the five free painting sessions users painted up to two hours and took breaks whenever needed. During free painting the participants indicated accidental choice of a not intended item by saying "no" and reported the actually intended item by saying, e.g., "red" or "circle". Even though speech of three users was restricted none of the users had difficulties with the communication of these single words. In addition, the pre- and post-set interval in free painting added up to 16 s (see below), so that speech did not interfere with item selection.

Before each session the matrices (and in the painting session the toolbox and the canvas) and the tasks were explained to the users. They were asked to focus their attention on the item they wanted to select by silently counting the number of times the item was flashing.

For all users and sessions one symbol selection required 30 flashes (15 flashes per row and column = 15 sequences). The duration of each flash was 62.5 ms with an inter-stimulus interval of 125 ms. Before (pre-set) and after (post-set) each block of flashes intervals of 3750 ms and 6250 ms were inserted during calibration and the copy spelling task, of 1000 ms and 10,000 ms for the copy painting task and of 1000 ms and 15,000 ms during free painting. The longer post-set interval in the copy painting was chosen as the user had to search and focus on symbols that – even though they were depicted in groups – followed no external rules like the alphabet in the copy spelling matrix. In the free painting condition the long post-set interval was chosen to allow the user to decide where to place and how to "compose" the object he or she wished to paint. Taken together, the selection of a symbol required 49.38 s during calibration and copy spelling, 50.38 s during copy painting and 55.38 s during free painting.

2.7. Effectiveness

Accuracy was used as an indicator of effectiveness. It was calculated by taking the number of correct selections (letters or symbols) and dividing this value by the total number of selections, thereby deriving a percent value ranging from 0 to 100% accuracy.

2.8. Efficiency

As objective measures of efficiency the information transfer rate (ITR) [36] and the utility metric (for details see below) [37] were calculated. Up to date it is not well understood how BCI users perceive the difficulty of using a BCI application and which aspects of workload contribute the most [17,38,39]. Therefore, after the copy tasks and the last free painting session subjective workload was assessed with the National Aeronautics and Space Administration Task Load Index (NASA TLX) [40].

2.8.1. ITR and utility metric

The accuracy (percent of correctly selected targets), the information content within each selection (number of different targets) and the time needed per selection determine the ITR in bits/min [36] thereby incorporating speed and accuracy in a single value. The ITR was calculated for the copy spelling, copy painting and the last free painting session to measure the speed of command selection according to Wolpaw and colleagues [36]

$$B = \log_2 N + P \log_2 P + (1-P)\log_2 \left[\frac{1-P}{N-1}\right]$$

with N being the number of possible selections in the matrix and P being the accuracy of the user. If an error occurred in the free painting sessions, the wrongly selected item, the intended "undo"-command and the selection of the "correct" item were included in the calculation of the ITR.

The ITR is a widely used measure and allows the comparison with other BCI studies but has shortcomings. Dal Seno and colleagues [37] argue that the ITR does not consider how the P300 matrix spellers function in practical applications. For example, accuracies of 50% or lower mean that every selected item is more likely to be wrong than correct. Using the ITR as a metric can lead to high values even though the BCI would have no practical value. When the user tries to correct an error by selecting the "undo"-function another (wrong) item is selected more likely than not and the expected time to select an item correctly goes to infinite. In this case the interface cannot be used in a meaningful way. They therefore suggest the so-called utility metric [37]

$$U = \frac{b_L}{T_L} = \frac{(2p-1)\log_2(N-1)}{c}$$

with b_L being the average benefit, carried by any correctly selected item, T_L being the expected time required to spell the correctly selected item, p being the accuracy, N being the number of possible selections in the matrix, and c being the duration of a single trial. If $p \leq 0.5$ the transfer rate is 0 bits/min indicating that reasonable communication or use of the BCI application is no longer possible. In this study we provide both the ITR for comparability and the utility metric to underline the practical application of the BCI presented in this study.

2.8.2. Subjective workload

Users' subjective workload was assessed with the NASA TLX [40] after the copy spelling, the copy painting and the last free painting session. It measures (1) the overall workload in the different tasks and (2) the main sources of workload. Workload in the NASA TLX is defined as a "hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance" [40, p. 140]. It is estimated on six dimensions (mental, physical, and temporal demand and performance, effort, and frustration). Subjective workload for each dimension was rated on twenty step scales with scores from 0 to 100. With a weighting procedure the six individual ratings were combined into a global score. To do so, the six scales were combined to 15 pairs. The participants had to indicate which scale of the pair contributed more to their workload. For computing an overall measure of workload (between 0 and 100) and the relative contribution of each subscale (between 0 and 33.3) to overall workload, a weighted average technique was used.

2.9. Satisfaction

User satisfaction refers to the perceived comfort and acceptability while using the product. It was assessed with the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST) 2.0 [41], a visual analogue scale (VAS) "overall device satisfaction" and the Device Form of the Assistive Technology Device Predisposition Assessment (ATD PA) [42] after the final free painting session. In a semi-structured interview users reported their needs and BCI requirements for use in daily life. Speech of three users was restricted and researchers had to practice understanding. Answers were repeated by the researchers to avoid misunderstandings. Only user B used her communication AT (head tracker) during some parts of the interview to give clear and non-ambiguous feedback.

2.9.1. Extended QUEST 2.0

For assessing satisfaction with different aspects of the BCI-controlled AT device Brain Painting, we used the QUEST 2.0 [41]. The questionnaire consists of 12 items such as "adjustment" and

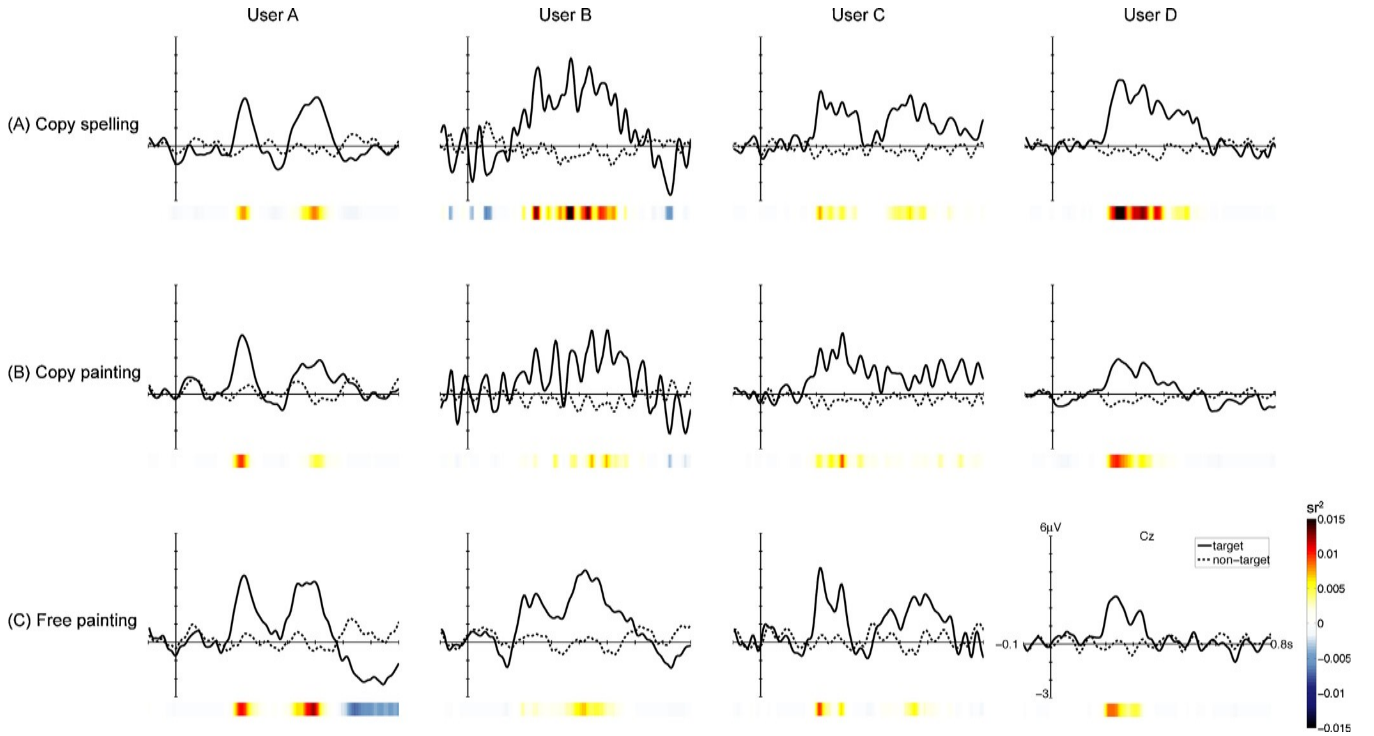

Fig. 2. Average EEG amplitudes (Cz) as a function of time and signed *r*-square values (sr^2, red indicates higher, blue lower sr^2 values; see color bar at the bottom right) for $N=4$ users depicted for copy spelling (A), copy painting (B), and the first 20 selections of the last free painting session (C). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

"service delivery". The QUEST 2.0 is considered invalid if scores for more than six (of 12) satisfaction items are missing. Four items (durability, service delivery, repairs/servicing, follow-up services) were not adequate for the evaluation of a BCI prototype and were thus, removed from the questionnaire (see Table 4). Demers and colleagues [43] explicitly invite researchers to add few items to render the questionnaire more suitable for a specific piece of technology [for an example see 44]. To render the QUEST 2.0 more suitable for evaluation of BCI-controlled AT the four items reliability, speed, learnability, and esthetic design were added. This BCI adapted QUEST 2.0 was referred to as "extended QUEST 2.0." [17].

Items had to be rated on a scale from 1 (not satisfied at all), 2 (not very satisfied), 3 (more or less satisfied), 4 (quite satisfied) to 5 (very satisfied). Users were asked to comment if they were not "very satisfied". The total satisfaction score can be obtained by summing across the QUEST 2.0 items and dividing the sum by the number of items. However, the added items cannot be integrated in the total score of the QUEST 2.0 [41,43]. Therefore, a separate total score was calculated for the four added items. Finally, users were asked to indicate the three most important items.

Table 2
Individual P300 amplitudes and signed *r*-squares of $N=4$ end users.

Task	User A	User B	User C	User D
	P300 amplitudes peak (μV)/latency (ms)			
CS	2.68/506	4.82/372	3.03/220	3.63/243
CP	3.24/235	3.51/501	3.35/298	1.93/239
FPlast	3.67/247	3.95/423	4.09/216	2.65/239
	Signed *r*-square (sr^2)/latency (ms)			
CS	0.008/498	0.018/369	0.007/213	0.016/244
CP	0.010/236	0.007/447	0.009/295	0.010/232
FPlast	0.012/494	0.007/416	0.010/213	0.009/228

Note. CS = copy spelling task, CP = copy painting task, FPlast = first 20 selections of the last free painting session.

2.9.2. ATD PA Device Form

The ATD PA is a set of questionnaires based on the Matching Person and Technology Model of Scherer [42]. For this study the ATD PA Device Form was used. It addresses characteristics of an AT device and asks respondents to rate their predisposition to using the AT under consideration. The questionnaire rates the AT-person match and the expected support in using the device, in other words the expected technology benefit. The 12 items such as "will help me to achieve my goals", "will fit in accustomed routine", had to be rated on a 5-point Likert-scale from 1 (not at all/0% of the time), 2 (sometimes/around 25% of the time), 3 (half the time, neutral/about 50% of the time), 4 (often/around 75% of the time) to 5 (all the time/100% of the time). Users had the option to indicate a "0" if the item was not applicable. By averaging the total of the items a mean score was calculated. The highest possible score was 5.0. Scores between 4.0 and 5.0 indicated a good match of person and AT device. Scores below 4.0 indicated that the match could be improved. If an item is scored 3 or less, there is a risk of device non-use [42].

2.9.3. VAS overall device satisfaction

Overall device satisfaction was assessed with a VAS, ranging from 0 (not at all satisfied) to 10 (absolutely satisfied).

2.9.4. Semi-structured interview

After the last free painting session, users reported in a short interview on their general impression of the prototype, requested improvements, and adjustments needed for use in daily life.

2.10. Statistical analysis

As the group of users was too small for inferential statistical analysis, individual data will be reported descriptively.

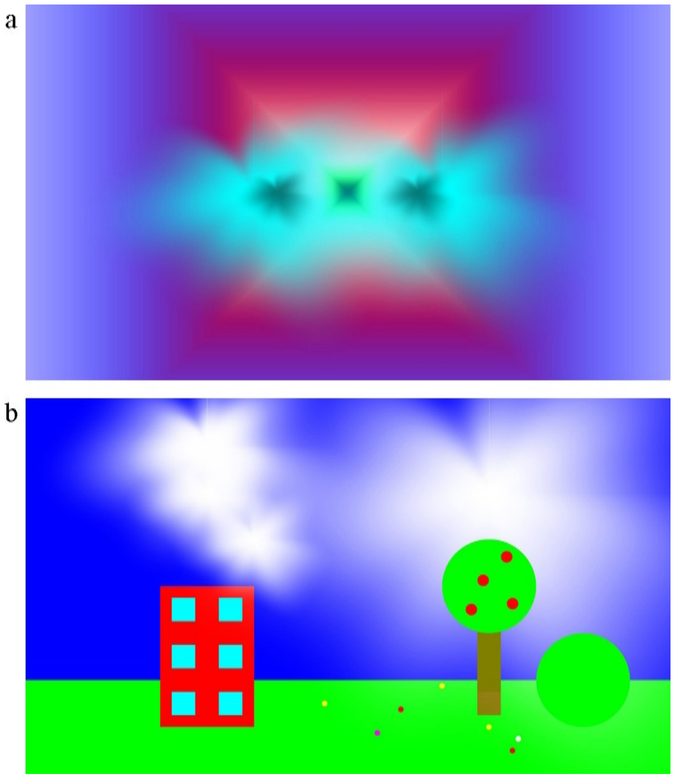

Fig. 3. (a) Painting by user C. (b) Painting by user D. This painting was actually the first which aimed at realistic as opposed to abstract content (an apple tree and skyscraper).

3. Results

All users completed the two copy tasks with 20 selections each and took part in the five free painting sessions that lasted up to two hours.

3.1. P300 amplitudes

Users' individual P300 amplitudes in the copy spelling, the copy painting, and the final free painting session are depicted in Fig. 2. To allow for an adequate comparison, only the first 20 selections of the final free painting session were included. For two users (A and C) lowest amplitudes were found in the copy spelling condition, the task considered least difficult. While for the other two users (B and D) lowest amplitudes were found in the copy painting condition, which was considered less difficult than the free painting task. With regards to ERP amplitudes user A showed a positive peak at approximately 250 and 500 ms in all three conditions. In the copy spelling and the free painting the later ERPs explained more of the variance while in the copy painting condition this was true for the earlier ERPs (for details see Fig. 2 and Table 2). The amplitudes in the copy painting of user B differed in shape and peak from those in the other two conditions (Fig. 2 and Table 2). This might be because on the day of the copy painting session user B suffered from an infected skin and the ground and reference electrodes could not be optimally placed. Likewise, user C presented with two positive peaks in the copy spelling and the free painting condition. User D had ERPs of similar shapes in all three conditions (for details see Fig. 2 and Table 2).

3.2. Effectiveness

For all users accuracy was at least 80% in the copy spelling, the copy painting and the last free painting session (see Table 3). Each user participated in five free painting sessions. Accuracies in all free painting sessions was between 55 and 100%. The average accuracy of the five sessions was 80% and above per user (see Table 3). Users selected between 31 and 129 items during a free painting session (see Table 3) and painted between one and five pictures during the study (user A: 3 pictures, user B: 2, user C: 5, user D: 1). For examples see Fig. 3.

3.3. Efficiency

3.3.1. ITR/utility metric

Highest ITR/utility metric was achieved by user B in the copy painting task with 6.65/6.61 bits/min (100% correct selections). In the free painting condition the pre- and post-set interval of 16 s added to the ITR/utility metric. Highest rates were 5.22/5.18 bits/min (user B with 93% correct selections). Due to high accuracies

Table 3
Individual accuracies, number of selections, information transfer rates, utility metrics, P300 amplitudes, and subjective workload ratings (NASA TXL) of $N = 4$ end users.

Task	User A	User B	User C	User D
Accuracy in % (number of correct selections/number of total selections)				
CS	90 (18/20)	95 (19/20)	80 (16/20)	95 (19/20)
CP	95 (19/20)	100 (20/20)	80 (16/20)	80 (16/20)
FPlast	90 (116/129)	93 (70/75)	88 (112/128)	86 (43/50)
Mean accuracy in % of all free painting sessions (SD; range)				
FPall	80 (15.3; 55–94)	89 (3.2; 86–94)	91 (7.3; 84–100)	80 (9.9; 67–92)
Mean number of selections in all free painting sessions (SD; range)				
FPall	78 (30.0; 54–129)	60 (17.0; 31–74)	92.8 (25.7; 69–128)	63.4 (17.0; 51–89)
Information transfer rate by Wolpaw and utility metric by Dal Seno (bits/min)				
CS	5.56/5.40	6.13/6.07	4.56/4.05	6.13/6.07
CP	5.97/5.95	6.65/6.61	4.47/3.97	4.47/3.97
FPlast	4.93/4.81	5.22/5.18	4.74/4.57	4.56/4.33

NASA TLX

	User A							User B							User C							User D						
	M	Ph	T	P	E	F	**TS**	M	Ph	T	P	E	F	**TS**	M	Ph	T	P	E	F	**TS**	M	Ph	T	P	E	F	**TS**
CS	13	0	1	6	9	3	**32**	2	23	1	3	4	1	**34**	2	0	3	2	13	1	**21**	2	0	3	4	8	1	**18**
CP	13	1	4	3	11	0	**32**	1	4	1	1	1	0	**7**	1	0	1	1	1	0	**5**	1	0	4	13	5	3	**26**
FPlast	20	2	5	6	16	0	**49**	2	17	5	5	4	0	**33**	4	0	17	1	3	1	**23**	2	0	1	8	2	8	**21**

Note. CS = copy spelling task, CP = copy painting task, FPlast = last free painting session, FPall = all five free painting sessions. NASA TLX = National Aeronautics and Space Administration Task Load Index: M = mental demand, Ph = physical demand, T = temporal demand, P = performance, E = effort, F = frustration, **TS** = total score.

Table 4
Satisfaction ratings of $N = 4$ end users in the extended QUEST 2.0 and VAS overall device satisfaction.

Items	User A	User B	User C	User D	User comments
		(1) Extended QUEST 2.0 (ratings from 1 to 5)			
Dimensions	5	3	3	4	"Should be smaller, have less parts", "should work without cables", "cap is eye-catching"
Weight	5	5	4	5	"would be better without cables"
Adjustment	4	4	4	3	Adjustment of cap: "uncomfortable", "time consuming", "laborious", "less electrodes and no cables would be better"; hardware: "a lot of different parts to set up"; software/BCI2000: "still very technical", "not intuitive", "could not manage on my own"
Safety	5	5	5	5	
Comfort	5[a]	4[a]	5[a]	4	Cap/gel: "not comfortable"; cables: "I cannot change sitting position in the wheelchair or move around"; "in the public I would feel a bit uncomfortable as the device is so big and eye-catching"
Ease of use	5[a]	4	4	4	"Long selection time"
Effectiveness	5[a]	4	5	3[a]	"Limited functions compared to other painting programs", "time consuming preparation", "time consuming selection"
Prof. services (information/instructions)	5	5[a]	5	5	
QUEST 2.0 total score	**4.9**	**4.3**	**4.3**	**3.3**	
Reliability	4	4	4	5[a]	"Electrodes need exact adjustment", "you have to concentrate all the time"
Speed	5	4	4[a]	2[a]	"Very slow", "five times faster would be acceptable", "eye tracking systems allow faster selections"
Learnability	5	5[a]	5	5	
Esthetic design	5	4	4[a]	3	"Cap and cables are not aesthetic", "could be more fashionable"
Added items total score	**4.8**	**4.3**	**4.4**	**4.1**	
		(2) VAS overall device satisfaction (0–10)			
	6.7	5.0	7.9	7.0	

Note. (1) QUEST 2.0 = Quebec User Evaluation of Satisfaction with assistive Technology: 1 = not satisfied at all, 5 = very satisfied; M = mean. (2) VAS = visual analogue scale: 0 = not satisfied at all, 10 = extremely satisfied.
[a] Three most important aspects out of the 12 items.

maximum difference between ITR and utility metric was only 0.51 bits/min (see Table 3).

3.3.2. Subjective workload

All users rated their subjective workload in two copy sessions and the last free painting session low to moderate. Workload did not increase with increased task difficulty, i.e. from copy spelling to free painting (Table 3).

In detail, user A rated his overall subjective workload with 32 (of 0–100) in both copy tasks. "Mental demand" ("How mentally demanding was the task?") and "effort" ("How hard did you have to work to accomplish your level of performance?") contributed most to his workload. These aspects of workload increased during free painting so that his overall subjective workload was then 49. User B reported a moderate subjective workload of 34 and 33 in the copy spelling and the free painting sessions. "Physical demand" contributed most to her workload as she suffered from problems with sitting during both sessions. In the copy painting condition she reported a very low workload of 7. User C was confronted with a BCI for the first time. Thus, his workload dropped from 21 in the copy spelling to 5 in the copy painting mode. In the free painting session "temporal demand" ("How hurried or rushed was the pace of the task?") contributed most to his workload rated 23. He reported that counting the flashes all the time during selection of the 128 items was the most challenging part for him. User D rated his subjective workload low to moderate in all three sessions (18, 26, 21). "Effort" contributed most to his workload in the copy spelling session. "Performance" ("How successful were you in accomplishing what you were asked to do?") was rated highest in the copy painting condition. In the free painting condition "performance" and "frustration" ("How insecure, discouraged, irritated, stressed, and annoyed were you?") both contributed most to his workload.

3.4. Satisfaction

3.4.1. Extended QUEST 2.0

According to the total score of the QUEST 2.0 and of the added items, three users (A, B, C) were "very satisfied" (user A: 4.9/4.8) or "rather satisfied" (user B: 4.3/4.3; user C: 4.3/4.4) with the different aspects of the Brain Painting BCI device. User D was "more or less satisfied" (3.3) with the aspects rated in the QUEST 2.0 and "rather satisfied" (4.1) with those in the added items.

In detail, the four users were "very satisfied" (5) with the "safety" of the device, the "professional services (information/instructions)" and the "learnability" (how easy or hard it was to learn to use the BCI application). High satisfaction ratings (4 or 5) were also found for "weight" and "comfort" of the device, "ease of use" and "reliability". Users B and C were only "more or less satisfied" (3) with "dimensions". User D was only "more or less satisfied" (3) with "effectiveness" and "esthetic design" of the device and "not very satisfied" (2) with "speed". None was "very satisfied" (5) with the adjustment of the device (see Table 4).

User comments revealed that reasons for dissatisfaction were mainly the time consuming and uncomfortable adjustment of the cap and the cables, restricted movement due to the cables, the number of parts of the device, the eye-catching and not esthetic design. Comments showed that some aspects of the Brain Painting BCI device such as the gel-based electrode cap and the cables had an influence on the satisfaction rating of several items (dimensions, weight, adjustment, comfort, reliability, and esthetic design).

Speed was judged too slow and was compared to the experience some users had with eye tracking devices. Furthermore, user C judged the program as limited compared to other painting programs (see Table 4) although he did not specify those.

Users were further asked to indicate the three most important items of the extended QUEST 2.0. Comfort was indicated by three users, effectiveness and speed by two users and the following items by one user each: ease of use, professional services (information/instructions), reliability, learnability and esthetic design.

3.4.2. ATD PA Device Form

The mean scores of user A (4.3) and user C (4.2) indicated a good match between these users and the device and revealed that they expected a high benefit from using it. User D (3.8) and user B (3.4) scores were lower indicating that there were areas in which the users were not sufficiently satisfied and perceived a mismatch between themselves and the device. All users presented with high ratings (5 or 4) for the following items: "This device will help me to achieve my goals", "I have the capabilities and stamina to use this device without discomfort, stress and fatigue" and "I will feel comfortable using this device around family".

Lower ratings (3 or 2) of two or more users were found for "this device will fit well with my accustomed routine" (users B, C, D), "the supports, assistance and accommodations exist for successful use of this device" (users B, C), "I will feel more secure (safe, sure of myself) when using this device" (users B, C), and "I will feel comfortable using this device around the community" (users B, D) (Table 5).

3.4.3. VAS overall device satisfaction

While user B indicated that her overall satisfaction with the Brain Painting device was moderate (5.0) only, the other users rated satisfaction between 6.7 and 7.9 (scale 0–10).

Table 5
Satisfaction ratings of $N=4$ end users in the ATD PA Device Form.

ATD PA Device Form				
Items	User A	User B	User C	User D
This device will help me to achieve my goals.	4	4	5	4
This device will benefit me and improve my quality of life.	4	4	4	3
I am confident I know how to use this device and its various features.	5	3	5	5
I will feel more secure (safe, sure of myself) when using this device	4	3	2	0
This device will fit well with my accustomed routine.	4	2	3	3
I have the capabilities and stamina to use this device without discomfort, stress and fatigue.	4	4	5	5
The supports, assistance and accommodations exist for successful use of this device.	5	2	2	4
This device will physically fit in all desired environments (car, living room, etc.).	4	2	4	4
I will feel comfortable using this device around family.	5	5	5	5
I will feel comfortable using this device around friends.	4	5	5	4
I will feel comfortable using this device at work.	0	0	5	3
I will feel comfortable using this device around the community.	0	3	5	2
Mean score	**4.3**	**3.4**	**4.2**	**3.8**

Note. ATD PA = Assistive Technology Device Predisposition Assessment, 5 = all the time/100% of the time, 4 = often/around 75% of the time, 3 = half the time, neutral/about 50% of the time, 2 = sometimes/around 25% of the time, 1 = not at all/0% of the time, 0 = not applicable.

3.4.4. Semi-structured interview

All users reported that they enjoyed using Brain Painting. They appreciated that there was a program that allowed people with severe disabilities creative expression through painting. Three users (A, C, and D) asked for improvement of the matrix (more colors, shapes, additional grid sizes). Users C and D reported that they would have preferred less sequences for selection to reduce workload or increase speed, thereby taking in account the risk of lower accuracies. With regard to use in daily life users asked for less electrodes, and no cables. Users A and B stated that they would suffer from skin problems if they would use the gel more than once a week. Adjustment and set-up time were judged as too long and complicated by all users. User B who lived at home on her own and was supported by professional caregivers (paid by the social security system) and user C who lived in a home for people with disabilities clearly indicated that they would not have sufficient support to use the BCI system in daily life. Furthermore, users B and D asked for an integration of Brain Painting in communication AT devices, user C in his own computer system. User D suggested using BCI as an additional input channel.

3.5. Overall user requirements for further improvement of Brain Painting

Taken together evaluation of the Brain Painting prototype revealed the following user requirements:

- improvement of the matrix, specifically more colors, shapes and grid sizes to make the application more comparable to conventional painting programs,
- less electrodes and gel to enhance comfort and reliability and most importantly to optimize adjustment procedures, thereby reducing the burden of users and caregivers and allowing users in different living and support situations the use of the BCI application,
- a simple set-up of the software to reduce burden of caregivers,
- no cables to enhance comfort and allow movement (changing of sitting position, moving the wheelchair) during use of the BCI application,
- improvement of the appearance of the EEG cap to make the device less eye-catching and encourage users to use it outside the context of family and friends,
- improvement of speed to make the BCI more attractive in comparison to other AT solutions, e.g., eye trackers,
- the possibility to integrate Brain Painting into existing communication systems, so that there is no need for two different computer based AT systems,
- the possibility to use BCI as an additional input channel in existing communication systems.

4. Discussion

The current study reports on a systematic user-centered evaluation according to the standards of the ISO by four users with severe disabilities. Users with three different diseases were included, living in two different environments with three different caregiving situations which fosters generalizability of our results. The Brain Painting prototype was evaluated in its target function free painting, and in addition compared to the well studied P300 speller, using a defined set of objective measures, standardized questionnaires and an interview.

The outcome of the usability testing by end users provided a proof of feasibility of the free painting function of the second BCI Brain Painting prototype and a description of user requirements to promote further development of the prototype and BCI in general. Evaluation yielded high effectiveness, and efficiency was comparable to that in other BCI studies with severely ill end users [1,4,17]. Users enjoyed painting and painted up to one picture per session. Three users would have liked to use Brain Painting in daily life once or twice a week. They reported high satisfaction with the learnability, ease of use, and reliability of the device. For daily use, users recommended less electrodes/gel, no cables, an improved appearance of the cap, a simple set-up procedure for the software, the possibility to integrate the Brain Painting application into their existing communication systems and the possibility to use BCI as an additional input channel.

4.1. Effectiveness

End users had high accuracies that were at least 80% in the copy spelling and the copy painting task proving that the spelling matrix and the black and white matrix of the second Brain Painting prototype allowed for comparable performance. Even more important, users achieved high effectiveness also in the free painting condition, where the challenge was not only to select the intended item but to “compose” the intended object from size, shape, color, etc. and, of course, to merge objects to a painting. Average performance in the five free painting sessions was 80% and above for all users and users selected up to 129 items per session. Novice BCI users (2) and novice Brain Painting users (3) presented with high accuracies already in the first free painting session thereby underlining the good learnability of the P300 Brain Painting application. This result was supported by high satisfaction ratings for learnability and ease of use. During the five free painting sessions users painted 11 pictures, thereby proving that it is possible for users with severe disabilities to use Brain Painting successfully on a regular basis.

4.2. Efficiency

With 4.47–6.65 bits/min ITRs in the copy tasks were comparable to other P300 studies with end users [1,4,17]. ITRs in the last free painting session were maximum 5.22 bits/min due to the long post-set interval of 15 s per selection. Due to high accuracies the differences between ITRs and utility metric were negligible. Users agreed to the long post-set interval that allowed them to decide where to place and how to “compose” the intended object.

However, two users would have preferred less than 15 sequences per selection. The reason for user C was that the counting of the flashes contributed most to his subjective workload, while for user D the improvement of speed was crucial. To ensure comparability we had a fixed number of sequences independent of the individual user. In real life situations the optimal ratio of sequences and accuracy would be adapted individually, thereby improving efficiency and user satisfaction.

The NASA TLX revealed that subjective workload was low to moderate (5–32 of 100) in the copy tasks and moderate (21–49) during free painting. This is very encouraging as the additional demand of thinking about what to paint, choosing the different symbols before an item can be "transferred" to the "canvas" and having the entire picture in mind, did not impose insurmountable workload. One goal of the study had been to understand which aspects contributed most to the subjective workload. For three users the same aspects contributed most to their workload in at least two of the three conditions (user A: mental demand and effort, user B: physical demand, user D: performance). In users C and D "effort" decreased from copy spelling (13 and 8 of 33) to free painting (3 and 2) which may be interpreted as a learning effect. In user B it stayed the same (4) and in user A it increased to 16 which is in the moderate range. Low "effort" ratings were also found in the evaluation study of a P300-controlled commercial AT software for text entry (average "effort" rating for spelling: 7, for emailing: 5, and for internet surfing: 4) [17]. Thus, "effort" to operate a P300-based BCI appears to be lower than for sensorimotor rhythms (SMR)-controlled BCI application. For example, Felton and colleagues [39] who evaluated a SMR BCI with two tasks (cursor movement in horizontal direction and cursor movement in vertical direction) reported "effort" ratings of on average 15 and 18. "Effort" and "mental demand" contributed most to the subjective workload of these disabled end users. However, ratings of users who evaluated a SMR BCI driven gaming application (Holz et al., 2013) were comparable to those achieved in our study. The NASA TLX also revealed that not only BCI operation contributed to subjective workload but also the lack of physical well-being as in the case of user B, who suffered from problems with sitting and spasms in her legs in the copy spelling and last free painting session.

P300 waveforms varied between users and for two users between conditions. However, variability in P300 amplitudes did not markedly affect performance. This result is comparable to other BCI studies and specifically true for motor impaired end users [1,3]. Lowest P300 amplitudes were found in the copy spelling and the copy painting condition, even though free painting was considered the most difficult task. One could expect that composing a picture added to the workload which could have resulted in lower amplitudes [45,46]. The long post-set interval of 15 s in the free painting condition might have contributed to this result as this break allowed the users to first decide what to do next and then concentrate on the desired item, thereby reducing workload which is confirmed by the NASA TLX results. Importantly, none of the users wanted this interval to be reduced even though two users asked for improvement of speed.

4.3. Satisfaction

Generally, users were rather satisfied with the second Brain Painting prototype. They indicated that Brain Painting helped them to achieve their goals, contributed to their quality of life, and that they had the capabilities and stamina to use it without stress. Users suggested minor changes of the matrix. Main reasons for dissatisfaction, risk of non-match of person and device, and risk of non-use were the dimensions of the device (number of electrodes and cables), the extensive adjustment procedure and need of support, and the eye-catching design that might prevent users from using the device outside the context of family and friends. These reasons for dissatisfaction were also found in the evaluation study of a P300-controlled commercial AT software [17] and in the focus group study of Blain-Moraes and colleagues [18] about user acceptance of BCIs. In the focus group study five out of eight social units (the end user and his/her caregiver (all of them family members)) were not satisfied with the gel-based EEG cap. The prospect of using the BCI in daily live was a source of tension in these social units such that the BCI would on the one hand allow communication and thereby contribute to their quality of life but on the other hand add to the burden of the caregivers. Our results indicated that the adjustment of hard- and software would specifically exclude users who depend exclusively on the support of professional caregivers either at home or in a home for people with disabilities. Therefore, simple procedures and "ease of use" are essential needs of any caregiver [31], such that BCI does not add to their burden [47]. Unlike in P300 communication devices [17] speed was a source of dissatisfaction only for user D, which indicates that users are more tolerant toward sub-optimal efficiency and efficacy if the AT device provides additional options apart from mandatory communication.

4.4. Limitations

Generalization of our results is limited because only four end users evaluated the prototype. Much effort is necessary to find and conduct "long-term" experiments with severely disabled individuals, especially on experimental interfaces. Due to the small sample size, data were reported descriptively only. However, the assessment of several satisfaction measures, including individual comments of users in the extended QUEST 2.0 and a semi-structured interview allowed an in-depth analysis of individual experience.

All included end users had experience with computers and AT for communication. This inclusion criterion was chosen to avoid a confounding of BCI evaluation with aspects of experience with computers. Our participants rated their workload low to moderate and indicated that they had the capabilities and stamina to use the Brain Painting without stress. It cannot be ruled out that potential users who are uncomfortable with technology might be concerned about their ability using a BCI device [18]. Another limitation of the current study was that caregivers were not included as observers and that BCI set-up was performed by BCI experts. Future studies will have to overcome these limitations and are currently implemented in single case studies [15]. Finally results are limited as the prototype was not evaluated in a long-term home use condition. A study with long-term independent home use evaluation of the Brain Painting application is currently conducted [15] and will provide valuable information to bridge the translational and reliability gap in BCI research [12].

4.5. Outlook: further development of the P300 Brain Painting application

Following the user-centered design process we further adapt the Brain Painting application according to the user requirements (S3) and continuously evaluate these against the requirements (S4). To better match user and technology, an extended matrix with more options (color, shapes, merging of objects, integrating external items) is currently developed. To foster independent use, the optimized computer interface for autocalibration developed by Kaufmann and colleagues will be implemented [21]. To increase effectiveness and efficiency the flashing face paradigm [23] has recently been integrated in the long-term independent home use system for Brain Painting [15]. To reduce the time for adjustment, only 8 EEG electrodes are used in the long-term study [15]. Dry electrodes and wireless signal transmission will further improve ease of use [48–50]. Finally, Brain Painting will be integrated in existing AT systems and hybrid BCIs [51].

5. Conclusion

The P300 BCI Brain Painting application was evaluated in its genuine target function – free painting – by potential end users with severe disabilities. Usability testing of the second prototype with the black and white matrix revealed high effectiveness (accuracy) and efficiency (ITR, subjective workload) in both the copy and the free painting conditions. This is very encouraging as the additional demand of thinking about what to paint, neither choosing the different symbols before an item can be "transferred" to the "canvas" nor having the entire picture in mind, did impose insurmountable workload nor impair performance. Furthermore, results revealed that users were rather satisfied with the device and very interested in the BCI application that allowed them creative expression and entertainment and users would have liked to use it in daily life up to twice a week. The user-centered design process was successfully applied to improve the Brain Painting application resulting in a second prototype tested within this study. The theoretical framework of the user-centered design process allows for systematic evaluation of usability in its aspects effectiveness, efficiency and satisfaction in a specified context of use and provides researchers with valuable information about user requirements. Such insights are of utmost importance for bringing BCI-controlled applications to end users.

Conflict of interest

None.

Acknowledgements

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

References

[1] Nijboer F, Sellers EW, Mellinger J, Jordan MA, Matuz T, Furdea A, et al. A P300-based brain–computer interface for people with amyotrophic lateral sclerosis. Clin Neurophysiol 2008;119(8):1909–16.
[2] Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, et al. A spelling device for the paralysed. Nature 1999;398(6725):297–8.
[3] Sellers EW, Donchin E. A P300-based brain–computer interface: initial tests by ALS patients. Clin Neurophysiol 2006;117(3):538–48.
[4] Mugler EM, Ruf CA, Halder S, Bensch M, Kübler A. Design and implementation of a P300-based brain–computer interface for controlling an internet browser. IEEE Trans Neural Syst Rehabil Eng 2010;18(6):599–609.
[5] Aloise F, Ferriero D, Ruiu A, Santucci G, Catarci T, Mattia D, et al. Controlling domotic appliances via a dynamical P300-based brain–computer interface. In: Emiliani PL, Burzagli L, Como A, Gabbanini F, Salminen AL, editors. Assistive Technology from Adapted Equipment to Inclusive Environments-AAATE 2009. Assistive Technology Research Series 25. Amsterdam: IOS Press; 2009. p. 200–3.
[6] Millan JdR, Falan F, Vanhooydonck D, Lew E, Phillips J, Nuttin M. Asynchronous non-invasive brain-actuated control of an intelligent wheelchair. In: Proc 31st A Int Conf IEEE Eng Med Biol Soc 2009. 2009. p. 3361–4, http://dx.doi.org/10.1109/IEMBS.2009.5332828.
[7] Tangermann M, Krauledat M, Grzeska K, Sagebaum M, Blankertz B, Vidaurre C, et al. Playing pinball with non-invasive BCI. In: Advances in Neural Information Processing Systems 21 (NIPS). Cambridge: MIT Press; 2009. p. 1641–8.
[8] Finke A, Lenhardt A, Ritter H. The MindGame: a P300-based brain–computer interface game. Neural Networks 2009;22:1329–33.
[9] Ramsey L, Tangermann M, Haufe S, Blankertz S. Practicing fast-decision BCI using a goalkeeper-paradigm. BMC Neurosci 2009;10(Suppl. 1):69.
[10] Münßinger JI, Halder S, Kleih SC, Furdera A, Raco V, Hösle A, et al. Brain Painting: first evaluation of a new brain–computer interface application with ALS-patients and healthy volunteers. Front Neurosci 2010:182, http://dx.doi.org/10.3389/fnins.2010.00182.
[11] Kübler A, Halder S, Furdea A, Hösle A. Brain Painting – BCI meets art. In: Müller-Putz GR, Brunner C, Leeb R, Pfurtscheller G, Neuper C, editors. Proc 4th Int brain–computer interface workshop and training course. Graz: Verlag der Technischen Universität Graz; 2008. p. 361–6 http://bci.tugraz.at/images/Proceedings_BCI_Conferences/Proceeding_BCI_Conference_2008.pdf [accessed 15 April 2013].
[12] Kübler A, Holz EM, Zickler C, Kaufmann T. A user centred approach for bringing BCI controlled applications to end-users. In: Fazel-Rezai R, editor. Brain–computer interface systems – recent progress and future prospects. InTech; published June 05, 2013 under CC BY 3.0 license; 2013., http://dx.doi.org/10.5772/50696 http://www.intechopen.com/articles/show/title/a-user-centred-approach-for-bringing-bci-controlled-applications-to-end-users [accessed 10 July 2013].
[13] Vaughan TM, McFarland DJ, Schalk G, Sarnacki WA, Krusienski DJ, Sellers EW, et al. The Wadsworth BCI Research and Development Program: at home with BCI. IEEE Trans Neural Syst Rehabil Eng 2006;14(2):229–33.
[14] Sellers EW, Vaughan TM, Wolpaw JR. A brain computer interface for long-term independent use. Amyotrop Lateral Scler 2010;11(5):449–55.
[15] Holz EM, Botrel L, Kübler A. Bridging gaps long-term independent BCI home-use by a locked-in end-user. In: Millán NL, Leeb R, Millán JdR, Guechoul N, Carlson T, Sobolewski A, et al., editors. Proc TOBI workshop IV. Practical brain computer interfaces for end-users: progress and challenges. 2013. p. 35–6 http://www.tobi-project.org/sites/default/files/public/Workshop/Proceedings TOBIworkshopIV.pdf [accessed 15 April 2013].
[16] Zickler C, Di Donna V, Kaiser V, Al-Khodairy A, Kleih S, Kuebler A, et al. BCI applications for people with disabilities: defining user needs and user requirements. In: Emiliani PL, Burzagli L, Como A, Gabbanini F, Salminen AL, editors. Assistive Technology from Adapted Equipment to Inclusive Environments-AAATE 2009. Assistive Technology Research Series 25. Amsterdam: IOS Press; 2009. p. 185–9.
[17] Zickler C, Riccio A, Leotta F, Hillian-Tress S, Halder S, Holz E, et al. A brain–computer interface as input channel for a standard assistive technology software. Clin EEG Neurosci 2011;42(4):222–30.
[18] Blain-Moraes S, Schaff R, Gruis KL, Huggins JE, Wren PA. Barriers to and mediators of brain–computer interface user acceptance: focus group findings. Ergonomics 2010;55(5):516–25.
[19] Polich J. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 2007;118(10):2128–48.
[20] Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol 1988;70(6):510–23.
[21] Kaufmann T, Völker S, Gunesch L, Kübler A. Spelling is just a click away – a user-centered brain–computer interface including auto-calibration and predictive text entry. Front Neurosci 2012;6(May):72, http://dx.doi.org/10.3389/fnins.2012.00072 http://www.readcube.com/articles/10.3389/fnins.2012.00072 [accessed 15 April 2013].
[22] Piccione F, Giorgi F, Tonin P, Priftis K, Giove S, Silvoni S, et al. P300-based brain computer interface: reliability and performance in healthy and paralysed participants. Clin Neurophysiol 2006;117(3):531–7.
[23] Kaufmann T, Schulz SM, Köblitz A, Renner G, Wessig C, Kübler A. Face stimuli effectively prevent brain–computer interface inefficiency in patients with neurodegenerative disease. Clin Neurophysiol 2013;124(5):893–900.
[24] Allison BZ, Pineda JA. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. IEEE Trans Neural Syst Rehabil Eng 2003;11(2):110–3.
[25] Kaufmann T, Hammer EM, Kübler A. ERPs contributing to classification in the P300 BCI. In: Müller-Putz GR, Scherer R, Billinger M, Kreilinger A, Kaiser V, Neuper C, editors. Proc 5th int brain–computer interface conf. Graz: Verlag der Technischen Universität Graz; 2011. p. 136–9 http://bci.tugraz.at/images/Proceedings_BCI_Conferences/Proceeding_BCI_Conference_2011.pdf [accessed 15 April 2013].
[26] ISO 9241-210. Ergonomics of human–system interaction – Part 210: human-centred design for interactive systems. Genf: ISO; 2010.
[27] Hoogerwerf EJ, Mongardi S, Staiger-Saelzer P, Zickler C. BCI research and user involvement: the role of independent AT centres in the TOBI project. In: Assistive technology – technology transfer. Proc of the AAATE workshop. 2010. p. 32–5 http://kt-equal.org.uk/uploads/aaateoct/aaatepreceedings.pdf [accessed 15 April 2013].
[28] Kübler A, Kotchoubey B, Kaiser K, Wolpaw JR, Birbaumer N. Brain–computer communication: unlocking the locked in. Psychol Bull 2001;127(3):358–75.
[29] Maguire MC. User-centred requirements handbook. WP5 D5.3 of the Telematics Applications Project TE-RESPECT: requirements engineering and specification in Telematics; 1998 http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/UserCenteredRequirementsHandbook.pdf [accessed 15 April 2013].
[30] Dix A, Finlay J, Abowd GD, Beale R. Human–computer interaction. 3rd ed. Harlow, England: Pearson Education Limited; 2004.
[31] Zickler C, Kaiser V, Al-Khodairy A, Kleih SC, Kübler A, Mattia D, et al. BCI-applications: needs and requirements of disabled end-users and professional users. Poster presented at the 4th int BCI meeting; 2010. http://www.tobi-project.org/sites/default/files/public/Publications/asilomar_2010_zickler.pdf [accessed 15 April 2013].

[32] Sharbrough F, Chatrian GE, Lesser RP, Lüders H, Nuwer M, Picton TW. American Electroencephalographic Society guidelines for standard electrode position nomenclature. J Clin Neurophysiol 1991;8:200–2.
[33] Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: a general-purpose brain–computer interface (BCI) system. IEEE Trans Biomed Eng 2004;51(6):1034–43.
[34] Krusienski DJ, Sellers EW, McFarland DJ, Vaughan TM, Wolpaw JR. Toward enhanced P300 speller performance. J Neurosci Methods 2008;167(1):15–21.
[35] Sutton S, Braren M, Zubin J, John ER. Evoked-potential correlates of stimulus uncertainty. Science 1965;150(700):1187–8.
[36] Wolpaw JR, Ramoser H, McFarland DJ, Pfurtscheller G. EEG-based communication: improved accuracy by response verification. IEEE Trans Rehabil Eng 1998;6(3):326–33.
[37] Dal Seno B, Matteucci M, Mainardi LT. The utility metric: a novel method to assess the overall performance of discrete brain–computer interfaces. IEEE Trans Neural Syst Rehabil Eng 2010;18(1):20–8.
[38] Riccio A, Leotta F, Bianchi L, Aloise F, Zickler C, Hoogerwerf EJ, et al. Workload measurement in a communication application operated through a P300-based BCI. J Neural Eng 2011;8(2):025028.
[39] Felton EA, Williams JC, Vanderheiden GC, Radwin RG. Mental workload during brain–computer interface training. Ergonomics 2012;55(5):526–37.
[40] Hart SG, Staveland LE. Development of NASA TLX (Task Load Index): results of experimental and theoretical research. In: Hancock PA, Meshkati N, editors. Human mental workload. Amsterdam: Elsevier (North-Holland); 1988. p. 139–83.
[41] Demers L, Weiss-Lambrou R, Ska B. Quebec User Evaluation of Satisfaction with assistive Technology QUEST version 2.0. An outcome measure for assistive technology devices. Webster, New York: Institute for Matching Person and Technology; 2000.
[42] Scherer MJ. The Matching Person & Technology (MPT) model. Manual and assessments. Webster, New York: Institute for Matching Person and Technology; 2007 [CD-ROM].
[43] Demers L, Weiss-Lambrou R, Ska B. Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0): an overview and recent progress. Technol Disabil 2002;14:101–5.
[44] Vincent C, Deaudelin I, Hotton M. Pilot on evaluating social participation following the use of an assistive technology designed to facilitate face-to-face communication between deaf and hearing persons. Technol Disabil 2007;19(4):153–67.
[45] Isreal JG, Chesney GL, Wickens CD, Donchin E. P300 and tracking difficulty: evidence for multiple resources in dual-task performance. Psychophysiology 1980;17(3):259–73.
[46] Isreal JG, Wickens CD, Donchin E. The dynamics of P300 during dual-task performance. Prog Brain Res 1980;54:416–21.
[47] Rabkin JG, Albert SM, Rowland LP, Mitsumoto H. How common is depression among ALS caregivers? A longitudinal study. Amyotroph Lateral Scler 2009;10(5/6):445–55.
[48] Gargiulo G, Calvo RA, Bifulco P, Cesarelli M, Jin C, Mohamed A, et al. A new EEG recording system for passive dry electrodes. Clin Neurophysiol 2010;121:686–93.
[49] Volosyak I, Valbuena D, Malechka T, Peuscher J, Gräser A. Brain–computer interface using water-based electrodes. J Neural Eng 2010;7(6):066007.
[50] Zander TO, Lehne M, Ihmes K, Jatzex IS, Correia JH, Kothe C, et al. A dry EEG-system for scientific research and brain–computer interface. Front Neurosci 2011;5:53, http://dx.doi.org/10.3389/fnins.2011.00053.
[51] Allison BZ, Brunner C, Kaiser V, Müller-Putz GR, Neuper C, Pfurtscheller G. Toward a hybrid brain–computer interface based on imagined movement and visual attention. J Neural Eng 2010;7(2):026007.