Real-Time Closed-Loop Control of Cognitive Load in Neurological Patients During Robot-Assisted Gait Training

Alexander Koenig, *Student Member, IEEE*, Domen Novak, Ximena Omlin, Michael Pulfer, Eric Perreault, *Member, IEEE*, Lukas Zimmerli, Matjaz Mihelj, *Member, IEEE*, and Robert Riener, *Member, IEEE*

***Abstract*—Cognitively challenging training sessions during robot-assisted gait training after stroke were shown to be key requirements for the success of rehabilitation. Despite a broad variability of cognitive impairments amongst the stroke population, current rehabilitation environments do not adapt to the cognitive capabilities of the patient, as cognitive load cannot be objectively assessed in real-time. We provided healthy subjects and stroke patients with a virtual task during robot-assisted gait training, which allowed modulating cognitive load by adapting the difficulty level of the task. We quantified the cognitive load of stroke patients by using psychophysiological measurements and performance data. In open-loop experiments with healthy subjects and stroke patients, we obtained training data for a linear, adaptive classifier that estimated the current cognitive load of patients in real-time. We verified our classification results via questionnaires and obtained 88% correct classification in healthy subjects and 75% in patients. Using the pre-trained, adaptive classifier, we closed the cognitive control loop around healthy subjects and stroke patients by automatically adapting the difficulty level of the virtual task in real-time such that patients were neither cognitively overloaded nor under-challenged.**

***Index Terms*—Bio cooperative control, cognitive control, Lokomat, psychophysiology, stroke rehabilitation.**

I. Introduction

ROBOT-assisted gait rehabilitation is becoming increasingly common in patients after stroke, spinal cord injury, traumatic brain injury or cerebral palsy. Cognitively challenging training sessions were shown to be key requirements for the success of motor learning in general and in rehabilitation [1]–[3]. In addition, research in healthy subjects suggests that motor learning decreases in the presence of a distracting cognitive task, which presents a cognitively over-challenging situation [4], [5]. Quantifying and controlling cognitive load in neuro-rehabilitation to avoid tasks that are cognitively too demanding or too easy, has the potential to increase motor learning and thereby the training efficiency and therapeutic outcome of neurological rehabilitation [1], [3]. In the context of robot-assisted gait training, we define cognitive load as the amount of attention and focus the patient has to dedicate towards the task in order to successfully fulfill this task.

Despite existing tools used to modulate patient motivation such as virtual environments [6], the rehabilitation environment does not yet adapt to the cognitive load of the patient. One major reason is that the current cognitive load of patients cannot be objectively assessed. Questionnaires can be used to obtain subjective information, but only at discrete time-points after training has ceased. They can therefore not be used in real time.

Psychophysiological measurements can provide real-time information on the cognitive load of subjects [7], [8], as physiological processes were shown to reflect behavioral-, cognitive-, emotional-, and social interaction [9]. Heart rate variability (HRV) was shown to decrease with cognitive load [10] and negative emotions [11]. Skin conductance has previously been used as a measure for arousal [12], [13] and was found to increase during demanding tasks compared to a rest period [14]. Breathing frequency was found to increase during cognitive effort [15], negative emotions [16] and also during physical activity [17]. However, not all physiological signals that provide information on cognitive load are unambiguous. Heart rate (HR) was found to increase due to stress or negative emotions [8], [18], but decreased in reaction to unpleasant stimuli [19]–[21]. Skin temperature decreased with increased cognitive load [22], but increased in response to positive emotions [23].

Psychophysiological measurements can be used to perform bio-cooperative control by putting the human in a psychological closed control loop [24]. Previously, mental engagement of neurological patients has been automatically quantified during robot assisted rehabilitation using psychophysiological signals [25], [26]. Real-time stress level estimation from analysis of HRV has previously been performed in healthy subjects [27]. However, these approaches were neither adaptive, nor was cognitive load controlled to a desired level.

Manuscript received November 08, 2010; revised March 24, 2011; accepted June 10, 2011. Date of current version August 10, 2011. This work was supported by the EU Project MIMICS funded by the European Community's Seventh Framework Program (FP7/2007-2013) under Grant 215756.

A. Koenig, X. Omlin, and R. Riener are with the Sensory-Motor Systems Lab, Department of Mechanical Engineering and Process Engineering, ETH Zurich, 8092 Zurich, Switzerland and also with the Spinal Cord Injury Center, Medical Faculty, University Hospital Balgrist, University of Zurich, 8008 Zurich, Switzerland (e-mail: koenig@mavt.ethz.ch; riener@mavt.ethz.ch).

M. Pulfer is with the Sensory-Motor Systems Lab, Department of Mechanical Engineering and Process Engineering, ETH Zurich, 8092 Zurich, Switzerland (e-mail: pulferm@ee.ethz.ch).

D. Novak and M. Mihelj are with the Faculty of Electrical Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia (e-mail: domenn@robo.fe.uni-lj.si; matjaz.mihelj@robo.fe.uni-lj.si).

E. Perreault is with the Department of Biomedical Engineering, Northwestern University, Chicago, IL 60611 USA (e-mail: e-perreault@northwestern.edu).

L. Zimmerli is with the Sensory-Motor Systems Lab, Department of Mechanical and Process Engineering, ETH Zurich, 8092 Zurich, Switzerland, and also with the 5Hocoma AG, 8604 Volketswil, Switzerland (e-mail: lukas.zimmerli@mavt.ethz.ch).

Digital Object Identifier 10.1109/TNSRE.2011.2160460

The first objective of this paper was to evaluate if psychophysiological signals would allow bio-cooperative control of cognitive load during robot assisted gait therapy in the presence of physical effort induced by walking. We induced different levels of cognitive load and used physiological measurements in combination with real-time machine learning techniques to objectively quantify the current cognitive load and control it to a desired level.

The second objective was to determine if performance metrics could be used as a proxy, instead of psychophysiological signals. While performance metrics might be less accurate, they are more practical to obtain in a clinical setting and might replace psychophysiological signals. We, therefore, compared a closed loop controller using physiological signals with a simple task performance controller that only controlled task success without the use of physiological signals.

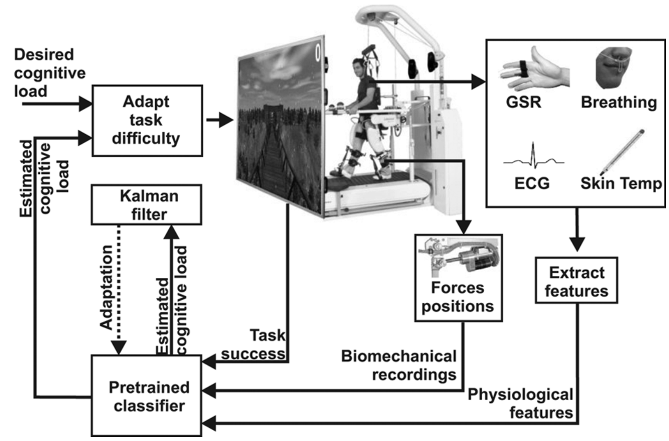


Fig. 1. Overview system setup. The features from physiological recordings of the subject, biomechanical recordings from the robot and performance information from the virtual environment were used to classify the current cognitive load of subjects during robot assisted treadmill training. A classifier, pre-trained on data obtained in open loop experiments, was set up with initial classification parameters. A Kalman filter adapted the weights of the classifier at run time to take patient specific responses in physiology into account.

II. Methods

We provided subjects with a virtual task, which was used to modulate cognitive load of subjects during robot-assisted gait training. Details on the possibility of using virtual environments to modulate cognitive load during robot-assisted gait training can be found in [25]. In brief, the task was either too easy such that cognitive load was very low, too difficulty such that cognitive load was very high, or adjusted such that subjects were cognitively capable of fulfilling the task if they concentrated on it. Psychophysiological recordings were used to objectively quantify the changes in cognitive load. We set up a linear classifier and enhanced it with a Kalman filter. This classifier was trained with data from open loop experiments in healthy subjects and stroke patients. After the initial training of the classifier, we performed bio-cooperative closed loop control of cognitive load in both healthy subjects and patients. In the following paragraphs, we first describe the training environment and the physiological signals that were used for automatic classification of cognitive load. Then we explain the open loop experiments which were performed for classifier training and the closed loop experiments used for bio-cooperative closed loop control.

A. Hardware

The experimental setup consisted of three parts: a commercially available driven gait orthosis (DGO) commonly used in gait rehabilitation, the virtual reality display system, and the measurement system for physiological signals (Fig. 1). As DGO, the Lokomat (Hocoma Inc., Volketswil, Switzerland) was used for the locomotion training. Drives on hip and knee joints provide torques to the subject and assist the locomotion on a treadmill by guiding the subject's legs along a predefined trajectory. The display system consisted of a 3×2 m back-projection screen in front of the gait robot and a 5.1 surround sound system. All physiological signals were recorded and amplified with the g.USBamp (Guger Technologies, Graz, Austria).

B. Input Data for Classification of Cognitive Load

We recorded physiological signals from the subject, force data from the DGO and task success data from the virtual environment. We extracted features from the physiological data as described below, took the mean and standard deviation over 30 s and fused the data into one feature vector. All signal processing software was written in Matlab 2008b (The Mathworks, Natick, MA).

Physiological signals recorded from subjects during Lokomat walking were HR, breathing frequency, skin conductance and skin temperature. In previous work, these four signals were established to carry the most information while being recordable with reasonable effort for the clinical personnel [25], [28]. The electrocardiogram was measured with three surface electrodes. One electrode was affixed 2 cm below the right clavicula between the first and the second rib, one was affixed at the fifth intercostal space on the mid axillary line on the left side of the body, and a ground electrode was affixed to the right acromion. HR was computed from electrocardiogram using a real time R wave detection algorithm [29]. HRV in the time domain was computed according to the recommendations of Malik [30] as the square root of the mean squared differences of successive normal-to-normal intervals (RMSSD). The frequency analysis of HRV was performed using the ratio of low-frequency components and high-frequency components (LF/HF). Using a thermistor flow sensor placed underneath the nose, we recorded the breathing of subjects and computed breathing frequency and its derivative using a peak detection algorithm. Changes in skin conductance were measured with the g.GSR sensor from Guger Technologies (Graz, Austria) using two electrodes attached on the proximal phalanx of the second and the fourth fingers on the left hand or the unaffected hand in stroke patients. Skin conductance responses (SCR) were detected from the skin conductance level (SCL) when signal amplitude increased by at least $0.05\ \mu$S in less than 5 s [14]. SCL was bandpass filtered with a 20–50 Hz Butterworth filter to remove sensory artifacts and sensory noise. Skin temperature was measured using the Guger g.Temp sensor on the distal phalanx of the fifth finger of the left hand or the unaffected hand in stroke patients. Signals were sampled at 512 Hz according to the recommendations of Malik *et al.* [30].

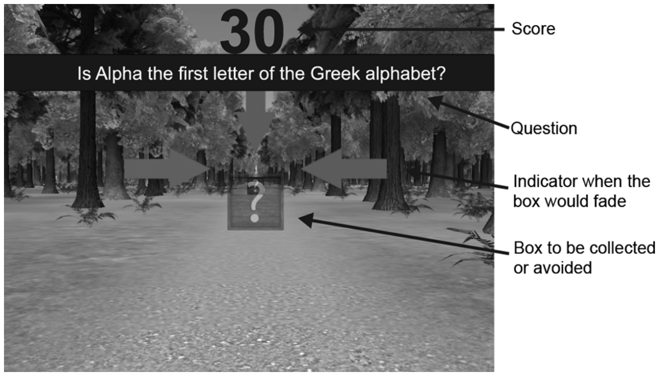


Fig. 2. Virtual scenario. A cognitive task (question) had to be answered via biomechanical effort. If the question was correctly posed (alpha is the first letter of the Greek alphabet), then the subject had to accelerate in the virtual world and had to collect the box in front of him/her, before it disappeared. Otherwise, the subject had to decelerate in the virtual world and wait, until the box had disappeared. The time until the box disappears was coded by arrows that pointed at the next object, which made the task easier to understand.

Force data from the DGO was weighted and summed for each step such that it reflected the current physical effort of the subjects [31]. From the virtual environment, we obtained the success rate of percent correctly avoided and collected objects and percent correctly answered questions.

C. Virtual Task

A virtual reality task with adjustable difficulty level was used to modulate cognitive load during training sessions. The walking speed in the scenario was controlled via subject's voluntary effort in the DGO. As the DGO was position controlled, the subject could produce voluntary forces, either pushing into the movement direction of the orthosis or resisting the gait movement of the orthosis. An increase in effort yielded to an increase in virtual walking speed; a decrease in effort resulted in a decrease in virtual walking speed. While the subject could influence the virtual walking speed in the scenario, the real walking speed in the DGO was kept constant.

In the virtual task, subjects had to collect and avoid objects which were placed on a straight line and disappeared slowly in front of them. By modulation of their physical effort in the DGO, the subject could collect objects by increasing effort and avoid objects by decreasing effort. In addition to this biomechanical task, subjects had to answer questions during the task, which were displayed in a box on the screen. If the statement was correct (e.g., $1 + 1 = 2$), subjects had to collect the box before it disappeared. If the statement was false (e.g., $1+1 = 3$), subjects had to avoid it by decreasing the walking speed until the box disappeared (Fig. 2). Subjects obtained immediate feedback on their performance via their score (Fig. 2, top), which increased or decreased by five points if an object or a question was answered correctly or incorrectly.

The task difficulty could be increased by increasing the question difficulty, by decreasing the time available to read and answer the question, by decreasing the distance between objects, and increasing the time until the objects disappeared. Conversely, the difficulty could be decreased by posing easier questions, allowing more time to read and react to the question, by increasing the distance between objects and decreasing the time until the objects disappeared. As subjects showed individual differences in how fast they read question, decelerated or accelerated in the virtual environment or decided how to answer a question, each of these variables needed to be adjusted for each subject individually.

There were over 200 questions in total, divided into nine categories: science, mathematics, history, geography, sports, art, nature, general, and music. They were presented from all categories evenly, and the same question was never given twice to the same subject. Since the questions all had yes/no answers, they were set up so that, at lower difficulties, the questions would not be a strong distractor (i.e., the answer was very obvious) while at higher difficulties both answers were probable and required the subject to think carefully. The difficulty of the questions was rated independently by two psychologists on a scale from 1 to 10, and a third rater was consulted when the opinions of the first two raters differed by more than two. While a perfect scaling of difficulty levels cannot be guaranteed due to inter-individual differences in knowledge, a definite trend thus exists.

D. Modulating Cognitive Load

We induced three distinct levels of cognitive load by adjusting the task difficulty. In condition one, subjects were bored and under-challenged; condition two provided a cognitive challenge which was difficult, but feasible; condition three over-challenged and overstressed subjects with an unfeasibly difficult task. Task difficulty was set individually: in the under-challenging condition, the task was adjusted such that subjects succeeded in over 90% of cases. The questions were very simple, the objects were placed far away and disappeared slowly such that subjects had a long time to think about the answer. In the challenging condition, question difficulty and the required reaction time were adjusted so that the success rate was between 40%–80%. In the over-challenging condition, subjects had very little time to answer very difficult questions with a success rate of maximally 20%. The task difficulty was set individually to take inter-subject differences into account: subjects with a high general level of knowledge or a lower level of impairment, for instance, would be able to answer more difficult questions. The selection of task success thresholds was established previously in our laboratory [25].

E. Questionnaires

We asked the subjects, how difficult they perceived the task in terms of physical effort as well as cognitive load on a five point scale. 1 was not physically exhausting and 5 was extremely physical exhausting. Similarly, 1 on the cognitive difficulty scale represented no cognitive challenge while 5 represented an extreme cognitive challenge. All questions were posed nonverbally as a pictorial questionnaire, as not to disturb the breathing frequency analysis by speaking and also to reduce the complexity of responding to the questionnaire for aphasic stroke patients or patients with cognitive impairments.

F. Classifier Training

1) Experimental Protocol of Training Dataset Recordings: Developing an algorithm that estimated only cognitive load re-

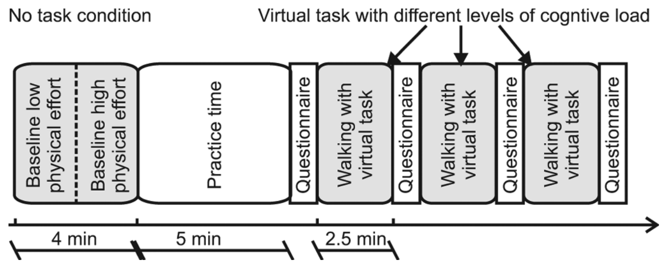


Fig. 3. Study protocol open loop experiments. The virtual task is presented three times, each time with a different task difficulty to induce three different levels of cognitive load. The order of the conditions is randomized.

quired us to verify that cognitive load and physical effort were dissociated. This was necessary, as the virtual task was controlled by modulation of physical effort. We, therefore, collect data in which cognitive load (task difficulty) and physical effort (required energy to walk and to control the virtual task) were not co-varied in our main protocol.

Before the beginning of the recording session, subjects were familiarized with the questionnaires. Each recording session started with a 4-min baseline period, in which physical effort was varied, but no cognitive task was present ("no task" condition, Fig. 3). During this initial period, subjects completed two walking behaviors: passive, such that the robot provided most of the physical effort and active, overemphasizing the gait pattern and expending additional energy.

The initial 4-min period of only physical effort was followed by 5 min of exercise time, during which subjects could get acquainted with the addition of the virtual task. Meanwhile, the experimenter determined the levels of cognitive load by adjusting the distance between objects and the question difficulty level such that the task success for each condition was reached as described above.

After the baseline measurement, three different cognitive load conditions were presented in randomized order, each 2.5 min long (cognitive task with randomized task difficulty, Fig. 3). The three different levels of cognitive load were induced by adjusting the difficulty of the task at the beginning and, as necessary, during the condition such that the subjects could reach a desired task success. Difficulty was modulated by question difficulty, distance between objects and the time before the next object would disappear.

After baseline and after each condition, subjects were asked to answer questionnaires on cognitive load, in order to verify if we really under-challenged, challenged and over-challenged the subjects cognitively. In addition, we asked the subjects, how difficult they perceived the task in terms of physical effort.

2) Correlation Between Physical Effort and Cognitive Load: To verify that cognitive load and physical effort were dissociated, we tested questionnaire results from physical effort and cognitive difficulty level for significant differences between the two baseline conditions (Fig. 3, No task condition). We also tested, if the change in physical effort was significantly higher than the change in cognitive load for the two baseline conditions. Both tests were performed using the Friedman test followed by a Wilcoxon test for paired comparison. To get further information, if cognitive load and perceived physical effort were dissociated, we computed the coefficient of determination R2 between cognitive load and perceived physical effort.

3) Classifier Setup: We investigated a classic linear discriminant analysis (LDA) [32] classifier as well as a Kalman adaptive version of the LDA. Both classifiers were trained to classify cognitive load from the recorded physiological variables and performed classification once every 30 s.

The classic LDA tries to separate two classes by mapping the input vector $\vec{x}$ to a one dimensional output y using the weights $\vec{w}$ after $y = \vec{x}\vec{w}\vec{w}$ is computed by maximizing the cost function

$$J(\vec{w}) = \frac{\vec{w}^T S_B \vec{w}}{\vec{w}^T S_W \vec{w}} \tag{1}$$

where S_B and S_W are the between-class covariance matrices and the within-class covariance matricies of the two classes that are to be separated. With four classes to be distinguished (no task, under-challenged, challenged, over-challenged), we trained four two-class LDA classifiers $J_i, i \in [1, 2, 3, 4]$. The class was then identified as $\max(J_i)$.

All data recorded in the "no task" condition, regardless of the level of physical effort, was labeled as baseline to the classifier. This ensured that the classifier estimated only cognitive load and not physical effort.

Kalman adaptive linear discriminant analysis (KALDA) is an adaptive version of the classic LDA classifier where the weights $\vec{w}$ are updated recursively using a Kalman filter when new data become available [33]. Every 30 s, when a new input vector $\vec{x}$ and its corresponding known output y become known, the weights $\vec{w}$ are updated according to the following equations:

$$H = [1, \vec{x}] \tag{2}$$
$$e = y - H \cdot \vec{w}_{k-1} \tag{3}$$
$$v = 1 - \mathrm{UPC} \tag{4}$$
$$Q = H \cdot A_{k-1} \cdot H^T + v \tag{5}$$
$$g_k = \frac{A_{k-1} \cdot H^T}{Q} \tag{6}$$
$$\vec{w}_k = \vec{w}_{k-1} + g_k \cdot e \tag{7}$$
$$\tilde{A}_k = A_{k-1} - g_k \cdot H \cdot A_{k-1} \tag{8}$$
$$A_k = \frac{\mathrm{trace}(\tilde{A}_k) \cdot \mathrm{UPC}}{p} + \tilde{A}_k \tag{9}$$

where $\vec{w}_{k-1}$ are the old weights, $\vec{w}$ are the updated weights, e is the one-step prediction error, Q is the estimated prediction variance, A_{k-1} is the old *a priori* state error correlation matrix, A_k is the new *a priori* state error correlation matrix, $\tilde{A}_k$ is an intermediate value needed to compute A_k, v is the variance of the innovation process, g_k is the Kalman gain, UPC is the update coefficient, and p is the number of elements of $\vec{w}$. The starting values of A_0 and $\vec{w}_0$ as well as the optimal value of UPC are computed from the training data set, with possible values of UPC limited to the interval of $[01)$.

Originally designed for analysis of electroencephalographic data [33], KALDA has already been used for two-class classification of physiological measurements and motor activity measurements in upper extremity rehabilitation [34]. The original KALDA was designed for two classes, but it can be expanded to four classes by running the update process for all four two-class classifiers in parallel. When an updated took place, we updated the classifier corresponding to the correct class.

An important problem with the original implementation of KALDA is that it is supervised; as can be seen from (3), the correct output class (y) is required to update the weights. Since this information is generally not available in practice, one alternative is to modify KALDA so that, once the current class is estimated as $\max(J_i)$, $\max(J_i)$ is passed to (3) in place of y. Thus, in essence, KALDA updates the classifiers using its own estimate of the output class rather than the correct output class. However, such an approach could also amplify errors and lead to instability if unchecked. If an incorrect estimate is used to update the discriminant function, the discriminant function will become worse.

We addressed this potential instability by only performing the update process if the classifier has a high probability of being correct. As previously mentioned, the estimated class is defined using all four classifiers as $\max(J_i),\ i \in [1 \ldots 4]$. If $\max(J_i)$ is much higher than other elements of J_i, it is more likely to be correct. If, on the other hand, all elements of J_i have roughly the same value, the estimated class should be considered unreliable. Our implementation of unsupervised KALDA was thus as follows:

- compute $\max(J_i)$ as per (1);
- if ($\max(J_i)$—second-largest value of J_i) $> T_R$, proceed with (2)–(9);
- otherwise proceed to the next time period without updating.

T_R was considered to be a "reliability threshold"—how probable the classification result needs to be. The optimal value of T_R was calculated from the training data set using a sensitivity analysis. Performing "leave one out" classification, we computed the classification results for $[0 <= T_R <= 1]$ for each subject. We averaged the results for each value of T_R over all subjects and selected T_R for the best average classification result.

We acknowledge that this unsupervised method does not have a strong theoretical basis and could become unstable with inappropriate values of *UPC* and T_R. It has already been tested with similar input variables in upper extremity rehabilitation and found to be more accurate than classic LDA [34], but this is only an empirical confirmation. Thus, our plan was to test both classic LDA and adaptive LDA with open-loop data. The one that would prove more accurate would be then used in the closed-loop phase.

4) Performance Evaluation: The quality of the classifier training was quantified by computing percent correctly classified between the estimated and the actual cognitive load. The actual cognitive load was labeled as cognitively under-challenging, challenging but feasible and over-challenging, according to the condition the subject was in, as explained in the section "Modulating cognitive load." We investigated how well the classifier could generalize across subjects by training the classifier on all but the ith subject and performing classification on the ith subject, commonly called "leave one out" classification. This was done separately for the data of healthy subjects and for the pooled data of healthy subjects and patients.

To investigate, if physiological signals alone would suffice to classify cognitive load, we trained the classifier with five different input vectors (conditions 1–5).

- C1: Physiological signals alone.
- C2: Physiological signals with task success data from the virtual environment.
- C3: Physiological signals with force data from the robot.
- C4: A joint input vector of physiological signals, task success, and force data.
- C5: Only task success from the virtual environment.

We compared the classification results for the five input vectors for healthy subjects and for patients. We then checked, if the KALDA would significantly improved classification compared to the classic LDA. All statistical tests were performed using the Friedman test. Afterwards, a post-hoc Wilcoxon test for paired comparison with Bonferroni correction was performed. Due to the paired comparison of five conditions, the significance level was set to 0.01.

G. Closed Loop Control of Cognitive Load

1) Adaptation of Virtual Environment: The goal of the closed loop experiment was to reach a challenging, but feasible task difficulty for each subject, independent of the subject's abilities and the initial settings of the virtual task. We had intentionally set up a four class classifier that could distinguish between three classes of cognitive load and a baseline. The virtual environment however only allowed making the task easier or harder. We, therefore, had to reduce three classes of cognitive load to the binary decision easier or harder.

If cognitive load was classified as under-challenging, task difficulty was increased with a large adaptation step in the virtual environment. If cognitive load was classified as over-challenging, task difficulty was decreased with a large adaptation step. If cognitive load reached a state in which it was classified as challenging but feasible, the classifier evaluated if the task was by trend too easy or too hard and then also performed an update of the task difficulty, but with smaller adaptation steps (Fig. 4). This allowed fast convergence to a state in which the subject was cognitively challenged and prevented oscillatory behavior of the task difficulty. Theoretically, the classifier could also detect baseline. While this situation never occurred, the adaptation rules stated that no change would be undertaken in this case.

2) Experimental Protocol: Subjects started to walk in the Lokomat and were given 5 min to exercise the task. The assistive force of the Lokomat was set to 100%, which corresponds to position control of the gait trajectory. Testing the controller based on physiological signals, we started the training session in an extreme condition (either too easy or too difficult), pseudo-randomized for each subject. Every 60 s, the classifier provided a real-time estimation of the current cognitive load, based on the last 30 s of data, and updated the virtual task difficulty. We allowed 10 update steps of the virtual environment, which resulted in 10 min Lokomat walking.

This protocol was run in two randomized conditions to evaluate the necessity of psychophysiological recordings for automatic classification of cognitive load: once, the experiment was

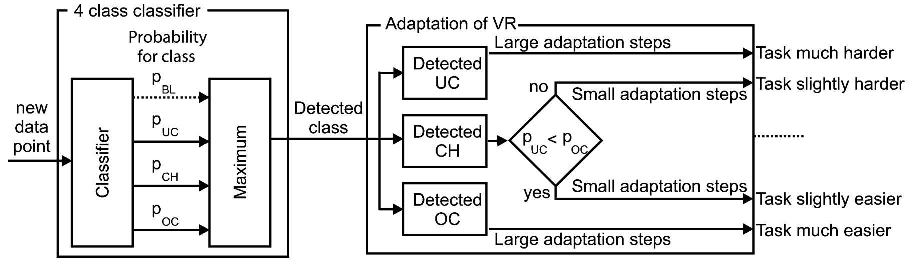


Fig. 4. Adaptation of virtual environment based on the result of the classifier. The classifier determines the probability p for each of the four classes (BL: baseline, UC: under-challenged, CH: challenged, OC: over-challenged) and determines the current cognitive load from the largest probability. In the extreme cases (UC , OC), large adaptation steps are done when adapting the virtual task. Adaptation also happens if the classifier detects the class CH (p_{CH} larger than all other probabilities). By comparing the p_{UC} versus p_{OC}, the algorithm determines if the subject tends towards under-challenged or over-challenged and performs small adaptation steps.

performed with the KALDA classifier; input to the system was the full feature vector as described above. Once, only task success without additional physiological recordings was controlled to a desired level. The controller tried to set the task level difficulty such that success rate reached 70%, which was defined as challenging, but not too difficult.

3) Performance Evaluation of Closed Loop Experiments: While in the open loop experiments, the subjects had time to answer the questions after each condition, in the closed loop experiments we did not want to interrupt the immersion and focus on the game. We, therefore, decided against asking subjects the full set of questionnaires on cognitive load and physical effort during the closed loop experiments. Assuming that we were able to really modulate cognitive load as suggested by the results of the open loop experiments, we only tested whether or not the subjects agreed with the adaptation step of the virtual environment. The correctness of the classifier's decision was verified by asking subjects at each update step if they would want the task to be easier or harder. This meant that evaluation of the closed loop experiments was only done with two decisions (easier/harder), compared to four classes in the open loop experiments (Fig. 4).

While the subject's answer to the questions was not taken into account for the classifiers decision, it allowed comparison between the subjects' opinion and the adaptation steps taken in the virtual environment. Also, for comparison, the experimenter rated the performance of subjects and noted, if the task should be easier or harder from the therapeutic point of view. To avoid a bias, the experimenter rated the performance before asking the subject. Also, the experimenter could not see the classifier's decision.

Comparing the percent match between subject-classifier and experimenter-classifier was of particular interest to quantify how patients perceived and rated their own performance. While we excluded patients with cognitive impairments in this study, patients might not be able to rate their own performance subjectively compared to objective expert-rating of the therapist.

Using a Friedman test with post-hoc Wilcoxon test, we compared the closed loop system with physiological data, robot data and score information to the system that only controlled score. Significance level was set to 0.05, as no Bonferroni correction was necessary.

H. Subject Data

Both open and closed loop experiments were performed with naïve subjects that had never seen the virtual environment before. Open loop experiments were performed in nine healthy subjects (5 female, 4 male, 29 years ± 5) and four stroke subjects (Table I, top). Closed loop experiments were performed in five healthy subjects (1 female, 4 male, 32 years ± 12) and five stroke patients (Table I, bottom). Subjects were excluded if cognitive impairments prevented them from reading and understanding the questions on the screen. Subjects were asked to refrain from coffee, tea and cigarette consumption four hours prior to the recording. Upon arrival, the task and the questionnaires were explained to all subjects. All subjects gave informed consent. Subjects were fixed into the DGO with a harness around the hip and cuffs around the legs and walked at 2 km/h, which was found to be a comfortable walking velocity for healthy subjects and a feasible walking velocity for all patients. All subjects walked at the same velocity to record physiological signals at comparable conditions. For safety reasons, all subjects were connected to the body weight support system. Approval for all studies was obtained from local ethics committees, and all subjects gave written informed consent before data collection.

TABLE I
CHARACTERISTICS OF PATIENTS FOR OPEN LOOP CLASSIFIER TRAINING AND CLOSED LOOP CONTROL OF COGNITIVE LOAD. GENDER: M=MALE, F=FEMALE

	Subj.	Sex	Age [y]	T. since inc. [m]	Lesion	Beta blockers
Open Loop	1	f	52	29	left ischemic	no
	2	m	43	5	left hemorrhagic	no
	3	f	37	22	left hemorrhagic	no
	4	m	66	29	left ischemic	no
Closed Loop	1	m	45	9	left ischemic	no
	2	m	47	14	right hemorrhagic	no
	3	m	63	516	right ischemic	no
	4	m	64	60	left ischemic	no
	4	f	54	54	left ischemic	no

III. Results

A. Training Experiments

Physical effort and perceived cognitive load were dissociated in both healthy subjects and patients. While not perfectly independent, the coefficient of determination R^2 between physical effort and cognitive load was 0.22 for healthy subjects and 0.33 for patients.

In healthy subjects, at two distinct levels of physical effort without virtual task, the reported physical effort of healthy subjects increased significantly for the harder physical effort condition compared to the easier physical effort condition

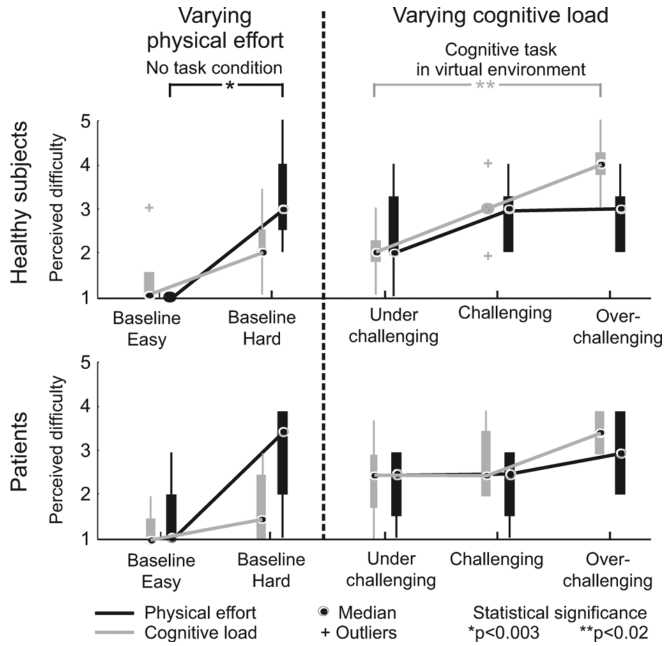


Fig. 5. Boxplots of perceived physical effort and cognitive load in healthy subjects and patients. The graph shows median and 25%–75% percentiles of questionnaire answers.

($p \leq 0.003$). Meanwhile the reported cognitive load increased, but not significantly (Fig. 5, top left) from easy to hard physical effort. While perceived physical effort and perceived cognitive load both increased, the increase in physical effort was significantly higher compared to the increase in cognitive load ($p \leq 0.02$). Conversely, in the cognitive task conditions, the reported physical effort did not increase significantly while the perceived cognitive load increased significantly from the under-challenged to the over-challenged condition ($p \leq 0.02$, Fig. 5, top right). Again, perceived physical effort and perceived cognitive load both increased. However, in the cognitive task condition, the increase in cognitive load was significantly higher compared to the increase in physical effort ($p \leq 0.005$).

In patients, changes in perceived physical effort and perceived cognitive load showed similar trends as in healthy subjects. None of the results in patients were statistically significant, which could possibly be attributed to the small sample size. The change in physical effort during the no task condition increased, but not significantly ($p = 0.13$, Fig. 5, bottom left). The increase in physical effort was almost significantly higher compared to the increase in cognitive load ($p = 0.08$). In the condition with cognitive task, cognitive load increased, but again not significantly ($p = 0.2$, Fig. 5 bottom right).

B. *Classification Performance of Training Experiments*

The classification results from all classification experiments for healthy subjects and patients are summarized in Table II. With four classes, 25% correct classification would correspond to chance. In healthy subject, the classic LDA as well as the KALDA only reached a level slightly above chance when used on physiological signals alone. Physiological signals in combination with force data from the robot did not improve classification results of the classic LDA; the Kalman filter however could improve the classification based on physiological signals and force data by 12%. A key input was the task success in the virtual environment (score), which raised the classification results to 88% correct classification. In patients, the LDA performed equally poor in data sets with physiological signals alone or physiology signals in combination with robot force data. However, the patients could benefit to a much larger extend from the KALDA approach compared to healthy subjects, as improvements of up to 25% correct classification could be achieved. Interestingly, in patients, score information from the virtual environment could not improve classification results of the KALDA classifier.

In healthy subjects and patients, the statistical tests showed a significant improvement of classification results if score information was present. Classifier C1 and C2 had score information available and classified cognitive load better than classifiers C3 and C4 that did not have score information available ($p < 0.007$). The KALDA did not significantly improve the classification results neither in healthy subjects ($p = 0.89$), nor in patients ($p = 0.11$).

The classifier with all information available (C1) did classify cognitive load better than the classifier base only on score (C5). However, results were not statistically significant after the Bonferroni correction was applied ($p = 0.026$ in healthy subject, $p = 0.14$ in patients).

C. *Classifier Performance During Closed Loop Control of Cognitive Load*

On average, classification of cognitive load in healthy subjects during closed loop control was achieved with $87\% \pm 8\%$ correct classification. As explained above, this number refers to the percent match of the classifiers result with the questionnaire answer of the subject. In healthy subjects, the experimenter rating of cognitive load coincided to over 95% with the decision of the subject. This system used a joint input vector of physiological data, force data from the robot and task success data from the virtual environment in combination with the KALDA classifier. Controlling only task success without the use of physiological signals, we obtained an average classification result of $64\% \pm 11\%$ correctly classified (Table III).

In patients, the percent match between the patient's decision and the KALDA classifier was very low However, the experimenter matched the decision of the classifier with 80% (Table III). In healthy subjects, the closed loop system with physiological data, robot data and score data performed significantly better than the system based only on score information ($p = 0.039$).

IV. Discussion

We performed bio-cooperative, closed loop control of cognitive load during robot assisted gait training in healthy subjects and neurological patients after stroke. Using psychophysiological measurements, robot force data and performance data from a virtual environment from open loop experiments, we trained a linear discriminant analysis classifier. As stroke is reported to cause disturbances in autonomic functions and therefore in physiological signals [35]–[37] we used a Kalman filter based, auto-adaptive system (KALDA), which automatically adapted

TABLE II
OPEN LOOP, "LEAVE ONE OUT" CLASSIFICATION RESULTS FOR HEALTHY SUBJECTS AND PATIENTS FOR FOUR DIFFERENT KINDS OF INPUT VECTORS. RESULTS ARE PRESENTED AS PERCENT CORRECTLY CLASSIFIED (MEAN $\pm$ STD)

		Input vector to classifier				
		All	Physiological signals and score	Physiological signals and force from robot	Only physiological signals	Only score
Healthy subjects	Classic LDA	88 ± 9	84 ±10	46±17	34±18	74 ± 7
	KALDA	88±11	85±10	58±11	38±18	75±6
Patients	Classic LDA	60±16	65±25	50±38	45±41	57±5
	KALDA	75±26	70±26	75±10	70±12	60±4

TABLE III
RESULTS OF CLOSED LOOP EXPERIMENTS IN HEALTHY SUBJECTS AND PATIENTS. RESULTS ARE PRESENTED AS PERCENT MATCH BETWEEN THE DECISION OF THE CLASSIFIER AND THE DECISION OF SUBJECT OR EXPERIMENTER. ALTHOUGH THE SUBJECTS AND EXPERIMENTERS WERE ASKED FOR THEIR PERSONAL RATING, THIS INFORMATION DID NOT INFLUENCE THE ADAPTATION OF THE CLOSED LOOP SYSTEM. RESULTS ARE PRESENTED AS PERCENT CORRECTLY CLASSIFIED (MEAN $\pm$ STD)

	With physiological data		Only task success (score)	
	Subject-classifier	Experimenter-classifier	Subject-classifier	Experimenter-classifier
Healthy subjects	87±8	85±10	64±11	68±12
Patients	53±33	80±8	68±17	74±17

the classifier to the physiological responses of subjects. The KALDA generated data labels based upon its own class probabilities and therefore did not need any further input from therapist or subject.

A. Classification Performance

In open loop experiments, 88% correct classification was possible in healthy subjects. The KALDA could only further improve the results of classification that did not rely on the score from the virtual environment. However, in patients, the KALDA allowed for up to 75% correct classification. The score information from the virtual environment was a key input for the classifier, in healthy subjects and patients alike. This appears logical, as the different levels of cognitive load were induced by adjusting task difficulty via task success.

While closed loop control of cognitive load could be achieved with 88% correct predictions in healthy subjects, patient results show only 53% correct classification (Table III). When comparing the results of the classifier with the information obtained from asking the patients, the controller only performed 3% above chance level, as we had asked subjects only if they wanted the task to be easier or harder. However, taking into account possibly decreased self assessment capabilities, the reported answers of patients did often not reflect the objective assessment of the therapist.

In this light, we argue that the $80\% \pm 8\%$ of correct match between the classifiers decision compared to the experimenters rating reflects the capabilities of the classifier more realistically (Table III). Patient 4 for example started the experiment with a virtual task that over-challenged him. Although he obtained a score of 0% in the first 3 min, he wanted the task to be more difficult.

A broader basis of patient data could have potentially improved the open loop and closed loop classification results. In healthy subject, gender [38], age [39], and the presence of other people [40] were shown to have an effect of psychophysiological responses. In stroke and traumatic brain injured patients, changes in heart rate and skin conductance as reaction to stress were shown to be reduced for patients with right hemispheric injury compared to patients with left hemispheric injury [41]. The classification results might therefore improve with a larger pool of patient data or with data from an age matched group of healthy subjects. The low sample size might be responsible for the large standard deviations in the classification results (Tables II and III). In addition, the low sample size might explain why the classic LDA classifier performed worse with all physiological data (Table II, first row) compared to classification with only physiological signals and score information (Table II, second row).

Furthermore, LDA assumes that the input data are normally distributed. However, this was later found not to be the case for several inputs, including score for healthy subjects. Despite the assumption, LDA is to some degree robust to violations of normality, as is for example evidenced by the good classification performance using only healthy subjects' score. Nonetheless, we acknowledge that future studies may wish to evaluate other classifiers that do not require this assumption.

B. The Influence of Physical Effort in Classifying Cognitive Load

If employed during robot-assisted gait training, psychophysiological recordings might be influenced by both motor and cognitive task. Synergistic physiological responses to combined physical and cognitive workload were for example investigated by Novak *et al.* [42] or Wasmund *et al.* [43]. Both found that physical effort influenced recordings during dual tasks, but that the effects on physiology were mostly decoupled.

In our experiments, physical and cognitive activity were dissociated to some degree, since question difficulty was independent of physical activity. However, they were also connected since the cognitive task had to be solved by performing a physical action and was thus embedded into the task to some degree. A full separation of physical effort and cognitive load would of course be desirable from classification point of view. It was, however, therapeutically not desired to fully separate physical effort and cognitive load, as physical effort was shown to be crucial for rehabilitation success [2]. We had, therefore, intentionally designed the setup such that patients had to solve the cognitive tasks with a modulation of motor effort.

While the dissociation of perceived physical effort and perceived cognitive load was not perfect, the change in physical effort was significantly larger than the change in cognitive load during the "no task" condition. Conversely, the change in cognitive load from under-challenged to over-challenged was significantly larger than the change in physical effort (Fig. 5).

This verified that the classifier did indeed classify cognitive load, despite the similar trends of both quantities. HR, HRV, breathing frequency, and skin temperature are influenced by physical effort. This was taken into account in the choice of the baseline measurement. During the baseline measurement, subjects had to vary their physical effort (Fig. 3, "no task condition"). This data was labeled as "baseline" to the classifier. Training the classifier on data that did not include the initial baseline measurement, the classification performance dropped to values below 30% (results not reported).

C. Necessity for Kalman Filters in Neurological Patients

A disturbance of the autonomic functions was often described to affect physiological processes in cerebro-vascular diseases [35]–[37]. In this context, a decrease in HRV (standard deviation of RR intervals, low frequency and high frequency) was found for stroke patients [35], [37] . In addition, it was shown that skin temperature was lower on the contralesional side after stroke [44] and that the sympathetic skin conductance was altered in amplitude and delay [45]. Furthermore, medication of stroke patients can influence physiological signals, as for example beta blockers, which alter the cardio-vascular response to psychological or physical stress. In addition, patients can get exhausted during training. As changes in physiological recordings caused by exhaustion can potentially occur during a long training session, the KALDA with its updating frequency of 1 Hz can take these changes into account.

The classical LDA only reached open loop classification result of 45%–65% correct classification (Table II) due to the large variety of possible changes in physiological responses compared to healthy subjects. Also, the task might have been physically much more demanding for patients, which would alter effort related physiological signals such as mean HR, HRV, or breathing frequency. The Kalman adaptive classifier (KALDA) could take alterations of physiological signals caused by the stroke into account and improved the classification by up to 25% to a maximal classification result of 75%. Note that, with four classes, 25% classification correctness would correspond to chance. For daily clinical use, the basis of patient data will have to be increased to include patients with a variety of different lesions and of different age. The fact that the Kalman Filter did not result in statistically significant improvements of classification compared to the standard LDA is likely the result of the small sample size.

The efficacy of the KALDA for classification of neurological patients became even more apparent when compared to classification results of healthy subjects: the classifier, trained on healthy and patient data, could achieve 88% correct classifier even in its none-adaptive version. Possibly, this is due to the

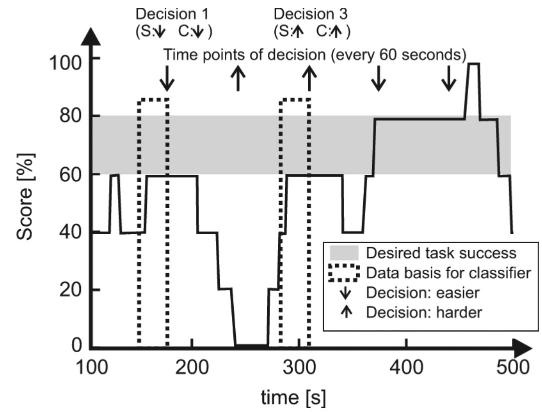


Fig. 6. Exemplary plot from data of healthy subject 4 from closed loop control of cognitive load using the KALDA classifier with physiological input. C: classifiers decision, S: subjects decision.

fact that most healthy subjects responded to our intervention in a similar manner.

However, we must acknowledge that our implementation of unsupervised KALDA was validated empirically rather than theoretically. With improperly selected parameters of the update process, instability could occur, leading to poor classification and decisions that could be detrimental to the patient. In our case, UPC and T_R were specifically trained to avoid such an occurrence, and KALDA thus resulted in better classification than classic LDA. Still, the stability of KALDA in an unsupervised setting should be more extensively validated both theoretically and empirically before it can be introduced into clinical practice.

D. The Necessity of Physiological Signals in a Clinical Setting

From the viewpoint of clinical applicability, a classifier based on task performance alone would be preferable compared to a classifier which also included physiological signals. Measuring physiological signals always included sensor placement, which was time consuming for the experimenter and uncomfortable for the subject. In particular, the time to place the sensors will be an issue during rehabilitation, as training-time is limited. With a classifier based on task performance, no further physiological measurements such as ECG, skin conductance, breathing, or skin temperature would be necessary. All information needed for the classification would be provided by the virtual environment.

We, therefore, examined the necessity of physiological signals for classification of cognitive load compared to a controller that adapted the virtual environment solely based on the subject's performance in the task (Table III). The score alone provided 74% correct classification in healthy subjects and 60% correct classification in patients. The physiological signals only improved the results by 13% and 15% in healthy subjects and patients, respectively (Table II). In healthy subjects, this increase was however significant. While this is only a small improvement compared to the additional effort of attaching the sensors, the importance of physiological data in the decision process of classification is exemplified with data from healthy subject 4 (Fig. 6). The classifier only takes the last 30 s of data into account (dashed box in plot). The classifier (C) with physiological signals as input decided in decision 1 to make the task easier, and in decision 3 to make the task harder. In both cases, the classifier's decision coincided with the decision of the subject (S). The score controller would have decided in both cases to make the task harder, as the optimal range of task success ($\geq$ 70%) was not yet reached and the data for the decision, acquired in the last 30 s, was similar in both cases. Therefore, the physiological data provided the deciding information on the cognitive load of the subject.

This shows that for control of cognitive load, physiological signals can be a necessary source of information required by the classifier. The score-based classifier showed good performance for extreme conditions under-challenged and over-challenged, but its performance dropped for situations, in which the subject reached a challenged state. However, for clinical use, down-scaling of the required physiological signals might provide a good tradeoff between effort for therapeutic staff involved with attachment of sensors and the benefit of assessing cognitive load. The large standard deviation in results of "only score" classification in patients (Table III) might result from the small sample size. For a definite answer if psychophysiological signals improve the classification statistically significantly, further recordings have to be performed.

V. Conclusion and Outlook

The key result of this study is that real time, objective assessment and control of cognitive load was possible by using a combination of psychophysiological measurements and task performance as source for state estimation. For the first time, closed loop control of cognitive load has been performed in neurological patients during robot-assisted gait training. Performance metrics can be used to replace psychophysiological recordings; this however decreases the classification quality and classification quality suffers.

By putting the human into the center of a cognitive control loop, the classical master–slave paradigm could be avoided, which requires the user to adapt to the robotic system. Focusing on integrating bio-cooperative closed loop control based on physiological signals reflecting psychological parameters could result in a setup, in which the robotic system adapts to the user. The use of adaptive algorithms for intelligent machine learning as described above could be the basis for future rehabilitation devices that automatically adapt to the specific needs and demands of the patient. In general, our adaptive algorithms for classification and control of cognitive load are not limited to rehabilitation, but could be used during any human–machine interaction.

Acknowledgment

The authors would like to thank K. Campen from the "Zentrum fuer Ambulante Rehabilitation" for help with patient recruitment and M. Munih for advice and proofreading of the manuscript.

References

[1] A. Kaelin-Lang, L. Sawaki, and L. G. Cohen, "Role of voluntary drive in encoding an elementary motor memory," *J. Neurophysiol.*, vol. 93, no. 2, pp. 1099–1103, 2005.

[2] M. Lotze, C. Braun, N. Birbaumer, S. Anders, and L. G. Cohen, "Motor learning elicited by voluntary drive," *Brain*, vol. 126, no. Pt 4, pp. 866–72, 2003.

[3] N. Maclean and P. Pound, "A critical review of the concept of patient motivation in the literature on physical rehabilitation," *Soc. Sci. Med.*, vol. 50, no. 4, pp. 495–506, 2000.

[4] J. A. Taylor and K. A. Thoroughman, "Motor adaptation scaled by the difficulty of a secondary cognitive task," *PLoS One*, vol. 3, no. 6, p. E2485, 2008.

[5] G. M. Redding, S. D. Rader, and D. R. Lucas, "Cognitive load and prism adaptation," *J. Motor Behavior*, vol. 24, no. 3, pp. 238–246, 1992.

[6] M. K. Holden, "Virtual environments for motor rehabilitation: Review," *Cyberpsychol. Behav.*, vol. 8, no. 3, pp. 212–9, 2005.

[7] V. G. Duffy, *Handbook of Digital Human Modeling: Research for Applied Ergonomics and Human Factors Engineering*. Boca Raton, FL: CRC, 2008.

[8] J. Andreassi, *Psychophysiology: Human Behavior and Physiological Response*, 5th ed. Mahwah: Lawrence Erlbaum Associates, 2007.

[9] K. Hugdahl, *Psychophysiology: The Mind-Body Perspective*. Cambridge, U.K.: Harvard Univ. Press, 1995.

[10] J. P. A. Delaney and D. A. Brodie, "Effects of short-term psychological stress on the time and frequency domains of heart-rate variability," *Perceptual Motor Skills*, vol. 91, no. 2, pp. 515–524, 2000.

[11] R. McCraty, M. Atkinson, W. A. Tiller, G. Rein, and A. D. Watkins, "The effects of emotions on short-term power spectrum analysis of heartrate- variability," *Am. J. Cardiol.*, vol. 76, no. 14, pp. 1089–1093, 1995.

[12] J. T. Cacioppo, L. Tassinary, and G. Berntson, *Handbook of Psychophysiology*, 2nd ed. Cambridge, U.K.: Cambridge Press 2000, 2000.

[13] A. F. Ax, "The physiological differentiation between fear and anger in humans," *Psychosomatic Med.*, vol. 15, no. 5, pp. 433–442, 1953.

[14] M. E. Dawson, A. M. Schell, and D. L. Filion, *Handbook of Psychophysiology*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007, pp. 159–181.

[15] D. Carroll, J. R. Turner, and R. Prasad, "The effects of level of difficulty of mental arithmetic challenge on heart rate ande oxygen-consumption," *Int. J. Psychophysiol.*, vol. 4, no. 3, pp. 167–173, 1986.

[16] F. A. Boiten, N. H. Frijda, and C. J. E. Wientjes, "Emotions and respiratory patterns: Review and critical analysis," *Int. J. Psychophysiol.*, vol. 17, no. 2, pp. 103–128, 1994.

[17] J. G. Rousselle, J. Blascovich, and R. M. Kelsey, "Cardiorespiratory response under combined psychological and exercise stress," *Int. J. Psychophysiol.*, vol. 20, no. 1, pp. 49–58, 1995.

[18] R. Levenson, P. Ekman, and W. Friesen, "Voluntary facial action generates emotion-specific autonomic nervous system activity," *Psychophysiology*, vol. 27, no. 4, pp. 363–384, 1990.

[19] P. Gomez, P. Zimmermann, S. Guttormsen-Schr, and B. Danuser, "Respiratory responses associated with affective processing of film stimuli," *Biol. Psychol.*, vol. 68, no. 3, pp. 223–235, 2005.

[20] D. Palomba, M. Sarlo, A. Angrilli, A. Mini, and L. Stegagno, "Cardiac responses associated with affective processing of unpleasant film stimuli," *Int. J. Psychophysiol.*, vol. 36, no. 1, pp. 45–57, 2000.

[21] M. Codispoti, P. Surcinelli, and B. Baldaro, "Watching emotional movies: Affective reactions and gender differences," *Int. J. Psychophysiol.*, vol. 69, no. 2, pp. 90–95, 2008.

[22] M. Ohsuga, F. Shimono, and H. Genno, "Assessment of phasic work stress using autonomic indices," *Int. J. Psychophysiol.*, vol. 40, no. 3, pp. 211–220, 2001.

[23] S. E. Rimm-Kaufman and J. Kagan, "The psychological significance of changes in skin temperature," *Motivation Emotion*, vol. 20, no. 1, pp. 63–78, 1996.

[24] R. Riener and M. Munih, "Guest editorial: Special section on rehabilitation via bio-cooperative control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 4, p. 2, Aug. 2010.

[25] A. Koenig, X. Omlin, L. Zimmerli, M. Sapa, C. Krewer, M. Bolliger, F. Mueller, and R. Riener, "Psychological state estimation from physiological recordings during robot assisted gait rehabilitation," *J. Rehabil. Res. Develop.*, vol. 48, no. 4, pp. 367–386, 2011.

[26] D. Novak, J. Ziherl, A. Olensek, M. Milavec, J. Podobnik, M. Mihelj, and M. Munih, "Psychophysiological responses to robotic rehabilitation tasks in stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 4, pp. 351–61, Aug. 2010.

[27] P. Rani, J. Sims, R. Brackin, and N. Sarkar, "Online stress detection using psychophysiological signals for implicit human-robot cooperation," *Robotica*, vol. 20, p. 13, 2002.

[28] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol Psychol*, vol. 84, p. 394421, 2010.

[29] I. Christov, "Real time electrocardiogram QRS detection using combined adaptive threshold," *BioMedical Engineering OnLine*, vol. 3, no. 28, p. 8, 2004.

[30] M. Malik, "Heart rate variability—standards of measurement, physiological interpretation and clinical use," *Eur. Heart J.*, vol. 17, p. 28, 1996.

[31] R. Banz, M. Bolliger, G. Colombo, V. Dietz, and L. Lunenburger, "Computerized visual feedback: An adjunct to robotic-assisted gait training," *Phys Ther*, vol. 88, no. 10, pp. 1135–45, 2008.

[32] C. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006, Information Science and Statistics.

[33] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 550–6, Mar. 2007.

[34] D. Novak, M. Mihelj, J. Ziherl, A. Olensek, and M. Munih, "Measuring motor actions and psychophysiology for task difficulty estimation in human-robot interaction," in *Proc. Measuring Behavior*, 2010, pp. 269–272.

[35] J. T. Korpelainen, K. A. Sotaniemi, H. V. Huikuri, and V. V. Myllyla, "Abnormal heart rate variability as a manifestation of autonomic dysfunction in hemispheric brain infarction," *Stroke*, vol. 27, no. 11, pp. 2059–2063, 1996.

[36] S. Meyer, M. Strittmatter, C. Fischer, T. Georg, and B. Schmitz, "Lateralization in autononic dysfunction in ischemic stroke involving the insular cortex," *NeuroReport*, vol. 15, no. 2, pp. 357–361, 2004.

[37] S. L. Tokgozoglu, M. K. Batur, M. A. Topcuoglu, O. Saribas, S. Kes, and A. Oto, "Effects of stroke localization on cardiac autonomic balance and sudden death," *Stroke*, vol. 30, no. 7, pp. 1307–1311, 1999.

[38] E. Carrillo, L. Moya-Albiol, E. Gonzlez-Bono, A. Salvador, J. Ricarte, and J. Gmez-Amor, "Gender differences in cardiovascular and electrodermal responses to public speaking task: The role of anxiety and mood states," *Int. J. Psychophysiol.*, vol. 42, no. 3, pp. 253–264, 2001.

[39] T. Ritvanen, V. Louhevaara, P. Helin, S. Visnen, and O. Hnninen, "Responses of the autonomic nervous system during periods of perceived high and low work stress in younger and older female teachers," *Appl. Ergonomics*, vol. 37, no. 3, pp. 311–318, 2006.

[40] A. C. Phillips, D. Carroll, K. Hunt, and G. Der, "The effects of the spontaneous presence of a spouse/partner and others on cardiovascular reactions to an acute psychological challenge," *Psychophysiology*, vol. 43, no. 6, pp. 633–640, 2006.

[41] S. Andersson and A. Finset, "Heart rate and skin conductance reactivity to brief psychological stress in brain-injured patients," *Journal of Psychosomatic Research*, vol. 44, no. 6, pp. 645–656, 1998.

[42] D. Novak, M. Mihelj, and M. Munih, "Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction," *Robotica*, pp. 1–8, 2010.

[43] W. L. Wasmund, E. C. Westerholm, D. E. Watenpaugh, S. L. Wasmund, and M. L. Smith, "Interactive effects of mental and physical stress on cardiovascular control," *J. Appl. Physiol.*, vol. 92, no. 5, pp. 1828–1834, 2002.

[44] H. Naver, C. Blomstrand, S. Ekholm, C. Jensen, T. Karlsson, and B. G. Wallin, "Autonomic and thermal sensory symptoms and dysfunction after stroke," *Stroke*, vol. 26, no. 8, pp. 1379–1385, 1995.

[45] J. Korpelainen, U. Tolonen, K. Sotaniemi, and V. Myllyla, "Suppressed sympathetic skin response in brain infarction," *Stroke*, vol. 24, no. 9, pp. 1389–1392, 1993.

Alexander Koenig received the M.Sc. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 2006, and the Ph.D. degree in mechanical engineering from the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, in 2011.

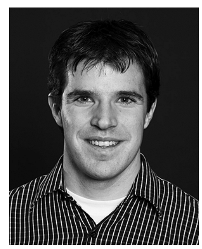
He currently holds a position as Associate Researcher at Harvard University in the Motion Analysis Lab. His research interests include theoretical models for stroke rehabilitation, gait rehabilitation robotics and psychophysiology.

Domen Novak received the degree in electrical engineering, in 2008, from the University of Ljubljana, Ljubljana, Slovenia, where he is now a Ph.D. student.

His research interests include applied psychophysiology, pattern recognition, and biomedical signal processing.

Ximena Omlin received the M.S. degree in human movement science from the Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, in 2009. She is currently working toward the Ph.D. degree at the Sensory-motor Systems Lab, ETH Zurich.

In 2009, she joined the Sensory-motor Systems Lab of Prof. R. Riener as a Research Assistant. Her work included the evaluation of novel control approaches for robot-assisted gait training in healthy subjects and neurological patients.



Michael Pulfer received the M.S. degree in electrical engineering and Information Technology at ETH Zurich, Switzerland, in 2010. The area of specialization was electrical power systems and mechatronics.

He is currently working at Helbling Technik AG, Switzerland as a Development Engineer in the field of new and innovative power tools.

Eric Perreault received the B.Eng. and M. Eng. degrees in electrical engineering from McGill University, Montreal, QC, Canada, in 1989 and 1991, respectively. After working in industry for four years, he enrolled in the Biomedical Engineering Department at Case Western Reserve University, Cleveland, OH, completing the Ph.D. degree in 2000.

He is an Associate Professor at Northwestern University, with appointments in the Department of Biomedical Engineering and the Department of Physical Medicine and Rehabilitation. He also is a member of the Sensory Motor Performance Program at the Rehabilitation Institute of Chicago, where his laboratory is located. From 2000 to 2002, he was a Postdoctoral Fellow in the Department of Physiology at Northwestern University. In 2010, he was a Visiting Professor at ETH Zürich. His current research focuses on understanding the neural and biomechanical factors involved in the control of multi-joint movement and posture and how these factors are modified following neuromotor pathologies such as stroke and spinal cord injury. The goal of his research is to provide a scientific basis for understanding normal and pathological motor control that can be used to guide rehabilitative strategies and user interface development for restoring function to individuals with motor deficits. Current applications include rehabilitation following stroke and tendon transfer surgeries, and user interfaces for neuroprosthetic control. He serves on the editorial boards for the *Journal of Motor Behavior* and the *Journal of Motor Control*.

Dr. Perreault currently serves as an Associate for the IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING. He also is a member of the IEEE Technical Committee on Rehabilitation Robotics.

Lukas Zimmerli received the M.Sc. degree in 2005. He started his doctorate studies in the field of rehabilitation at the Sensory-Motor Systems Lab, ETH Zurich, in 2008.

He is currently employed as a software engineer by the medical engineering company Hocoma AG, Volketswil, Switzerland. His research interests include the development of virtual reality environments for therapeutical purposes and their deployment to improve motivational aspects of current rehabilitation techniques.

Matjaz Mihelj received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia, 1999 and 2002, respectively.

Since 2009 he has been affiliated with the university as an Associate Professor. In 2001 he was a Visiting Researcher at Sendai University, Japan. From May 2005 to April 2006, he was a Visiting Researcher at the Automatic Control Lab, ETH Zurich with the Marie Curie Intra-European fellowship. His rehabilitation robotics research performed at ETH Zurich was awarded with the Swiss Technology Award. His current research interests are in man–machine interaction, haptic interfaces as well as modeling and control of biological systems with the main focus on restoration of lost functions in people with disabilities.

Robert Riener studied mechanical engineering at TU München, Munich, Germany, and University of Maryland, College Park. He received a Ph.D. degree in engineering from the TU München, in 1997. After postdoctoral work from 1998 to 1999 at the Centro di Bioingegneria, Politecnico di Milano, he returned to the TU München, where he completed his habilitation in the field of biomechatronics. in 2003.

From 2003 to 2006, he was Assistant Professor of Rehabilitation Engineering at the Automatic Control Laboratory of the ETH Zurich and Spinal Cord Injury Center of the University Hospital Balgrist ("double-professorship"). From 2006 to 2010, he has been Associate Professor and since June 2011 Full Professor for Sensory-Motor Systems at the Mechanical Engineering Department of ETH Zurich, while maintaining the double professorship with the University of Zurich. From 2008 to 2011, he was the Head of the Institute of Robotics and Intelligent Systems (IRIS) at ETH Zurich. Since 2009 he is the Deputy Head of the Mechanical Engineering Department, ETH Zurich. His research interests include rehabilitation robotics, virtual reality in medicine, biomechanics, and biomechatronics. He has authored and co-authored more than 300 peer-reviewed journal and conference articles, 18 book chapters, as well as 19 patents. He was or still is the principal investigator of 20 competitive national and international research projects. He was a Director of the IFESS and President of the AUTOMED. Currently, he is an editorial member of six international journals, including *Applied Bionics and Biomechanics*. Furthermore, he serves as reviewer for many international journals as well as for the Swiss National Science Foundation (SNF), the German Ministry of Research (BMBF), the Deutsche Forschungsgemeinschaft (DFG) and the Engineering and Physics Sciences Research (EPSRC), United Kingdom.

Dr. Riener has received more than 15 personal distinctions and awards including the humanTech Innovation Prize in 2005, the Swiss Technology Award in 2006, and the IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING (TNSRE) Best Paper Award 2010. He is editorial member the IEEE TNSRE