Psychophysiological Measurements in a Biocooperative Feedback Loop for Upper Extremity Rehabilitation

Domen Novak, Matjaž Mihelj, Jaka Ziherl, Andrej Olenšek, and Marko Munih, *Member, IEEE*

***Abstract*—This paper examines the usefulness of psychophysiological measurements in a biocooperative feedback loop that adjusts the difficulty of an upper extremity rehabilitation task. Psychophysiological measurements (heart rate, skin conductance, respiration, and skin temperature) were used both by themselves and in combination with task performance and biomechanics. Data fusion was performed with discriminant analysis, and a special adaptive version was implemented that can gradually adapt to a subject. Both healthy subjects and hemiparetic patients participated in the study. The accuracy of the biocooperative controller was defined as the percentage of times it matched the subjects' preferences. The highest accuracy rate was obtained for task performance (approximately 82% for both healthy subjects and patients), with psychophysiological measurements yielding relatively low accuracy (approximately 60%). The adaptive approach increased accuracy of psychophysiological measurements to 76.4% for healthy subjects and 68.8% for patients. Combining psychophysiology with task performance yielded an accuracy rate of 84.7% for healthy subjects and 89.4% for patients. Results suggest that psychophysiological measurements are not reliable as a primary data source in motor rehabilitation, but can provide supplementary information. However, it is questionable whether the amount of additional information justifies the increased complexity of the system.**

***Index Terms*—Biocooperative robotics, human factors, multimodal interfaces, psychophysiological measurements, rehabilitation robotics.**

I. Introduction

ROBOTIC interfaces are becoming increasingly common in motor rehabilitation [1]. In the long term, exercise with such devices yields results comparable to intensive exercise with a therapist [2]. Additionally, they offer an objective estimation of the patient's motor performance and functional improvement [3]. Frequently, they are combined with virtual environments in order to make rehabilitation more interesting and motivational [4].

Manuscript received October 14, 2010; revised February 23, 2011; accepted June 13, 2011. Date of publication June 23, 2011; date of current version August 10, 2011. This work was supported in part by the EU Information and Communication Technologies Collaborative Project MIMICS grant 215756 and in part by the Slovenian Research Agency.

The authors are with the Faculty of Electrical Engineering, University of Ljubljana, SI-1000 Ljubljana, Slovenia (e-mail: domen.novak@robo.fe.uni-lj.si; matjaz.mihelj@robo.fe.uni-lj.si; jaka.ziherl@robo.fe.uni-lj.si; andrej.olensek@robo.fe.uni-lj.si; marko@robo.fe.uni-lj.si).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2011.2160357

Early rehabilitation robots were able to provide active assistance to the patient, but did not adapt their movement to the activity (or passivity) of the patient. Instead, the affected limb was moved along a predefined, fixed trajectory. This problem was addressed by patient-cooperative or "assist as needed" control techniques. By recognizing the patient's movement intentions and motor abilities, these techniques adapt the robotic assistance to the activity (or passivity) of the patient. They have been successfully used for rehabilitation of both the lower (e.g., [5]) and upper extremities (e.g., [6]). Recently, the concept of patient-cooperative robotics has been extended to biocooperative robotics, which take into account not only the forces and movements applied by the subjects, but also psychological states.

Psychological factors such as motivation are known to be very important to the success of rehabilitation. Numerous reviews have shown that it is important to start intensive therapy as early as possible and that therapeutic outcome improves with increasing training intensity (e.g., [7]). Encouraging unmotivated patients thus improves the likelihood of their eventual recovery [8], [9]. The goal of biocooperative rehabilitation is thus to automatically adjust the therapy parameters so that the patient is challenged in a moderate but engaging and motivating way without causing undue stress or harm, thus hopefully resulting in longer and more intensive therapy [10]. However, measuring psychological states has proven to be more difficult than measuring forces and movements. Questionnaires are not a good solution, as they require therapy to be interrupted and only provide information "after the fact." In biocooperative robotics, an unobtrusive, real-time method of measuring psychological states would be very useful.

A possible solution to indirectly measure the subject's psychological state would be through psychophysiological measurements, which can be defined as the measurements of physiological responses to changes in psychological state. Perhaps the best-known psychophysiological responses are increased sweating and changes in heart rate as a result of anxiety, but psychophysiological responses have been connected to other emotions such as anger, fear, and sadness. A thorough review of psychophysiological responses to different emotions was recently performed by Kreibig [11].

Psychophysiological measurements can be taken without the subject's active cooperation, providing a convenient, objective, and unobtrusive method of estimating arousal, stress, engagement, etc. Because of these advantages, they have been implemented in situations such as flight simulators [12], interaction with mobile robots [13], and interaction with haptic

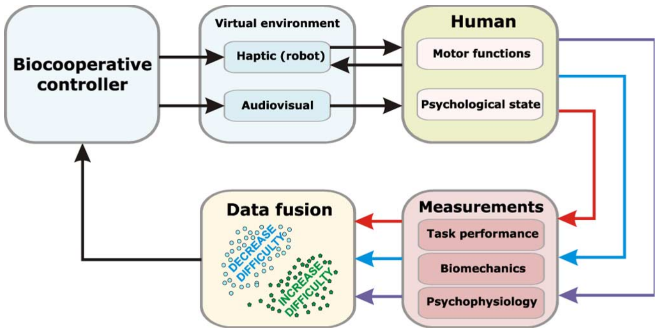

Fig. 1. The principle of a biocooperative feedback loop.

robots [14]. After they were suggested for use in motor rehabilitation [15], psychophysiological measurements were found to provide useful information about stroke patients' psychological states during robot-aided upper extremity rehabilitation [16]. They could thus be used in a biocooperative feedback loop, offering information about the patient that cannot be obtained from forces and movements.

Biocooperative systems have already attempted to either predict heart rate as a result of physical activity [17] or change the level of assistance provided by a rehabilitation system based on heart rate [18]. However, these systems used only one measurement (heart rate). It has long been known that multiple psychophysiological measurements need to be combined in order to obtain a better estimate of a person's psychological state [19]. So far, no closed-loop biocooperative system that combines multiple psychophysiological measurements has been implemented in motor rehabilitation, and such psychophysiological closed loops also represent a significant research challenge in other fields (see [20] for a multidisciplinary review).

Our paper presents a biocooperative feedback loop for upper extremity rehabilitation that adapts the difficulty of a task based on a fusion of task performance, biomechanical measurements (forces and movements), and four psychophysiological signals (heart rate, skin conductance, respiration, and peripheral skin temperature). Fusion is performed using discriminant analysis [21]. Since psychophysiological responses exhibit high intersubject variability, we also propose a method of adapting the system to a particular subject. The system is first trained using data from other subjects, then gradually adapts to the current subject as it obtains more and more data about the subject.

The goal of our study was to determine how much information psychophysiological measurements can provide in a biocooperative feedback loop, both when used by themselves and when combined with other sources of information.

II. Materials and Methods

The basic building blocks of a biocooperative feedback loop are shown in Fig. 1: the virtual environment, patient, measurements, data fusion, and biocooperative controller. The hardware underlying the feedback loop is described in Section II-B. Section II-C describes the virtual environment used by our subjects while Section II-D describes the study protocol and the measurements. Section II-E gives details on the subjects who participated in the study. Section II-F describes the measurement processing and the features extracted from the raw measurements. Using methods described in Section II-G, these features are then fused into an estimate of whether task difficulty should be increased or decreased in order to optimally challenge the patient. The ability of the biocooperative controller to adapt task difficulty was tested using the approach described in Sections II-H and II-I.

A. Ethical Approval

Before the study began, ethical approval was obtained both from the National Medical Ethics Committee of the Republic of Slovenia and from the Medical Ethics Committee of the University Rehabilitation Institute of the Republic of Slovenia.

B. Hardware

The HapticMaster robot [22], developed by Moog FCS, was used as the haptic interface. This robot offers movement with three degrees of freedom. Its end-point is equipped with force sensors as well as a two-axis gimbal with a two-degree-of-freedom passive grasping module. The subject's arm was additionally supported by two cuffs fastened above and below the elbow. These cuffs were connected to a motorized pulley which applied a constant pulling force in order to compensate for the gravity acting on the subject's arm. A 1.4×1.4-m screen was used to display visual data. Subjects sat approximately 1.25 m in front of the screen, with the robot situated between the seat and the screen.

Physiological signals were sampled at 1.2 kHz using a g.USBamp signal amplifier (g.tec Medical Engineering GmbH). The electrocardiogram (ECG) was recorded using four disposable surface electrodes placed in a configuration suggested by the manufacturer of the signal amplifier (two on the chest, one on the abdomen, and one on the back). Skin conductance was measured using a g.GSR sensor (g.tec). The electrodes were placed on the medial phalanges of the second and third fingers of the idle hand. Respiratory rate was obtained using a thermistor-based SleepSense Flow sensor placed beneath the nose. Peripheral skin temperature was measured using a g.TEMP sensor (g.tec) attached to the distal phalanx of the fifth finger of the idle hand.

A recent review of psychophysiological studies found that heart rate and skin conductance are by far the most commonly studied psychophysiological measurements of the autonomic nervous system, with respiration and skin temperature also frequently used [11]. They were thus included in our study. Measurements such as electroencephalography and facial electromyography were initially considered, but we chose to focus on autonomic nervous system measurements where the sensors are unobtrusive and can be attached or removed quickly.

C. Virtual Rehabilitation Task

The rehabilitation task was previously used in a study of stroke patients' psychophysiological responses to robot-aided rehabilitation [16]. It combines reaching and grasping exercise. In the center of the screen, there is a table sloped toward the subject. At the beginning of the task, a ball appears at the top of the slope and starts rolling downward. The subject's goal is

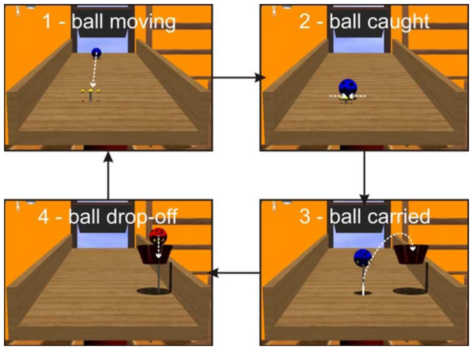

Fig. 2. The virtual rehabilitation task. A ball appears on the top of a sloped table (1) and begins to roll down. The subject then catches it (2) and carries it toward a basket that appears above the table (3). Once the ball is above the basket (4), the subject drops it into the basket and a new ball appears.

to catch the ball before it reaches the lower end of the table. Once the ball is grasped, a basket appears above the table. The subject must then place the ball into the basket. Once the ball is dropped into the basket or falls off the table, another ball appears at the top of the table, the basket disappears and the task continues. Screenshots of the task are shown in Fig. 2. The robot's haptic feedback allows the subject to feel the forces associated with each virtual item. A photograph of a subject performing the task and interfacing with the robot is available in the previous study [16].

Though various modes of active robotic support were offered in the original task, only one was used for our study. If a subject is unable to open or close his or her hand, the robot can automatically grasp the ball as long as the subject's hand is in the correct position.

Seven different difficulty levels were implemented, with higher levels featuring progressively smaller and faster balls. While the first level is very easy (the ball is very large and requires approximately 15 s to cross the table), the seventh is almost impossible (the ball crosses the table in less than 3 s and has a radius of 1/5 the radius from the first level). The third level is the one that was used in the previous study. The ultimate goal of the biocooperative feedback loop was to change the difficulty level so that the subject is optimally challenged.

D. Study Protocol

The study was divided into two phases: the open-loop phase (where task difficulty is adjusted manually by the subject and experiment supervisor) and the closed-loop phase (where task difficulty is adjusted by the biocooperative controller). The open-loop phase was conducted first, with the goal of obtaining a larger set of data for analysis and for training a biocooperative controller. It was performed first with healthy subjects, then with hemiparetic patients. The open-loop phase was necessary since the connections between psychophysiological responses and psychological states are still controversial to a large degree and are, to some degree, task-specific. Although it is possible to identify psychological states from psychophysiological responses using expert-defined rules, without the need for any training data (e.g., [23]), such an approach can be risky since the defined rules may be inaccurate. Thus, we felt that it would be more reliable to first obtain a large open-loop data set which could be used to train a biocooperative controller. After training the biocooperative controller using the open-loop data, the controller was tested in the closed-loop phase with a smaller number of both healthy subjects and hemiparetic patients.

The experiment procedure for both phases was similar. The experiment was conducted in a dedicated room at the University Rehabilitation Institute of the Republic of Slovenia. Three people were present: the subject, experiment supervisor, and occupational therapist. Upon arrival, subjects were informed of the purpose and procedure of the experiment, then signed an informed consent form. Then, they were seated in front of the robot. One arm (the paretic arm for patients, the right arm for healthy subjects) was strapped into the cuffs and grasping device, and the physiological sensors were attached. The third level of the task was demonstrated, and subjects were allowed to practice it briefly.

After practice, the subject rested for 2 min while baseline physiological measurements were recorded. Then, the subject began performing the task at level 3, 4, or 5 (randomly chosen). After 2 min of performing the task at that difficulty level, the task was paused briefly and the subject was asked whether he or she would prefer the difficulty of the task to increase or decrease. Subjects were not given the option to stay at the same difficulty level. Obviously, it is possible that a subject finds the current difficulty to be "just right" and does not wish to change it. However, we chose to offer only two choices for two reasons. First, this simplifies data fusion by reducing the problem to two choices rather than three. Second, we found in pretesting that subjects tended to disproportionately keep difficulty at the same level if offered the option, even if visibly frustrated or bored and even if encouraged by the experimenter to change the difficulty. This was likely due to a desire to please the experimenter and therapist by not reporting any dissatisfaction with the system.

Before asking the subject about his or her preference, the experimenter also noted his own opinion of whether difficulty should increase or decrease. We thus obtained a second, more objective opinion of what difficulty would be appropriate for the subject. The issue of the reliability of self-report measures has been previously raised in psychophysiology, and the opinions of an observer have been suggested as an alternative or validation measure [24]. The experimenter's opinion was, of course, also subjective to a degree and was based on factors such as the patient's task performance, level of physical exertion, verbal comments, and facial expressions.

In the open-loop phase, once the subject had stated his or her preference, the difficulty changed by one or two levels in the direction chosen by the subject. This randomness was introduced in order to expose subjects to a wider range of difficulty levels and create a more robust training data set. If difficulty had always changed by one level, the system would have most likely quickly reached a "steady state" where difficulty alternated between increasing and decreasing. In the closed-loop phase, the difficulty changed in the direction chosen by the biocooperative controller.

After task difficulty was changed, the task began again at the new difficulty. In total, the subject went through six 2-min periods, with the subject's preference noted and the difficulty changing after each one. After the final task period, the experiment was concluded.

E. Subjects

Twenty-four healthy subjects (20 males, four females, age 31.1 ± 10.9 years, age range 21–61) and 11 hemiparetic patients (eight males, three females, age 43.2 ± 13.5 years, age range 22–69) participated in the open-loop phase of the study. Ten healthy subjects (nine males, one female, age 33.9 ± 12.6 years, age range 22–62) and six hemiparetic patients (four male, two female, age 58.3 ± 6.3 years, age range 54–67) participated in the closed-loop phase of the study. All patients were undergoing motor rehabilitation at the University Rehabilitation Institute of the Republic of Slovenia and were tested with the Functional Independence Measure (FIM) [25] and Mini-Mental State Examination (MMSE) [26] within a week of the experiment session. All patients scored at least 26 out of a possible 30 on the MMSE and can thus be considered cognitively intact. None of the patients had been diagnosed with visual neglect.

The patients in the open-loop group were hemiparetic as a result of intracerebral hemorrhage (three subjects), cerebral infarction (four subjects), or surgery of a neoplasm of the brain (four subjects). Time since stroke onset or surgery was 216 ± 228 days (minimum 14, maximum 749). Score on the FIM was 103 ± 14 (out of a possible 126). Six suffered from hemiparesis of the left side of the body and five suffered from hemiparesis of the right side of the body.

The patients in the closed-loop group were hemiparetic as a result of subarachnoid hemorrhage (one subject), intracerebral hemorrhage (two subjects), cerebral infarction (two subjects), or surgery of a neoplasm of the brain (one subject). Time since stroke onset or surgery was 166 ± 34 days (minimum 110, maximum 202). Score on the FIM was 108 ± 5. Three suffered from hemiparesis of the left side of the body and three suffered from hemiparesis of the right side of the body.

A majority of the patients had received secondary stroke prevention drugs (including antihypertensives) prior to participation in the study. Seven patients in the open-loop group and one patient in the closed-loop group had received low doses of psychotropics that had no noticeable side-effects.

With 24 healthy subjects and 11 patients in the open-loop phase, there were thus 144 task periods for healthy subjects and 66 task periods for patients in the open-loop phase. With 10 healthy subjects and six patients in the closed-loop phase, there were thus 60 task periods for healthy subjects and 36 task periods for patients in the closed-loop phase.

F. Feature Extraction

Twenty-six features were calculated from the raw signals for each 2-min task period. They can be divided into three groups: task performance (four features), biomechanics (eight features), and psychophysiology (14 features).

1) Task Performance: Performance features describe how well a subject did during a particular time period and how long he or she had been performing the task. The four features used were the *difficulty level* (1–7), the *time period* (1—first, 6—last), the *percentage of caught balls*, and the *percentage of balls placed into the basket.*

2) Biomechanics: Biomechanical features describe the forces and movements applied by the subjects. The eight features used were *mean absolute force, mean absolute velocity, mean absolute acceleration, total work, mean frequency of the position signal, mean frequency of the velocity signal, mean frequency of the acceleration signal, and mean frequency of the force signal*. All of these were calculated only for movement in the horizontal plane when the subject is trying to catch the ball (since the part of the task where the subject is placing the ball into the basket remains the same in all difficulty levels). Mean frequencies were calculated using Welch's method of modified periodograms.

3) Psychophysiology: Four physiological signals were recorded: the electrocardiogram, skin conductance, respiration, and skin temperature. From the ECG, the intervals between two normal heartbeats (NN intervals) were extracted. Then, *mean heart rate* as well as several measures of heart rate variability (HRV) were calculated: *the standard deviation of NN intervals (SDNN), the square root of the mean squared differences of successive NN intervals (RMSSD), the number of interval differences of successive NN intervals greater than 50 ms divided by the total number of NN intervals (pNN50), total power in the high-frequency heart rate band*, and *total power in the low-frequency heart rate band*. More information about all these measures of HRV is available in [27].

The skin conductance signal can be divided into two components: the skin conductance level (SCL) and skin conductance responses (SCRs). The SCL is the baseline level of skin conductance in the absence of discrete environmental events. *Mean SCL* and *mean derivative of SCL* were calculated. SCRs are transient increases in skin conductance whose amplitude exceeds $0.05\ \mu$S and whose peak occurs less than 5 s after the beginning of the increase. *SCR frequency* and *mean SCR amplitude* were calculated.

Mean respiratory rate and *standard deviation of respiratory rate* were calculated from the respiration signal.

Final skin temperature was calculated as the mean temperature during the last 5 s of each period. Additionally, the *mean derivative of skin temperature* was calculated over the entire time period.

Due to large intersubject differences in baseline values, absolute values of most psychophysiological features were not used in data fusion. Instead, relative values were calculated by subtracting the baseline value from the absolute value or by subtracting the baseline value from the absolute value and dividing the result by the baseline value. The second definition was used for features where baseline values varied among subjects by a factor of more than two: *SCR frequency*, *standard deviation of respiratory rate*, and all measures of HRV except *pNN50*. There were only two psychophysiological features where absolute values were used: the *mean derivative of SCL* and *mean derivative of skin temperature* (which are already time-derivatives).

G. Data Fusion

After the feature extraction step, a number of different features are available for each time period. These features must then be fused into a common estimate of how task difficulty should be changed. As previously mentioned in Section II-D, we felt that the easiest approach to data fusion would be to create a training data set where the subject's preference is known and then use statistical or machine learning methods on this data set to build the data fusion rules.

During the open-loop phase of the study, subjects were regularly asked whether they would prefer the next task difficulty to be easier or harder, and their responses were noted. Assuming that the responses were true and accurate, this gave us a training data set with known inputs (performance, biomechanics, and psychophysiology) and known desired outputs (subject's preference). Since there are only two possible outputs (harder/easier difficulty), it is possible to use any of several available classification methods (e.g., discriminant analysis, neural networks, support vector machines, etc.) in order to translate input to output. After testing several classification methods on the open-loop cross-validation data from healthy subjects, we decided to focus on linear discriminant analysis since it outperformed more advanced methods such as support vector machines or neural networks and since some of its variants can be used to address challenges such as intersubject variability and online adaptation (as explained in Sections II-G2 to II-G5).

1) Linear Discriminant Analysis: Originally developed by Fisher [21], linear discriminant analysis (LDA) is a well-known method for feature extraction and classification. It is used to find a linear combination of features which best separate data points into two or more classes. The LDA equation for classification of data into two different classes (class 1 and class 2) can be written as

$$D(\boldsymbol{x}) = b + \boldsymbol{w}^T \cdot \boldsymbol{x} \tag{1}$$

$$b = -\boldsymbol{w}^{\mathrm{T}} \cdot \frac{1}{2} \cdot (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \tag{2}$$

$$\boldsymbol{w} = (\boldsymbol{S}_1 + \boldsymbol{S}_2)^{-1} \cdot (\boldsymbol{\mu}_2 + \boldsymbol{\mu}_1) \tag{3}$$

$$C(\boldsymbol{x}) = \begin{cases} 1; D(\boldsymbol{x}) < 0 \\ 2; D(\boldsymbol{x}) \geq 0 \end{cases} \tag{4}$$

where $\boldsymbol{x}$ is the vector of input features, $D(\boldsymbol{x})$ is the discriminant function, b and $\boldsymbol{w}$ are the weights of $D(\boldsymbol{x})$, $\boldsymbol{S}_k$ is the covariance matrix for class k, $\boldsymbol{\mu_k}$ is the mean value for class k, and $C(\boldsymbol{x})$ is the class to which $\boldsymbol{x}$ is assigned.

In our case, LDA can be used to classify the multiple input features into an estimate of how task difficulty should be changed (easier or harder difficulty). However, there are two problems. First, there are a very large number of available features, some of which may not even be relevant. If all of them are used in discriminant analysis, a very large training set is required to obtain an accurate discriminant function. Two possible solutions are discussed in Sections II-G2 and II-G3.

Second, biomechanical and especially psychophysiological features exhibit high intersubject variability. A discriminant function trained using data from multiple subjects will be generally accurate, but may fail for some subjects. As a subject performs rehabilitation exercises, it would be useful for the biocooperative feedback loop to gradually adapt to that subject and become more accurate. A possible method for this is Kalman adaptive LDA, described in Sections II-G4 and II-G5.

2) Stepwise Linear Discriminant Analysis: Stepwise LDA is a variant of LDA where, instead of all the features being entered simultaneously, a discriminant function is built step-by-step. Starting with no features in the function, at each step all the features are evaluated to determine which one will contribute most to the discrimination between classes. That feature is included in the function, and the process starts again. At each step, a feature already in the function can also be removed if it does not contribute sufficiently to discrimination. The stepwise procedure is usually guided by the statistical F-value of a feature, which indicates its statistical significance in discrimination between classes. The feature selection process ends when no feature has a sufficiently high F-value to be added to the function or a sufficiently low F-value to be removed. This approach has been used for analysis of different types of data, including psychophysiological responses [28].

In our biocooperative feedback loop, we can use stepwise LDA on the training data set to select the most relevant features for task difficulty adaptation. The threshold F-value to add a feature was 3.5 while the threshold F-value to remove a feature was 3. An exception was made if no features exceeded the threshold F-value to add a feature. This often occurred when only psychophysiological data was entered into the stepwise procedure. In this case, both thresholds were lowered in steps of 0.5 until at least one feature's F-value exceeded the threshold.

3) Diagonal Linear Discriminant Analysis: Diagonal LDA is a special case of LDA where all class densities are assumed to have the same diagonal covariance matrix. Thus, the class of an input vector $\boldsymbol{x}$ can be determined as

$$C(\boldsymbol{x}) = \arg\min_k \sum_{i=1}^{G} \frac{(x_i - \mu_{ik})^2}{\sigma_i^2} \tag{5}$$

where μ_{ik} is the mean value of element i for class k, σ_i is the standard deviation of element i, G is the number of elements of $\boldsymbol{x}$, and $C(\boldsymbol{x})$ is the class to which $\boldsymbol{x}$ is assigned.

Though diagonal LDA ignores correlations between input features, it is very effective for classification [29]. In our biocooperative feedback loop, it may be a useful alternative to normal LDA since it usually requires a smaller training set.

4) Kalman Adaptive Linear Discriminant Analysis: Originally developed for electroencephalography [30], Kalman adaptive linear discriminant analysis is a variant of LDA where the discriminant function is initialized using the training data set, then recursively updated online as new information becomes available (in our case, after every 2 min of the task). The update process adjusts the weights of the discriminant function (the contribution of each input feature) using Kalman filtering. For this process, (1)–(4) are expanded with [30]

$$\boldsymbol{H}_k = \left[1, x_{\boldsymbol{k}}^{\mathrm{T}}\right] \tag{6}$$

$$e_k = y_k - \boldsymbol{H}_k \cdot \hat{\boldsymbol{w}}_{k-1} \tag{7}$$

$$Q_k = \boldsymbol{H}_k \cdot \boldsymbol{A}_{k-1} \cdot \boldsymbol{H}_k^{\mathrm{T}} + (1 - \mathrm{UC}) \tag{8}$$

$$\boldsymbol{k}_k = \frac{\boldsymbol{A}_{k-1} \cdot \boldsymbol{H}_k^{\mathrm{T}}}{Q_k} \tag{9}$$

$$\hat{\boldsymbol{w}}_k = \hat{\boldsymbol{w}}_{k-1} + \boldsymbol{k}_k \cdot e_k \tag{10}$$

$$\tilde{\boldsymbol{A}}_k = \boldsymbol{A}_{k-1} - \boldsymbol{k}_k \cdot \boldsymbol{H}_k \cdot \boldsymbol{A}_{k-1} \tag{11}$$

$$\boldsymbol{A}_k = \frac{\mathrm{trace}(\tilde{\boldsymbol{A}}_k) \cdot \mathrm{UC}}{p} + \tilde{\boldsymbol{A}}_k \tag{12}$$

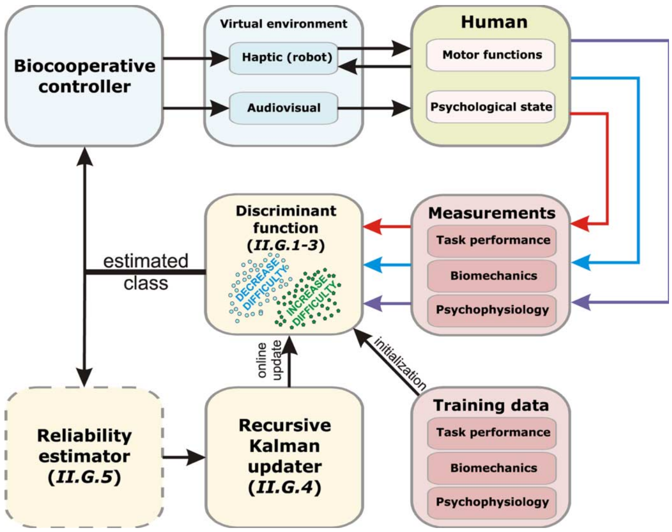

Fig. 3. Our implementation of the biocooperative feedback loop which can adapt to the current subject. Previously recorded and labeled training data is used to initialize the fusion rules, which are then recursively updated (with an optional reliability estimator for unsupervised learning) while the subject is interacting with the virtual environment.

where e_k is the one-step prediction error, y_k is the current class label, $\boldsymbol{x}_k$ is the current input vector, $\hat{\boldsymbol{w}}_k$ is the state vector ($\hat{\boldsymbol{w}}_k = [b, \boldsymbol{w}^T]$, the estimated weights for the LDA), Q_k is the estimated prediction variance, $\boldsymbol{A}_k$ is the *a priori* state error correlation matrix, $\tilde{\boldsymbol{A}}_k$ is an intermediate value needed to compute $\boldsymbol{A}_k$, $\boldsymbol{k}_k$ is the Kalman gain, UC is the update coefficient and p is the number of elements of $\hat{\boldsymbol{w}}_k$. The starting values of $\boldsymbol{A}_0$ and $\hat{\boldsymbol{w}}_0$ as well as the optimal value of UC are computed from the training data set.

In our biocooperative feedback loop, adaptive discriminant analysis can be used as follows. The discriminant function is first initialized with data from the training set. The subject performs the task for 2 min, and then the biocooperative system estimates how difficulty should be changed. Once the estimate has been made, the subject is asked whether he or she would prefer the task to be easier or harder. The system then updates the discriminant function based on the difference between the system's estimate and the subject's response. The function is updated in this way after every task period. In this way, we obtain an adaptive feedback loop which gradually adapts to the current subject (Fig. 3). Adaptive LDA can also be used with either stepwise LDA or diagonal LDA. For stepwise LDA, the stepwise method is used to select the most important input features, and the adaptation only includes those features. For diagonal LDA, the initial weights of the discriminant function are calculated with a diagonal covariance matrix and then entered into the update process.

5) Unsupervised Kalman Adaptive Linear Discriminant Analysis: A limitation of KALDA is that it is a supervised learning method: as can be seen from (7), the subject's actual preference (y_k) is required to update the weights. Since our goal was to see how useful psychophysiological responses could potentially be, we used this supervised KALDA to evaluate whether a biocooperative system can adapt to a particular subject in a "best case" scenario: if given accurate data about that subject. However, a supervised approach is inappropriate in practice—if the subject's preference is available, no automated feedback loop is necessary.

Thus, we have also modified KALDA so that it updates the weights using its own estimate of the subject's preference rather than the subject's actual preference (as seen in Fig. 3). While this makes KALDA unsupervised, it needs to be done carefully since such an approach can also amplify errors. If an incorrect estimate is used to update the discriminant function, the function will become worse. Our method of addressing this was to generate a measure of how "reliable" the estimate is. The system then only updates the discriminant function if the estimate is sufficiently reliable. The reliability criterion was relatively simple. As is evident from (4), the input vector is assigned to one class if $D(\boldsymbol{x}) < 0$ and to the other class if $D(\boldsymbol{x}) \geq 0$. If the absolute value of $D(\boldsymbol{x})$ is very close to zero, the estimate is likely to be unreliable. We considered the estimate to be sufficiently reliable (and updated the weights) if the absolute value of $D(\boldsymbol{x})$ was larger than a certain reliability threshold. The optimal value of this threshold is calculated from the training data set. While the modification used to make KALDA unsupervised is very simple, it allows us to gauge how accurate an adaptive algorithm is likely to be when working with realistically available data.

H. Open-Loop Cross-Validation

Data recorded during the open-loop phase of the study (where task difficulty is adjusted according to the subject's preferences) was evaluated using leave-one-out cross-validation. Discriminant functions were created using data from all subjects except one, then tested on the remaining subject. This was done as many times as there were subjects.

Discriminant functions were judged according to how often their estimate matched the subject's preference regarding task difficulty (easier/harder) in cross-validation. The accuracy rate of a discriminant function was defined as the number of matches divided by the number of all estimates made.

Several different discriminant functions were created. They varied according to the type of input data (performance, biomechanics, psychophysiology, all) and according to the type of discriminant analysis used (normal, stepwise, diagonal, adaptive, adaptive stepwise, adaptive diagonal LDA), for a total of 24 discriminant functions. In all cases, the supervised type of adaptive LDA (as described in Section II-G4) was used. The goal was to see how accurate psychophysiological data would be compared to performance and biomechanical data as well as how much the accuracy rate could be improved using stepwise, diagonal or adaptive methods. After calculating accuracy rates for all three variants of supervised adaptive LDA, the most accurate of the three variants was also used to perform unsupervised adaptation (as described in Section II-G5).

Discriminant functions were first built and cross-validated with data from only healthy subjects, then separately built and cross-validated with data from only hemiparetic subjects. Finally, we also built the discriminant functions using data from all healthy subjects and tested them on data from hemiparetic subjects. This allowed us to see whether information obtained from healthy subjects can be applied to patients. Since stroke patients, for instance, show long-lasting abnormalities in sweating and heart rate variability [31] that have also been noted as a response to rehabilitation tasks [16], we expected that, at the very least, discriminant functions incorporating psychophysiological data could not be directly transferred from healthy subjects to patients.

Additionally, we were interested in knowing which specific combination of features would be most useful in a discriminant function. The five most relevant features were determined for both healthy subjects and hemiparetic patients as the first five features selected by stepwise LDA, whether or not they exceed the F-to-enter threshold chosen in Section II-G2.

I. Closed-Loop Validation

The discriminant function that yielded the highest accuracy rate in open-loop cross-validation was selected for implementation in a closed-loop biocooperative controller. Due to expected differences between healthy subjects and patients, two discriminant functions were trained: one for healthy subjects and one for patients. The functions were trained using data from the open-loop cross-validation phase.

As mentioned in Section II-D, the closed-loop measurement protocol was similar to the open-loop protocol. At the end of each period, the biocooperative controller output whether the task difficulty should be increased or decreased. The output was shown on the screen to the experimenter, but not to the subject. The subject was asked about his or her preference, but task difficulty was changed according to the output of the controller. Accuracy rate was again calculated as the number of matches divided by the number of all estimates made.

The goal of closed-loop testing was not to compare different methods or features; this was done with the larger set of data from the open-loop phase. Instead, the goal was to demonstrate that discriminant analysis can be used online for task difficulty adaptation in a biocooperative feedback loop.

III. Results

A. Open-Loop Cross-Validation

Accuracy rates (in percentages) for open-loop cross-validation of different types of LDA performed on different types of input data are shown in Table I for healthy subjects and Table II for hemiparetic patients. Table III shows accuracy rates for open-loop cross-validation of different types of LDA trained on data from healthy subjects and then tested on patients. In all tables, the highest accuracy rates for nonadaptive and adaptive methods are bolded in each column. The experimenter and subject agreed on whether difficulty should be increased or decreased in 87.6% of all cases for healthy subjects and in 97.0% of all cases for patients.

The unsupervised adaptive approach was tested only on psychophysiological data since supervised adaptive LDA offered little (if any) increase in accuracy over nonadaptive LDA when other data types were used (as seen in Tables I and II). For healthy subjects, unsupervised adaptive diagonal LDA achieved an accuracy rate of 70.8% (compared to 76.4% in the supervised approach and 60.4% in the nonadaptive approach). For patients, unsupervised adaptive LDA achieved an accuracy rate of 65.2%

TABLE I
RATES FOR OPEN-LOOP CROSS-VALIDATION OF DIFFERENT TYPES OF LINEAR DISCRIMINANT ANALYSIS ON DIFFERENT TYPES OF INPUT DATA FROM HEALTHY SUBJECTS

	performance	biomechanics	psychophysiology	all
LDA	**81.9**	**75.0**	56.9	75.7
stepwise LDA	81.9	73.6	56.9	**84.7**
diagonal LDA	80.6	74.3	**60.4**	77.8
adaptive LDA	**82.6**	75.7	71.5	75.7
adaptive stepwise LDA	81.9	73.6	56.9	**84.7**
adaptive diagonal LDA	82.6	**80.6**	**76.4**	83.3

TABLE II
RATES FOR OPEN-LOOP CROSS-VALIDATION OF DIFFERENT TYPES OF LINEAR DISCRIMINANT ANALYSIS ON DIFFERENT TYPES OF PATIENT INPUT DATA

	performance	biomechanics	psychophysiology	all
LDA	**81.8**	**75.8**	54.5	75.8
stepwise LDA	81.8	75.8	**60.6**	**89.4**
diagonal LDA	81.8	71.2	60.6	75.8
adaptive LDA	**81.8**	**75.8**	**68.2**	75.8
adaptive stepwise LDA	81.8	75.8	68.2	**89.4**
adaptive diagonal LDA	81.8	71.2	68.2	76.5

TABLE III
RATES FOR OPEN-LOOP CROSS-VALIDATION OF DIFFERENT TYPES OF LINEAR DISCRIMINANT ANALYSIS ON DIFFERENT TYPES OF INPUT DATA. DISCRIMINANT FUNCTIONS WERE BUILT USING DATA FROM HEALTHY SUBJECTS, THEN TESTED ON PATIENT DATA

	performance	biomechanics	psychophysiology	all
LDA	78.8	63.6	**53.0**	71.2
stepwise LDA	**83.3**	60.6	53.0	**81.8**
diagonal LDA	81.8	**65.2**	50.0	69.7
adaptive LDA	78.8	**68.2**	**66.7**	68.2
adaptive stepwise LDA	**83.3**	65.2	56.0	**81.8**
adaptive diagonal LDA	78.8	63.6	60.6	63.6

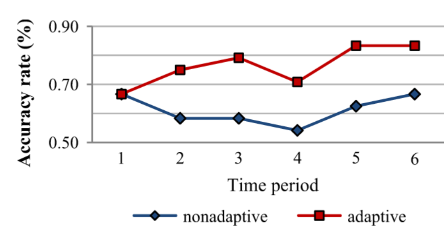

Fig. 4. Accuracy rate as a function of *time period* for open-loop cross-validation of nonadaptive and supervised adaptive diagonal LDA. The inputs are psychophysiological features from healthy subjects.

(compared to 68.2% in the supervised approach and 60.6% in the nonadaptive approach).

As an illustration of how adaptive methods improve accuracy, Fig. 4 shows a comparison of nonadaptive and supervised adaptive diagonal LDA as a function of *time period* when used on psychophysiological data from healthy subjects. Although both nonadaptive and adaptive diagonal LDA yield the same accuracy rate during the first task period, accuracy is higher for the adaptive approach afterwards.

As an illustration of how the size of the training set improves classification accuracy, Fig. 5 shows the accuracy rate of the best nonadaptive method as a function of training set size for different types of data from healthy subjects. Furthermore, Fig. 6 shows the accuracy rate of all three nonadaptive methods as

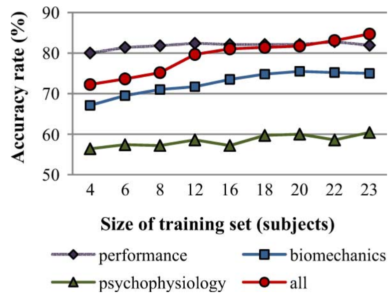

Fig. 5. Accuracy rate as a function of training set size for different types of input data in open-loop cross-validation. Accuracy rate is taken for the best non-adaptive method. All data are from healthy subjects.

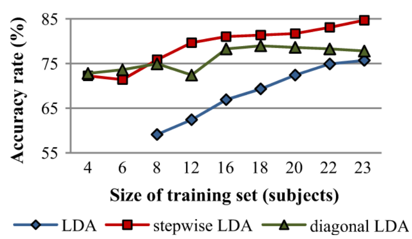

Fig. 6. Accuracy rate as a function of training set size when all input data types are entered in open-loop cross-validation. Results are shown only for nonadaptive methods. All data are from healthy subjects. Note that regular LDA fails for small training sets with more input features than observations.

a function of training set size when all input data types from healthy subjects are used.

In stepwise LDA of data from healthy subjects, the first five sequentially entered features were the *percentage of caught balls* (F to remove $=$ 129.87), *mean SCR amplitude* (F to remove $=$ 3.94), *pNN50* (F to remove $=$ 4.19), *total power in the low-frequency heart rate band* (F to remove $= 2.25$) and *mean derivative of skin temperature* (F to remove $= 2.08$).

In stepwise LDA of data from hemiparetic patients, the first five sequentially entered features were the *percentage of balls placed into the basket* (F to remove $= 100.55$), *standard deviation of respiratory rate* (F to remove $= 3.71$), *total power in the high-frequency heart rate band* (F to remove $= 7.73$), *RMSSD* (F to remove $=$ 5.02) and *final skin temperature* (F to remove $= 2.82$).

B. Closed-Loop Testing

As seen in the open-loop phase, the most accurate type of discriminant function was stepwise (adaptive or nonadaptive) LDA with all data types. Thus, stepwise LDA was chosen for closed-loop testing in both healthy subjects and patients. For healthy subjects, three features were included: the *percentage of caught balls*, *mean SCR amplitude and pNN50*. For patients, four features were included: *percentage of balls placed into the basket*, *standard deviation of respiratory rate, total power in the high-frequency heart rate band* and *RMSSD*. In closed-loop testing, stepwise LDA yielded an accuracy rate of 88.3% for healthy subjects and 88.9% for patients. The experimenter and subject agreed on whether difficulty should be increased or decreased in 91.7% of all cases for healthy subjects and in 97.2% of all cases for patients.

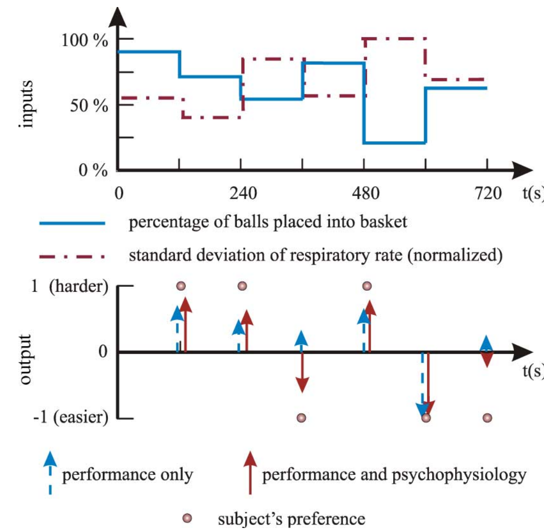

Fig. 7. One hemiparetic patient in the closed-loop phase: two input features (one performance, one psychophysiological), the output ($D(\boldsymbol{x})$), and the subject's preferences. High performance and a low *standard deviation of respiratory rate* (even, regular breathing) indicate an easy task. For the first, second, fourth and fifth task periods, task performance would have been sufficient to change the difficulty. During the third period, task performance is moderately high, but breathing becomes very uneven, indicating stress. If only task performance had been taken into account in this case, the incorrect decision would have been made (the patient was successful at the task, but was stressed and wanted difficulty to decrease). During the last period, both performance and psychophysiology are unreliable, and the patient stated that he would most prefer difficulty to stay the same.

In a follow-up offline analysis, the closed-loop data was also passed through the most accurate discriminant function based only on performance data (also trained using data from the open-loop phase). Performance data yielded an accuracy rate of 86.7% for healthy subjects and 83.3% for patients. For an example of psychophysiology increasing accuracy, see Fig. 7. Additionally, in a second follow-up offline analysis, the closed-loop data was passed through supervised adaptive stepwise LDA. However, the adaptive version yielded the same accuracy rates as the nonadaptive stepwise LDA for both healthy subjects and patients.

IV. Discussion

A. Usefulness of Different Data Types

Task performance was clearly the most accurate type of data in open-loop cross-validation, with an accuracy rate of over 80% for both healthy subjects and patients. Biomechanical data similarly had an accuracy rate of over 75% for both healthy subjects and patients. Psychophysiological measurements, on the other hand, yielded noticeably worse results. Nonadaptive methods yielded an accuracy rate of 60.4% for healthy subjects and 60.6% for patients. Supervised adaptive methods were able to improve the accuracy rate of psychophysiological measurements to 76.4% for healthy subjects and 68.2% for patients, but these results are still worse than results for task performance. This suggests that psychophysiological measurements by themselves are not reliable in a biocooperative feedback loop.

Combining multiple types of data using either LDA or diagonal LDA actually lowers the overall accuracy rate. This is most likely due to the small sample size problem: with a large number of features (26 in total) and a limited training set, it is difficult to find an accurate discriminant function. This is especially noticeable in Figs. 5 and 6, where the accuracy rate rises steadily as the size of the training set increases. Stepwise LDA appears to be the most robust method with regard to sample size. In open-loop cross-validation, combining multiple data types using stepwise LDA increases the accuracy rate from 82.6% (performance data only) to 84.7% for healthy subjects and from 81.8% to 89.4% for patients. In closed-loop validation, combining multiple types of data increases the accuracy rate from 86.7% (performance data only) to 88.3% for healthy subjects and from 83.3% to 88.9% for patients. While the stepwise approach identifies task performance as the most important source of data, several psychophysiological features are also entered into the stepwise discriminant function, suggesting that they can provide some supplementary information.

The question here is whether the increase in accuracy rate due to psychophysiology is sufficient to justify the increased complexity of the system. If measures of task performance are readily available and relevant, psychophysiological measurements are most likely unnecessary. Designers could take this into account by creating virtual environments in which performance is easy to quantify, although this may be difficult to achieve in non-game scenarios such as activities of daily living. In such cases, psychophysiology could prove useful since task performance measures are often not obtainable or not connected to the subject's psychological state. It could also be used to change elements other than the difficulty of the task—for instance, to change the visual appearance of a scenario or to select the music played.

Of course, an accuracy rate of 100% is most likely unrealistic. In a number of cases, subjects were uncertain how they wanted the difficulty to change (if at all), and responded with comments such as "I don't know, either is fine." In such a case, the best choice may have been not to change the task difficulty at all. During the closed-loop phase, it was observed (though only on a subjective, qualitative level) that the output of the discriminant function ($D(\boldsymbol{x})$ in (1) tended to be closer to zero in such cases as well, suggesting that the output of the discriminant function was also "uncertain" in a way.

The reliability of the subject's opinion was also taken into account by comparing the subject's opinion to the experimenter's opinion. These matched in over 90% of cases, with most disagreements being due to either the subject wanting to try a difficulty level that he/she had never encountered before or the subject being tired despite doing well. Thus, the relatively poor accuracy of psychophysiological measurements cannot be (only) due to subjects' inaccurate opinions.

Finally, a word on biomechanical measurements: as reported in Section III-A, the first five features in stepwise LDA include only task performance and psychophysiology. This does not mean that biomechanical measurements are useless. Before any features are included, the F-value (criterion for inclusion) of biomechanical features is higher than that of psychophysiological features. However, once the first feature (a task performance feature for both healthy subjects and patients) has been taken into account, biomechanical features offer less additional information than psychophysiological ones. This again suggests that psychophysiological measurements offer information that cannot be obtained from forces and movements.

B. Adaptive Linear Discriminant Analysis

In open-loop cross-validation, supervised adaptive LDA offers practically no improvement over nonadaptive LDA in the case of performance features (accuracy rate increases from 81.9% to 82.6% for healthy subjects, but not for patients) and only slight improvement in the case of biomechanical features (accuracy rate increases from 75.0% to 80.6% for healthy subjects, but not for patients). In the case of psychophysiological features, however, supervised adaptive LDA increases the accuracy rate from 60.4% to 76.4% for healthy subjects and from 60.6% to 68.8% for patients. Unsupervised adaptive LDA also increases the accuracy rate, though to a lesser degree.

It is currently uncertain why the improvement is greater for psychophysiological features than for other features, though we believe that the reasons are the high intersubject variability of psychophysiological features and the low initial accuracy. In any case, our results show that the system can gradually adapt itself to a given subject to some degree. Since rehabilitation is usually a long-term process, it would be interesting to see what kind of improvement adaptive methods could provide over multiple sessions.

In the supervised adaptive LDA, we provided the system with the subject's preference so that it could adapt the discriminant function with accurate information. Since this information is generally unavailable, we also demonstrated an unsupervised version where the discriminant function is adapted online using the system's own estimate of the subject's preference. Though our modification is probably not the optimal unsupervised adaptive LDA, it is a possible practical implementation of adaptive LDA. We also foresee two other possibilities.

In one alternative implementation of adaptive LDA, the patient's first session with the system is a supervised session where the patient regularly inputs his or her preference into the system, enabling accurate adaptation. In later sessions, the adaptation is turned off. Thus, the system uses the first session to adapt to the patient to some degree, and this information is incorporated into the system during later sessions.

In a second alternative implementation of adaptive LDA, the discriminant function would not be adapted on its own, but the subject could manually input his or her own preference at any time. The system would then not only change the difficulty of the task, but also update its discriminant function with the subject's input. Another possibility would be for the system to explicitly ask the subject for input if certain potentially erroneous trends are detected (e.g., if the system repeatedly estimates that the task is too easy even though the subject has reached a very high difficulty level).

C. Differences Between Healthy Subjects and Patients

Based on previous studies that have shown weakened psychophysiological responses as a result of stroke and other pathological conditions [16], [31], we expected that fusion of psychophysiological measurements would be less accurate in patients than in healthy subjects. However, this does not appear to be the case; the accuracy rate for nonadaptive methods is similar in healthy subjects and patients. Interestingly, accuracy rates are similar for both groups even though the patient group is much smaller.

As Table III shows, discriminant functions based on biomechanical or psychophysiological measurements cannot be transferred from healthy subjects to patients without a noticeable decrease in accuracy rate. Stepwise LDA also selects different features in healthy subjects and patients.

It is easy to understand why results of biomechanical measurements are different between groups: hemiparetic patients, by definition, cannot move their affected limb as well as healthy subjects can. This was evident, for instance, in their response to high difficulty levels. While all healthy subjects reacted to very fast balls by rapidly moving around the virtual table trying to catch the ball, many patients preferred to simply stay in one area of the table and catch only the balls passing through that area. Psychophysiological measurements are, to some degree, obviously different due to the aforementioned effects of stroke and other pathological conditions. Additionally, it is possible that, for patients, higher task difficulty levels are also physically demanding and thus evoke stronger physiological responses. It has been previously shown that, during haptic interaction, heart rate and skin conductance are affected by both cognitive and physical workload [14]. Thus, it is probable that the physiological responses in our study convey not only psychological information, but also information about physical activity.

D. Study Limitations

In the course of our study, a few limitations became apparent. First of all, we used only four psychophysiological signals (heart rate, skin conductance, respiration, and skin temperature). Data gleaned from these signals may not paint a complete picture of the patient's psychophysiological state. Additional signals such as facial electromyography or eye movements may make a biocooperative feedback loop more accurate since they have been successfully used together with autonomic nervous system responses in psychophysiology (e.g., [23], [32]). Furthermore, by using LDA, we assumed that connections between the measured features and subjects' preferences were linear. As previously mentioned, LDA was chosen over other methods for its good accuracy rate on the open-loop cross-validation data. However, if additional physiological signals and/or a larger sample were available, nonlinear methods may result in greater accuracy. Artificial neural networks, for instance, could be a useful alternative that has been tested with psychophysiological data in other settings [28]. In our case, however, neural networks tested on the open-loop data yielded a slightly lower accuracy than LDA, possibly due to the small sample size.

Another limitation of the study is that subjects were only given two choices: to "prefer easier" or to "prefer harder" task difficulty. Obviously, it is possible that a subject finds the difficulty to be 'just right' and does not wish to change it. A follow-up study would be useful to see how accurate a system would be if it also had the option not to change the task difficulty. For this, it may be necessary to define a third class in discriminant analysis, but it may also be enough to expand (4) with a simple rule: do not change the task difficulty if $D(\boldsymbol{x})$ is sufficiently close to zero (as seen in Fig. 7). Although we used a threshold to define two possible outputs of the discriminant function, the function by itself outputs a continuous value, and certain ranges of the output could correspond to different changes in task difficulty.

A final limitation is the choice of rehabilitation task. Since few rehabilitation tasks have been studied from a psychophysiological perspective, we chose to build on a task that has already been used in previous psychophysiological work [16]. However, one component of the task (placing the ball in the basket) does not depend on the difficulty level since the difficulty level only affects the size and speed of the ball. Psychophysiological differences between difficulty levels thus may not have been as large as they would have been if all task components had been affected by the difficulty level, and this may have contributed to the limited usefulness of psychophysiological measurements. Future psychophysiological studies may prefer to focus on a task with only a single component (e.g., only horizontal reaching).

V. Conclusion

The four psychophysiological responses evaluated in our study are not very accurate when used on their own, although adaptive methods that adapt to each individual subject can improve their accuracy. Psychophysiological responses can be used as a supplementary source of information in combination with measurements such as task performance and biomechanics, although it is uncertain whether they provide enough additional information to justify the increased cost and complexity of the system. They may also be a useful source of information in tasks and environments where task performance or biomechanical measurements are either not available or are not at all connected to the subject's mood.

Acknowledgment

The authors would like to thank the volunteers who participated in this study. Special thanks go to N. Goljar, MD and M. Javh, OT, of the University Rehabilitation Institute of the Republic of Slovenia for their assistance. Moog FCS loaned one of the HapticMasters used in our research.

References

[1] G. B. Prange, M. J. Jannink, C. G. Groothuis-Oudshoorn, and H. J. Hermens, "Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke," *J. Rehabil. Res. Develop.*, vol. 43, pp. 171–184, Mar. 2006.

[2] A. C. Lo *et al.*, "Robot-assisted therapy for long-term upper-limb impairment after stroke," *New Eng. J. Med.*, vol. 362, pp. 1772–1783, May 2010.

[3] W. S. Harwin, J. L. Patton, and V. R. Edgerton, "Challenges and opportunities for robot-mediated neurorehabilitation," *Proc. IEEE*, vol. 94, no. 9, pp. 1717–1726, Sep. 2006.

[4] M. K. Holden, "Virtual environments for motor rehabilitation: Review," *CyberPsychol. Behav.*, vol. 8, pp. 187–211, Jun. 2005.

[5] R. Riener, L. Lünenburger, S. Jezernik, M. Anderschitz, G. Colombo, and V. Dietz, "Patient-cooperative strategies for robot-aided treadmill training: First experimental results," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 380–394, Sep. 2005.

[6] E. T. Wolbrecht, V. Chan, D. J. Reinkensmeyer, and J. E. Bobrow, "Optimizing compliant, model-based robotic assistance to promote neurorehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 3, pp. 286–297, Jun. 2008.

[7] G. Kwakkel *et al.*, "Effects of augmented exercise therapy time after stroke: A meta-analysis," *Stroke*, vol. 35, pp. 2529–2539, Nov. 2004.

[8] I. Robertson and J. Murre, "Rehabilitation of brain damage: brain plasticity and principles of guided recovery," *Psychol. Bull.*, vol. 125, pp. 544–575, Sep. 1999.

[9] N. Maclean, P. Pound, C. Wulfe, and A. Rudd, "The concept of patient motivation: A qualitative analysis of stroke professionals' attitudes," *Stroke*, vol. 33, pp. 444–448, Feb. 2002.

[10] M. Munih *et al.*, "MIMICS: Multimodal immersive motion rehabilitation of upper and lower extremities by exploiting biocooperation principles," in *IEEE Conf. Rehabil. Robotics*, Kyoto, Japan, 2009, pp. 127–132.

[11] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, pp. 394–421, Jul. 2010.

[12] A. Haarmann, W. Boucsein, and F. Schaefer, "Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight," *Appl. Ergon.*, vol. 40, pp. 1026–1040, Nov. 2009.

[13] P. Rani, N. Sarkar, C. A. Smith, and L. D. Kirby, "Anxiety detecting robotic system—Towards implicit human-robot collaboration," *Robotica*, vol. 22, pp. 85–95, Jan. 2004.

[14] D. Novak, M. Mihelj, and M. Munih, "Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction," *Robotica*, vol. 29, pp. 367–374, May 2011.

[15] M. S. Cameirão, S. B. Badia, K. Mayank, C. Guger, and P. F. Verschure, "Physiological responses during performance within a virtual scenario for the rehabilitation of motor deficits," in *Proc. PRESENCE*, Barcelona, Spain, 2007, pp. 85–88.

[16] D. Novak *et al.*, "Psychophysiological responses to robotic rehabilitation tasks in stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 4, pp. 351–361, Aug. 2010.

[17] A. C. Koenig *et al.*, "Model-based heart rate prediction during lokomat walking," in *Proc. 31st Annu. Conf. IEEE Eng. Med. Biol. Conf.*, Minneapolis, MN, 2009, pp. 1758–1761.

[18] C. R. Guerrero, J. F. Marinero, J. P. Turiel, and P. R. Farina, "Bio cooperative robotic platform for motor function recovery of the upper limb after stroke," in *Proc. 32nd Annu. IEEE Eng. Med. Biol. Conf.*, Buenos Aires, 2010, pp. 4472–4475.

[19] J. T. Cacioppo and L. G. Tassinary, "Inferring psychological significance from physiological signals," *Am. Psychol.*, vol. 45, pp. 16–28, Jan. 1990.

[20] S. H. Fairclough, "Fundamentals of physiological computing," *Interact. Comput.*, vol. 21, pp. 133–145, Jan. 2009.

[21] R. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, pp. 179–188, 1936.

[22] R. Q. Van der Linde and P. Lammertse, "HapticMaster—A generic force controlled robot for human interaction," *Ind. Robot*, vol. 30, pp. 515–524, Oct. 2003.

[23] R. L. Mandryk and S. M. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *Int. J. Hum. Comput. Stud.*, vol. 65, pp. 329–347, Apr. 2007.

[24] A. Schwerdtfeger, "Predicting autonomic reactivity to public speaking: Don't get fixed on self-report data!," *Int. J. Psychophysiol.*, vol. 52, pp. 217–224, May 2004.

[25] D. Kidd, G. Stewart, J. Baldry, J. Johnson, D. Rossiter, A. Petruckevitch, and A. J. Thompson, "The functional independence measure: A comparative validity and reliability study," *Disabil. Rehabil.*, vol. 17, pp. 10–14, Jan. 1995.

[26] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.*, vol. 12, pp. 189–198, Nov. 1975.

[27] "Task force of the European society of cardiology and the North American society of pacing and electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, pp. 354–381, Mar. 1996.

[28] G. F. Wilson and C. A. Russell, "Operator functional state classification using multiple psychophysiological features in an air traffic control task," *Hum. Factors*, vol. 45, pp. 381–389, Jan. 2003.

[29] S. Dudoit, J. Fridlyand, and T. P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," *J. Amer. Stat. Assoc.*, vol. 97, pp. 77–87, Mar. 2002.

[30] C. Vidaurre, A. A. Schlögl, 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. 550–556, Mar. 2007.

[31] J. T. Korpelainen, K. A. Sotaniemi, and V. V. Myllylä, "Autonomic nervous system disorders in stroke," *Clin. Auton. Res.*, vol. 9, pp. 325–333, Dec. 1999.

[32] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," *Int. J. Hum. Comput. Stud.*, vol. 65, pp. 724–736, Aug. 2007.

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.

Matjaz Mihelj received the Ph.D. degree from the Faculty of Electrical Engineering, University of Ljubljana, Ljubljana, Slovenia, in 2002.

He was a visiting researcher at Tohoku University, Japan, and at ETH Zurich, Switzerland. Since 2009, he has been an Associate Professor at the Faculty of Electrical Engineering, University of Ljubljana. His research interests include man–machine interaction, haptic interfaces, and medical robotics.

Jaka Ziherl 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 haptic interfaces, rehabilitation robotics, and virtual reality.

Andrej Olenšek graduated from the University of Ljubljana, Ljubljana, Slovenia, in 2003 and defended his Ph.D. thesis in electrical engineering at the same university in 2008.

His primary research interest is multimodal human–robot interaction for rehabilitation purposes.

Marko Munih (M'88) received the Ph.D. degree in electrical engineering from the University of Ljubljana, Ljubljana, Slovenia.

His research interests include functional electrical stimulation, measurement and control, biomechanics, and robotics. In 1995–1996, he was a visiting scientist at University College London. He is now Full Professor and Head of the Laboratory for Robotics and Biomedical Engineering at the University of Ljubljana.