Adaptation of Task Difficulty in Rehabilitation Exercises Based on the User's Motor Performance and Physiological Responses

Navid Shirzad, H. F. Machiel Van der Loos
Department of Mechanical Engineering
The University of British Columbia
Vancouver, Canada
n.shirzad@alumni.ubc.ca, vdl@mech.ubc.ca

Abstract—**Although robot-assisted rehabilitation regimens are as effective, functionally, as conventional therapies, they still lack features to increase patients' engagement in the regimen. Providing rehabilitation tasks at a "desirable difficulty" is one of the ways to address this issue and increase the motivation of a patient to continue with the therapy program. Then the problem is to design a system that is capable of estimating the user's desirable difficulty, and ultimately, modifying the task based on this prediction. In this paper we compared the performance of three machine learning algorithms in predicting a user's desirable difficulty during a typical reaching motion rehabilitation task. Different levels of error amplification were used as different levels of task difficulty. We explored the usefulness of using participants' motor performance and physiological signals during the reaching task in prediction of their desirable difficulties. Results showed that a Neural Network approach gives higher prediction accuracy in comparison with models based on k-Nearest Neighbor and Discriminant Analysis methods.**

Keywords—robotic reaching exercise; error amplification; desirable difficulty; machine learning; motor performance; physiological signals; stroke therapy

I. INTRODUCTION

Recent findings of Neuroscience indicate that to stimulate brain plasticity and acquisition of new motor skills, intensive training and high dose of repetition are required [1]. The field of robot-aided stroke therapy has steadily progressed in the past two decades by incorporating this principle. Studies with different upper extremity rehabilitation robots such as MIT-Manus [2], Mirror Image Movement Enabler (MIME) [3], and GENTLE/s [4] have validated the positive effects of training with such devices on motor recovery and functional improvement. Studies with wrist rehabilitation robots such as InMotion3 [5] and RiceWrist [6] resulted in similar outcomes. Additionally, in robotic therapy interventions, the rehabilitation process and the patient's progress can be quantified. This suggests the potential of these robotic devices to be utilized as a mainstream motor rehabilitation tool.

Research supported by the National Science and Engineering Council of Canada (NSERC) and the Peter Wall Solutions Initiative, Vancouver, BC.

To maximize the effect of repetitive training, research in both artificial intelligence and motor learning proposes an error-driven process that supports learning [7, 8]. Combining visual and haptic error amplification with repetitive training improves the rate of motor adaptation in both healthy [9] and stroke populations [10].

While therapy is based on motor learning principles and relies on the physical, strengthening effect of exercise, a patient's motivation is also a key factor in the success of a therapy regimen to maintain engagement in the exercise for a longer period [11]. The issue of sustaining engagement has been theorized in the fields of game design and psychology. From a motor learning point of view, to avoid boredom or frustration, one needs to be kept at the upper limit of one's ability by meaningful manipulation of exercise difficulty (i.e., exercising at one's challenge point [12] or desirable difficulty). This desirable difficulty can be dependent on both task performance and a person's affective state. While quantifying task performance in robotic therapy regimens is relatively easy, measuring affect is more challenging. Physiological signals processing to gauge affect has gained popularity in recent years. Efficiency of several machine learning methods in predicting affect based on physiological signals is presented in [13]. In this study, Rani et al. used a human-robot interaction task to elicit different affective states.

Liu et al. [14-15] have studied the relation between affective state and physiological responses. In separate studies, they propose methods for prediction of a participant's affective state by using physiological responses and artificial intelligence. They showed that modifying tasks based on participant affect (e.g., anxiety and liking) leads to higher performance. Wang et al. [16], using a performance-based assist-as-needed robotic training, have reported higher overall performance for participants training with this adaptive exercise.

In a former study [9], we showed that doing a reaching exercise with different levels of error amplification leads to different levels of motor adaptation and affective states. Moreover, participants reported different levels of perceived difficulty for each error amplification level. Building on these findings and to investigate the potential of predicting a user's challenge point, in this paper we present a comparison of three machine learning approaches – neural networks, k-nearest neighbor and discriminant analysis – in predicting participants' desirable difficulty in a robotic reaching task. We applied these to three sets of participant data: motor performance, physiological signals, and a hybrid of both measures. Results of this study can be used in future work to design a closed-loop control system that is capable of predicting the user's challenge point to provide a more engaging exercise experience.

II. Methods

A. Research Ethics

This study was approved by the Clinical Ethics Research Board of the University of British Columbia, and all participants provided written informed consent.

B. Study Participants

Twenty-four healthy adult (average age 23.8) subjects participated in this study; the male/female ratio was 12/12. In order to confirm participants were free of neurological impairment, a minimum score of 24 in Folstein Mini-Mental Test was required. In addition, only participants with normal or corrected eyesight were recruited to make sure changes in performance were not attributable to anything but changes in task difficulty.

C. Hardware

A custom-built five-bar robot was used for this study [17]. The robot's end-effector moves in the horizontal plane in an area of approximately 50 cm × 35 cm. A handle instrumented with a 6-axis force/torque sensor (ATI Mini) measures force inputs by users, while readings from encoders at the base joints are used to calculate end-effector position. End-effector position is projected on a flat screen monitor as a moving dot using TargetDisplay software (MathWork Inc.). The two base joints are actuated by two motors (Parker-Compumotor Dynaserv DR1060B) to provide haptic feedback, and the elbow joints are passive.

Physiological signals are captured at 256 Hz using a ProCompInfiniti Physiology Suite (Thought Technology Inc.). In this study we recorded three physiological signals: skin temperature (Temp), skin conductance response (SCR), and respiration rate (Resp). Temperature was measured using a sensor strapped around the distal phalange of the ring finger of participant's dominant hand. SCR was recorded using two electrodes strapped around the distal phalanges of the index and middle fingers of the dominant hand. The breathing rate sensor was placed on a strap around the participant's chest.

Fig. 1 shows a participant wearing the SCR sensor while sitting in front of the robot (left) and a schematic top view of the robot's working area (right).

D. Rehabilitation Task, Distortion, and Error Amplification

The robot was used to exercise reaching to targets in the horizontal plane as a motor learning task. Targets were visually presented to participants on a monitor and they were instructed to use the robot with their non-dominant hand to move a cursor to these visual targets. Participants were told the cursor represents robot's end-effector position. The participants' hands were concealed with a cover in order to ensure visual feedback is only provided via the monitor.

Three targets were placed radially, with the same distance from the middle of the monitor and the other targets. When a target was presented, participants had to move the robot's handle to place the moving dot over the target. Once the target was reached, they had to move the cursor back to the center of the screen. We use the term "cycle" for the consecutive reaching motions (i.e., trials) to the three targets (in a random order).

To cause an initial error and a decrease in performance, and thus to initiate motor learning and motor adaptation, a visual distortion was implemented during training blocks (see Section E below). This visual distortion was implemented as a 30° rotation between the end-effector coordinates (i.e., actual hand position) and the monitor coordinates (i.e., visual target). Participants practiced with five different error amplification (EA) levels to learn reaching within this visual distortion.

The five EA levels used were (ordered from the easiest perceived difficulty by participants to the hardest [9]): control (no EA), low gain visual EA, high gain visual EA, low gain haptic/visual EA, high gain haptic/visual EA. Details of the EA implementation can be found in [9].

E. Study Protocol

After providing consent to participate in the study, participants were introduced to the robotic manipulandum and the reaching task. Subjects were then seated in front of the robot and physiological sensors were put on them. Each experiment, on average, took 80 minutes. Participants were given rest times (in addition to the existing rest periods) upon request. They were told that they could stop participating in the experiment at any point.

The experiment consisted of six exercise blocks. In the first block, participants practiced reaching in the virtual environment to get familiar with the task. In this block (14 cycles) both visual distortion and error amplification were turned off (i.e., plain motion). The familiarization block was followed by five training blocks (exercise blocks 2 to 6).

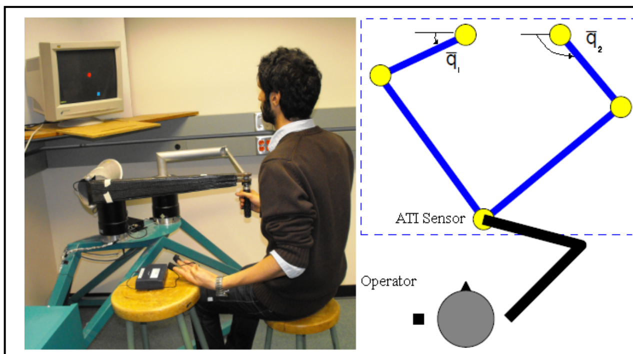

Fig. 1. Left: a participant holds robot's handle while SCR data is being recorded from his left hand's fingertips. Right: schematic top view of the workspace.

Each training block was divided into three sub-blocks: de-adaptation, rest period, challenge. The de-adaptation sub-blocks consisted of 10 cycles of plain motion. This was to ensure training effects (i.e., adaptation to visual distortion) from previous challenges did not carry over to the following challenge. In the rest period, participants were asked to remain seated and with the least possible motion for one minute. Physiological signals recorded in this period were used as a baseline in computing physiological measures. Finally, the challenge sub-block consisted of reaching to visual targets within the visually distorted environment with one of the five EA levels (13 cycles). The order of appearance of the five EA levels throughout the training blocks (blocks 2 to 6) was random and predefined for each participant.

At the end of each of the training blocks (after finishing challenge sub-blocks), participants were asked to report whether they wanted their next trial to be easier (recorded as −1) or harder (+1), in other words, whether their "desirable difficulty" was easier or harder than the challenge they just practiced. To make the question clear an example was given: "Assume that you want to play chess. If you play with a five-year-old, you will be winning all the time without feeling any challenge, and you will get bored. If you play with the world champion, you will be losing all the time without any hope for a win and that will make you frustrated. But, if you play with an average player the game will be engaging to you and you will enjoy the game. We call that desirable difficulty."

F. *Motor Performance Features*

Motor performance is measured by motor adaptation and accuracy. As mentioned, a cycle consists of consecutive reaching trial to the three target points. For each cycle, we define the performance measure as the highest value between maximum errors of each of the three trials. The maximum error of each trial is calculated as the maximum deviation of the reaching trajectory from the line between the start point and the target point (i.e., the error vector that is used for EA). Fig. 2 shows a plot of these maximum errors over the course of exercise block number 2 (first training block) for participant 7. In this challenge sub-block, the participant trained with low gain haptic/visual gain.

As shown in Fig. 2 (right side), at the beginning of exposure to visual distortion, performance decreases (i.e., large initial error). With more practice, the participant adapts to the distortion and learns to perform reaching within the distorted environment with lower error. Adaptation can be modeled with an exponential decay function of this form [9]:

$$y = ae^{-x/b} + c \qquad (1)$$

In (1), y is the maximum error in a cycle, x is the cycle number, a is the total decrease in the maximum error (i.e., the difference between the maximum error in the first cycle and the last cycle), b is the time constant and represents the speed of adaptation, and finally c can be interpreted as the final maximum error after many cycles of practice (i.e., maximum error will eventually converge to c). For each challenge sub-block, a set of a, b, and c is calculated and used as three features of motor performance in predicting the participant's desirable difficulty.

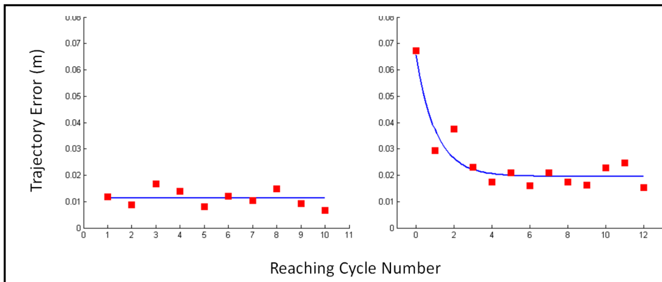

Fig. 2. Left: following the familiarization block and during the first de-adaptation sub-block, the participant shows little variation in maximum trajectory error (about ±1mm). Right: During practice with challenge, as the participant adapts to the visual distortion, error decays.

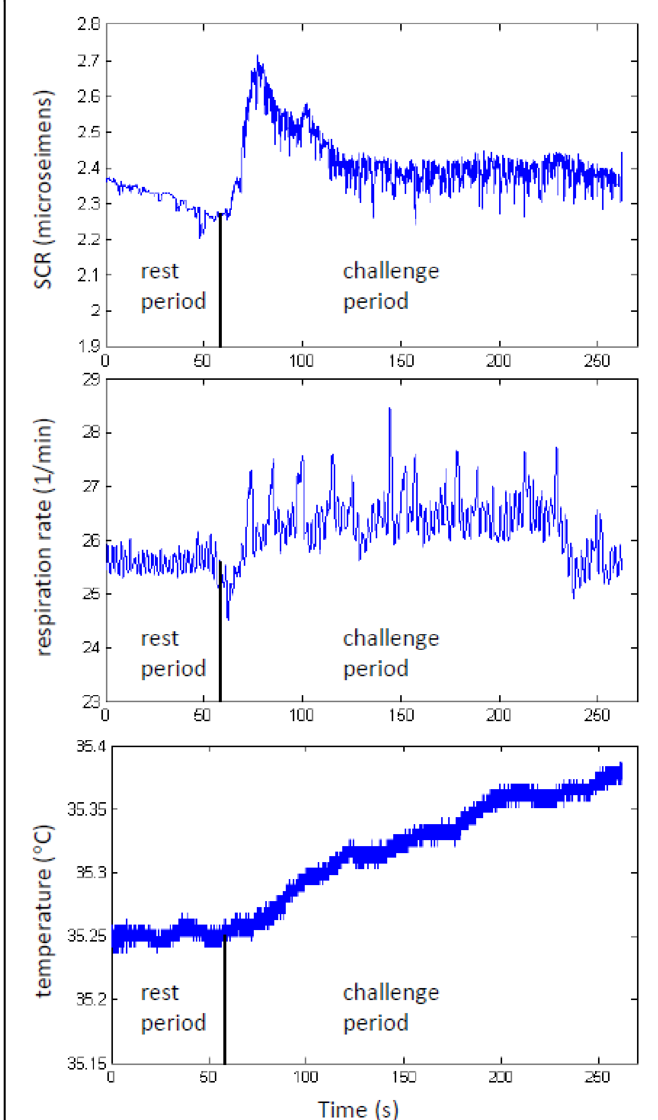

Fig. 3. Raw physiological responses recorded from participant 7 during the first training block. Top: raw SCR signal. Middle: raw respiration rate signal. Bottom: raw skin temperature signal.

G. *Physiological Features*

During each rest period and challenge sub-block, three physiological signals were recorded: skin temperature, respiration rate, and skin conductance rate (SCR). From these raw signals, 11 physiological features were extracted. Fig. 3 shows a plot of these raw signals over the course of exercise block number 2 (first training block) for participant 7. In this challenge sub-block, the participant trained with low gain haptic/visual gain.

From the SCR signal, 5 features are extracted: **normalized SCR**, defined as mean SCR in the challenge period divided by mean SCR in the rest period, **SCR peak value**, defined as the difference between minimum SCR and maximum SCR during the challenge period, **normalized derivative of SCR (dSCR)**, defined as mean dSCR in the challenge period divided by mean dSCR in the rest period, **SCR change**, defined as the difference between the mean SCR in rest period and mean SCR during the challenge period, and **dSCR variability**, defined as the standard deviation of dSCR in the challenge period.

From the respiration rate (Resp) signal, 3 features are extracted: **normalized respiration rate,** defined as mean Resp in the challenge period divided by mean Resp in the rest period, **respiration rate change,** defined as the subtraction of mean Resp during the rest period from mean Resp during challenge period, and **respiration rate variability**, defined as the standard deviation of respiration rate in challenge period.

From the skin temperature (Temp) signal, 3 features are extracted: **normalized skin temperature,** defined as mean Temp in the challenge period divided by mean Temp in the rest period, **skin temperature change**, defined as the subtraction of mean Temp during the rest period from mean Temp during the challenge period, and **skin temperature variability**, defined as the standard deviation of skin temperature in the challenge period.

H. Applied Machine Learning Algorithms

The goal of this study was to investigate the potential of predicting reported desirable difficulty of the reaching task (easier or harder) based on motor performance and physiological features. This was framed as a classification problem with desirable difficulty as the target function and motor performance and physiological features as the predictor variables.

In this study we compared the accuracy of three machine learning algorithms in this classification problem by using three sets of predictor variables. The machine learning algorithms are k-Nearest Neighbor, Neural Network, and Discriminant Analysis. The three sets of predictor variables are the motor performance features (3 attributes), physiological features (11 attributes), and a hybrid of performance and physiological features (14 attributes). Comparing the results of prediction based on each of these predictor variable sets informs us about the usefulness of including or excluding each of the feature sets from the prediction process.

III. RESULTS

A. Training Machine Learning Algorithms

From the human subject study (24 participants) we collected a dataset with a total of 107 instances. Each instance consists of 14 numeric values for the predicting variables (i.e., motor performance and physiological features) and one self-report for desirable difficulty (−1 for easier and +1 for harder). Three prediction models (one for each prediction variable set) were trained per machine learning algorithm (3×3=9 prediction models) using 80 of the 107 instances i.e., 75% of the dataset. The other 27 instances were used for validation of the models and comparison of their prediction accuracy.

To train the k-Nearest Neighbor models, we used k = 7 neighbors with Standardized Euclidean distances and equal distance weighting. After experimenting with different flavors of Discriminant Analysis methods, Diagonal Quadratic Discriminant Analysis yielded the best results. Finally, we used Neural Networks with one hidden layer.

After training each of the 9 prediction models, we fed it the same 80-instance prediction variable set that was used for training. The number of correct predictions calculated from this process was used in defining the algorithm's ability to learn the training set. This was defined as a percentage and calculated as the ratio of the number of correct predictions divided by 80 (Table I).

B. Validation of the Trained Models

To validate the trained models and investigate their ability to generalize, the remaining 27 instances (25% of the available data points) were fed into each of the 9 models and their prediction accuracies were calculated (hold-out cross-validation). Results of this process are presented in Table II.

To provide a common ground in comparing prediction accuracies, we used a random number generator as the baseline predictor model. We generated 20 sets of 107 randomly-generated (−1, +1) pairs and used them as the output of the predictor for desirable difficulties. On average, the accuracy of randomly guessing a participant's desirable difficulty was found to be 49%, close to the expected value of 50%.

TABLE I. ABILITY TO LEARN THE TRAINING SETS

Machine Learning Algorithm	**Prediction Variable Set**		
	Performance features	***Physiological features***	***Hybrid of all features***
k-Nearest Neighbor	54/80 = 68%	50/80 = 63%	52/80 = 65%
Neural Network	62/80 = 78%	48/80 = 60%	76/80 = 95%
Discriminant Analysis	58/80 = 73%	47/80 = 59%	56/80 = 70%

TABLE II. PREDICTION ACCURACY OF NEW INSTANCES

Machine Learning Algorithm	**Prediction Variable Set**		
	Performance features	***Physiological features***	***Hybrid of all features***
k-Nearest Neighbor	21/27 = 78%	15/27 = 56%	18/27 = 67%
Neural Network	21/27 = 78%	16/27 = 59%	19/27 = 70%
Discriminant Analysis	20/27 = 74%	16/27 = 59%	18/27 = 67%
Random Number Generator	49%[a]		

[a.] Average of 20 trials using all 107 instances

IV. Discussion

k-Nearest Neighbor (k-NN) is a simple non-parametric classification method that searches for similarities between the input data and the existing training data. However, this method showed the lowest ability to learn the training set using two of the variable sets (i.e., performance and hybrid feature sets). On the other hand, it showed comparable results when it was tested with the validation dataset. Neural Network (NN) is a rather computationally intensive parametric machine learning method. It showed the highest rate in learning the training dataset, which resulted in better generalization of the data and thus better performance in the validation process. Discriminant Analysis (DA), probably the most computationally expensive model between the three methods, showed slightly lower performance compared to NN. Although the three methods showed different levels of ability to learn the dataset, they showed competitively close accuracies in predicting desired difficulties based on new data. NN, with a moderate level of computational expense and followed by DA and k-NN, had the best accuracies in learning the full (hybrid) training dataset and predicting desirable difficulties in the validation phase.

To evaluate the trained models and assess their ability to generalize to new input data, we used the holdout cross-validation (HCV) method. Leave-one-out cross-validation (LOOCV) is more common in the evaluation of machine learning methods used for affect recognition [13]. However, given the dataset size, use of LOOCV would be more computationally expensive compared to use of holdout cross-validation (training 107 models vs. 1 model). In HCV, each data point is used either for training or validation, whereas in LOOCV data points are used for both training and validation. This means that by use of HCV, performance of one model is assessed by several new data points. In comparison to LOOCV where average performance of several models over several validation datasets is used in the cross-validation process, HCV gives a more objective estimate of accuracy and generalizability.

To assess the usefulness of the three different prediction variable sets, results presented in Table 2 were compared column by column. Prediction based on motor performance features has the highest accuracy rate (78%) among all three machine learning methods and it is followed by prediction based on a hybrid of all the features (70%) and then physiological features (59%). Lower accuracy of prediction based on physiological features can be attributed to the noisy nature of these features (compared to motor performance features). This noticeable superiority of prediction based on motor performance suggests the possibility of simplifying the robotic therapy system by removing the physiological sensors and avoiding the physiological signal analysis altogether.

However, given the dependency of the challenge point on the affective state of the user [12-14], the authors believe that physiological signals can provide supplementary information in predicting desirable difficulty in rehabilitation tasks. In future, we aim to explore additional ways of combining motor performance and physiological features to achieve higher accuracy rates in predicting each user's desirable difficulty level.

V. Conclusion

To investigate the possibility of estimating a user's desirable difficulty during rehabilitation exercise, we presented the accuracy of three commonly used machine learning algorithms in a comparative study. Data from a human subject study with 24 healthy participants were used to train models that map participants' motor performance and physiological signals to their desirable difficulties. To elicit different responses (i.e., motor performance and physiological changes) participants were asked to exercise with a robotic rehabilitation reaching task. From the most accurate to the least accurate, the three machine learning algorithms are Neural Network, Discriminant Analysis, and k-Nearest Neighbor. All three Artificial intelligence methods showed greater performance in comparison with random prediction. Additionally, results showed use of motor performance attributes as an input to the prediction models yields higher accuracy. However, we believe use of physiological attributes in conjunction with motor performance features can provide a richer source of information for predicting user's preference. Future work following this validation phase will include the exploration of more sophisticated machine learning methods such as Support Vector Machines (SVM) [13], as well as human subject studies with stroke survivors to investigate long-term effects of training with a system capable of providing exercises at the user's desirable difficulty level. We hypothesize that utilizing such systems will increase patients' motivation for doing their therapy exercises and ultimately realize a higher self-efficacy and more intense therapy program.

Acknowledgment

The Authors would like to thank Dr. Elizabeth Croft for her feedback on this work, Arnold Yeung and Sara Sheikholeslami for their help in physiological signals processing, and the individuals who participated in this study.

References

[1] J.D. Schaechter, "Motor rehabilitation and brain plasticity after hemiparetic stroke," Progress in Neurobiology, vol. 73, pp. 61-72, 2004.

[2] M.L. Aisen, H.I. Krebs, N. Hogan, F. McDowell, and B.T. Volpe, "The effect of robot-assisted therapy and rehabilitative training on motor recovery following stroke," Archives of Neurology, vol. 54, 1997, pp. 443-446.

[3] P.S. Lum, C.G. Burgar, P.C. Shor, M. Majmundar, and H.F.M. Van der Loos, "Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke," Archives of Physical Medicine and Rehabilitation, vol. 83, Jul. 2002, pp. 952-959.

[4] F. Amirabdollahian, R. Loureiro, E. Gradwell, C. Collin, W. Harwin, and G. Johnson, "Multivariate analysis of the Fugl-Meyer outcome measures assessing the effectiveness of GENTLE/S robot-mediated stroke therapy," Journal of NeuroEngineering and Rehabilitation, vol. 4, 2007, p. 4.

[5] S. Charles, H.I. Krebs, B. Volpe, D. Lynch, and N. Hogan, "Wrist rehabilitation following stroke: initial clinical results," in Proc. IEEE International Conference on Rehabilitation Robotics ICORR, 2005, pp. 13-16.

[6] A. Gupta, M.K. O'Malley, V. Patoglu, and C. Burgar, "Design, Control and Performance of RiceWrist: A Force Feedback Wrist Exoskeleton for Rehabilitation and Training," The International Journal of Robotics Research, vol. 27, 2008, pp. 233-251.

[7] D.M. Wolpert, Z. Ghahramani, and M.I. Jordan, “An internal model for sensorimotor integration,” Science, vol. 269, 1995, pp. 1880–1882.

[8] D. B. Willingham, “The neural basis of motor-skill learning,” Current Directions in Psychological Science , vol. 8, no. 6 , 1999, pp. 178-182.

[9] N. Shirzad, and H.F.M.Van der Loos, "Error amplification to promote motor learning and motivation in therapy robotics," in Proc. IEEE International Conference on Engineering in Medicine and Biology Society EMBC, 2012, pp. 3907-3910.

[10] Y. Wei, P. Bajaj, R. Scheidt, and J. Patton, “Visual error augmentation for enhancing motor learning and rehabilitative relearning,” in Proc. IEEE International Conference on Rehabilitation Robotics ICORR, 2005, pp. 505-510.

[11] N. Maclean, P. Pound, C. Wolfe, and A. Rudd, “Qualitative analysis of stroke patients’ motivation for rehabilitation,” BMJ (Clinical research ed.), vol. 321, 2000, pp. 1051-1054.

[12] M.A. Guadagnoli, and T.D. Lee, “Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning,” Journal of Motor Behavior, vol. 36, 2004, pp. 212–224.

[13] P. Rani, C. Liu, N. Sarkar, E. Vanman, “An empricial study of machine learning technique sfor affect recognition in human-robot interaction,” Pattern Analysis and Applications, vol. 9, 2006, pp. 58-69.

[14] C. Liu, P. Rani, and N. Sarkar, “Affective state recognition and adaptation in human-robot interaction: a design approach,” IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 3099-3106.

[15] C. Liu, K. Conn, N. Sarkar, and W. Stone, "Online affect detection and robot behavior adaptation for intervention of children with autism," IEEE Transactions on Robotics, vol. 24, no. 4, 2008, pp. 883-896.

[16] F. Wang; D.E. Barkana, and N. Sarkar, "Impact of visual error augmentation when integrated with assist-as-needed training method in robot-assisted rehabilitation," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 18, no. 5, 2010, pp. 571-579.

[17] W.J. Atsma, “Inference of central nervous system input and its complexity for interactive arm movement,” Ph.D. dissertation, Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada, 2006.