

# A Decision-Theoretic Approach in the Design of an Adaptive Upper-Limb Stroke Rehabilitation Robot

Rajibul Huq<sup>1</sup>, Patricia Kan<sup>1</sup>, Robby Goetschalckx<sup>2</sup>, Debbie Hébert<sup>3,5</sup>, Jesse Hoey<sup>4</sup>, and Alex Mihailidis<sup>1,3,5§</sup>

<sup>1</sup>Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada

<sup>2</sup>School of Computing, University of Dundee, Dundee, Scotland

<sup>3</sup>Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, Canada

<sup>4</sup>School of Computer Science, University of Waterloo, Waterloo, Canada

<sup>5</sup>Toronto Rehabilitation Institute, Toronto, Canada

<sup>§</sup>Corresponding author

Email addresses: rajibul.huq@utoronto.ca, patricia.kan@utoronto.ca, robby@computing.dundee.ac.uk, hebert.debbie@torontorehab.on.ca, jhoey@cs.uwaterloo.ca, and alex.mihailidis@utoronto.ca

**Abstract**—This paper presents an automated system for a rehabilitation robotic device that guides stroke patients through an upper-limb reaching task. The system uses a partially observable Markov decision process (POMDP) as its primary engine for decision-making. The POMDP allows the system to automatically modify exercise parameters to account for the specific needs and abilities of different individuals, and to use these parameters to take appropriate decisions about stroke rehabilitation exercises. The performance of the system was evaluated through various simulations and by comparing the decisions made by the system with those of a human therapist for a single patient. In general, the simulations showed promising results and the therapist thought the system decisions were believable.

**Keywords** - upper-limb stroke rehabilitation; robotic reaching exercise; artificial intelligence, POMDP

## I. INTRODUCTION

Stroke is the leading cause of physical disability and death around the world [1, 2]. Research has shown that post-stroke impairments and disabilities can be reduced by intensive, repetitive, and goal-directed rehabilitation, which improves motor function and cortical reorganization in stroke patients with both acute and long-term (chronic) impairments [3]. The recovery process, however, is typically slow and labor-intensive, usually involving extensive interaction between a therapist and a patient. One of the main motivations for developing rehabilitation robotic devices is to automate repetitive and physically demanding interventions, which can alleviate strain on therapists. This technology makes it possible for a single therapist to supervise multiple patients simultaneously, and for home-based therapy, which can contribute in the reduction of health care costs.

The upper extremities are typically affected more than the lower extremities after stroke [4]. There have been several types of robotic devices designed to deliver upper-limb rehabilitation for people with paralyzed upper extremities. The Assisted Rehabilitation and Measurement (ARM) Guide [5] presents a technique called “active assist therapy” that helps a user to complete a desired reaching task. The Mirror Image Movement Enabler (MIME) therapy system [6] uses a six-

degree of freedom (DOF) robot manipulator that applies forces through an orthosis to the user’s affected arm in order to accomplish both unimanual and bimanual goal-directed movements in 3-dimensional space. The GENTLE/s system [7] uses a 3-DOF robot that allows pronation/supination of the elbow as well as flexion and extension of the wrist through a gimbal mechanism. The rehabilitation robotic device that has received the most clinical testing is the Massachusetts Institute of Technology (MIT)-MANUS [8]. It consists of a 2-DOF robot manipulator that assists shoulder and elbow movements by moving the user’s hand in the horizontal plane.

Recent work has also employed virtual reality games [9] and artificial intelligence (AI) methods, e.g. fuzzy logic [10] and artificial neural network [11], to improve upon the active assistance techniques found in the previous systems mentioned above. While these robotic systems have shown promising results, none of them is able to autonomously adjust different exercise parameters according to each individual’s needs. The rehabilitation systems discussed above also do not account for physiological factors, such as fatigue, which can provide information as to when the user should take a break and rest, which may benefit rehabilitation progress [12].

The research described in this paper aims to fill these existing gaps by using stochastic modeling and decision theoretic reasoning to autonomously facilitate upper-limb reaching rehabilitation for moderate level stroke patients, tailor the exercise parameters for each individual, and estimate user fatigue. This paper is an extension of our previous work [13] that presents an intelligent controller using POMDP (partially observable Markov decision process) for reaching rehabilitation task. This paper also presents early pilot data to show the efficacy of the new system.

## II. REHABILITATION SYSTEM OVERVIEW

The automated upper-limb stroke rehabilitation system consists of three main components: the exercise (Fig. 1), the robotic system (Fig. 2(a)), and the POMDP agent (Fig. 2(b)). As the user performs the reaching exercise on the robot, data from the robotic system are used as input to the POMDP, which decides on the next action for the system to take.

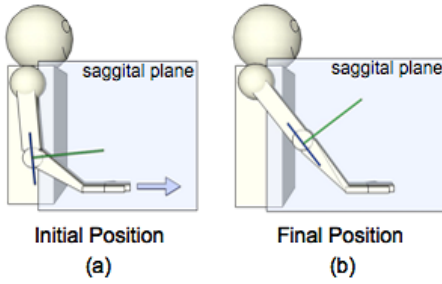


Figure 1. The reaching exercise

#### A. The exercise

A targeted, load-bearing, forward reaching exercise was chosen by consulting the therapists of this project. Reaching is one of the most important abilities to possess, as it is the basic motion involved in many activities of daily living [12]. Fig. 1 provides an overview of the reaching exercise. The reaching exercise is performed in the sagittal plane (aligned with the shoulder) and begins with a slight forward flexion of the shoulder, and extension of the elbow and wrist (Fig. 1(a)). Weight is translated through the heel of the hand as it is pushed forward in the direction indicated by the arrow, until it reaches the target position (Fig. 1(b)). The return path brings the arm back to the initial position. The goal is to have patients gradually reach the furthest target at maximum resistance, while performing the exercise with control (e.g. no deviation from the straight path) and proper posture (e.g. without trunk rotation, shoulder abduction/internal rotation).

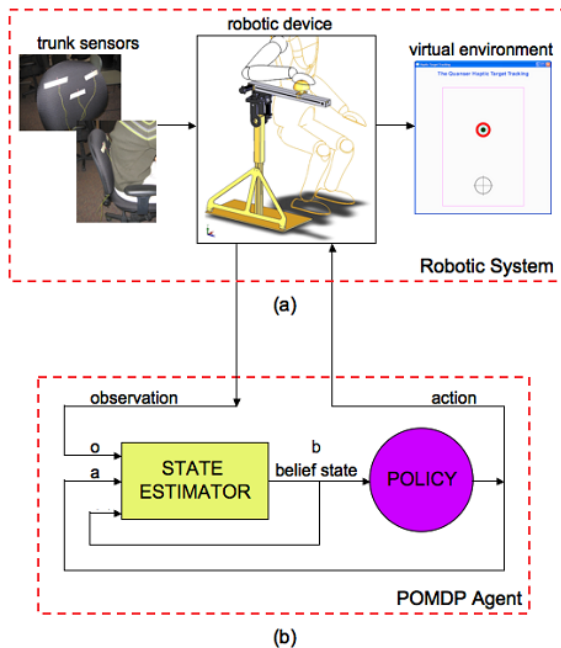


Figure 2. The reaching rehabilitation system

#### B. Robotic system

Fig. 2(a) shows the robotic system, which is comprised of three main components: the robotic device, the postural sensors, and the virtual environment.

The non-restraining robotic platform, shown in Fig. 3(a), was built by Quanser Inc., a robotics company in Toronto. It

has two degrees of freedom, which allow the reaching exercise to be performed in 2D space. The robotic device incorporates haptic technology to provide resistance and boundary guidance for the user during the exercise.

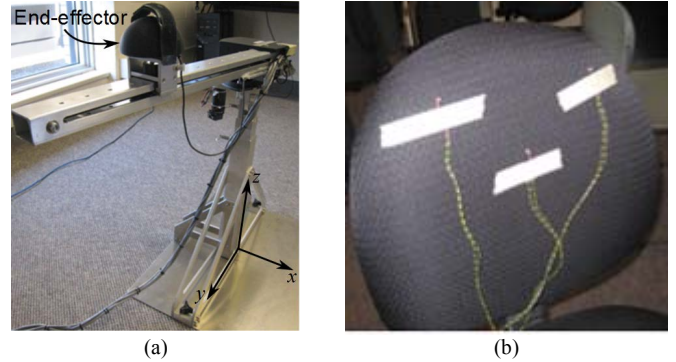


Figure 3. Robotic platform

Encoders in the end-effector of the robotic device provide data to indicate hand position and shoulder abduction/internal rotation (i.e. compensation) during the exercise. The unobtrusive trunk sensors (Fig. 3(b)) provide data to indicate trunk rotation compensation. The trunk sensors are comprised of three photoresistors taped to the back of a chair, each in one of three locations: the lower back, lower left scapula, and lower right scapula. The detection of light during the exercise indicates trunk rotation, as it means a gap is present between the chair and user.

Finally, the virtual environment represents the reaching exercise in the form of a 2D bull's eye game. The goal of the game is for the user to move the robot's end-effector, which corresponds to the cross-tracker in the virtual environment, to the bull's eye target. The rectangular box is the virtual (haptic) boundary, which keeps the cross-tracker within those walls during the exercise.

#### C. POMDP agent

The POMDP agent (Fig. 2(b)) is the decision-maker of the system. Observation data from the robotic device are passed to a state estimator that estimates the progress of the user as a probability distribution over the possible states, known as a belief state. A policy then maps the belief state to an action for the system to execute, which can be either setting a new target position and resistance level or stopping the exercise.

#### D. Partially observable Markov Decision Process

A POMDP is a decision-theoretic model defined by: a finite set of world states  $S$ ; a finite set of actions  $A$ ; a finite set of observations  $O$ ; a transition function  $T : S \times A \rightarrow \prod(S)$ , where  $\prod(S)$  denotes a probability distribution over states  $S$ , and  $P(s' | s, a)$  denotes the probability of transition from state  $s$  to  $s'$  when action  $a$  is performed; an observation function  $Z : S \times A \rightarrow \prod(O)$ , with  $P(o | a, s')$  denoting the probability of observing  $o$  after performing action  $a$  and transiting to state  $s'$ ; and a reward function  $R : S \times A \times O \rightarrow \mathbb{R}$ , with  $R(s, o, a)$  denoting the expected reward or cost (i.e. negative reward) incurred after performing action  $a$  and observing  $o$  in state  $s$ .

The POMDP agent is used to find a policy (i.e. course of action) that maximizes the expected discounted sum of rewards attained by the system over an infinite horizon, to monitor beliefs about the system state in real time, and to use the computed policy to decide which actions to take based on the belief states [14].

#### E. Justification for using a POMDP to model reaching rehabilitation

The rehabilitation system for the reaching task does not use any physical sensors to directly measure user fatigue, which is considered a “hidden” state. Hence, a POMDP-based formalism is suitable for estimation of user fatigue by capturing observable states such as user compensation and control. The reaching exercise is also a stochastic decision problem where choosing a particular action (target distance and resistance) at a particular state does not always produce the same outcome (user fatigue), instead, the action has a random chance of producing a specific result with a known probability. A POMDP can account for the uncertainty associated with an action through its transition probabilities and reward functions. By knowing the probabilities and rewards of the outcomes of taking an action in a specific state, the POMDP agent can estimate the likelihood of future outcomes to determine the optimal course of action to take in the present. The reaching task also needs to adapt to each individual patient’s needs and abilities over time. A POMDP has the capability of incorporating user abilities autonomously in real-time by keeping track of which actions have been observed to be the most effective in the past. For example, the POMDP may decide to keep the target closer for a longer period of time for patients who are progressing slowly, but may increase the target location further at a quicker rate for those who are progressing faster. Overall, the autonomous decision-making capability of a POMDP may help to reduce the health care cost since the decision-making process does not involve direct input of the therapists.

### III. POMDP MODEL

Fig. 4 shows the POMDP model (without rewards) as a dynamic Bayesian network (DBN). The following sections describe different components of the system.

#### A. Actions, variables, and observations

The system can take 10 possible actions, where nine actions are specified as different combinations of target distance  $d \in \{d1, d2, d3\}$  and resistance level  $r \in \{none, min, max\}$ , and the last action is specified to stop the exercise when the user is fatigued.

The following variables were chosen to represent the exercise: 1)  $fatigue = \{yes, no\}$  describes the user’s level of fatigue, 2)  $n(r) = \{none, d1, d2, d3\}$  describes the range (or ability) of the user at a particular resistance level,  $r \in \{none, min, max\}$ . The range is defined as the furthest target distance,  $d \in \{d1, d2, d3\}$ , the user is able to reach at a particular resistance. For example, if  $r=min$  and the furthest target the user can reach is  $d=d2$ , then the user’s range is  $n(min)=d2$ . 3)  $stretch = \{+9, +8, +7, +6, +5, +4, +3, +2, +1, 0, -1, -2\}$  describes the amount the system is asking the user to go beyond his/her

current range. It is a deterministic function of the system’s choice of resistance level ( $a_r$ ) and distance ( $a_d$ ), which measures how much this choice is going to push a user beyond his/her range, and is computed as follows:

$$stretch = [a_d + n_{a_r}] + \sum_{r=1}^{a_r-1} [3 - n_r] \quad (1)$$

where  $r \in \{1=none, 2=min, 3=max\}$ ,  $a_r \in \{1, 2, 3\}$ , and  $a_d \in \{1, 2, 3\}$ . Variables  $n_r \in \{0, 1, 2, 3\}$  and  $n_{a_r} \in \{0, 1, 2, 3\}$  indicate ranges at  $r$  and  $a_r$ . 4)  $learnrate = \{lo, med, hi\}$  describes how quickly the user is progressing during the exercise.

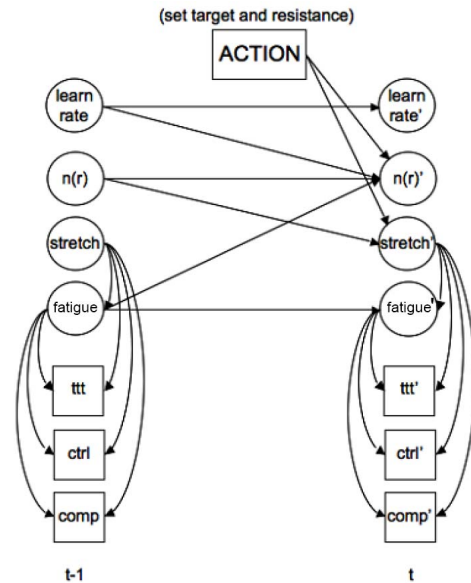


Figure 4. POMDP model as a DBN

The observations were chosen as follows: 1)  $ttt = \{none, slow, norm\}$  describes the time it takes the user to reach the target, 2)  $ctrl = \{none, min, max\}$  describes the user’s control level by his/her ability to stay on the straight path, and 3)  $comp = \{yes, no\}$  describes any compensatory actions (i.e. improper posture) performed.

Note that, although the observations are fully observable, the states are still not known with certainty since the fatigue, user range, stretch, and learning rate variables are unobservable and must be estimated.

#### B. Dynamics

The dynamics of all variables were specified manually using simple parametric functions of  $stretch$  and the user’s  $fatigue$ . The functions relating  $stretch$  and  $fatigue$  levels to user performance are called *pace functions*. The pace function,  $\phi$ , is a function of the stretch,  $s$ , and fatigue,  $f$ , and is a sigmoid function defined as follows:

$$\phi(s, f) = \frac{1}{1 + e^{-\left[\frac{s-m(f)}{\sigma_s}\right]}} \quad (2)$$

where  $m$  is the mean stretch (the value of stretch for which the function  $\phi$  is 0.5 when the user is not fatigued),  $m(f)$  is a shift function that is dependent on the user's fatigue level (e.g. 0 if the user is not fatigued), and  $\sigma_s$  is the slope of the pace function. There is one such pace function for each variable, and the value of the pace function at a particular stretch and fatigue level gives the probability of the variable in question being true in the following time step. Fig. 5 shows an example of pace function for  $comp=yes$ . It shows that when the user is not fatigued and the system sets a target with a stretch of 3 (upper pace limit), the user might have a 90% chance to compensate. However, if the stretch is -1 (lower pace limit), then this chance might decrease to 10%. The pace limits decrease when the user is fatigued (at the same probability). In other words, the user is more likely to compensate when fatigued.

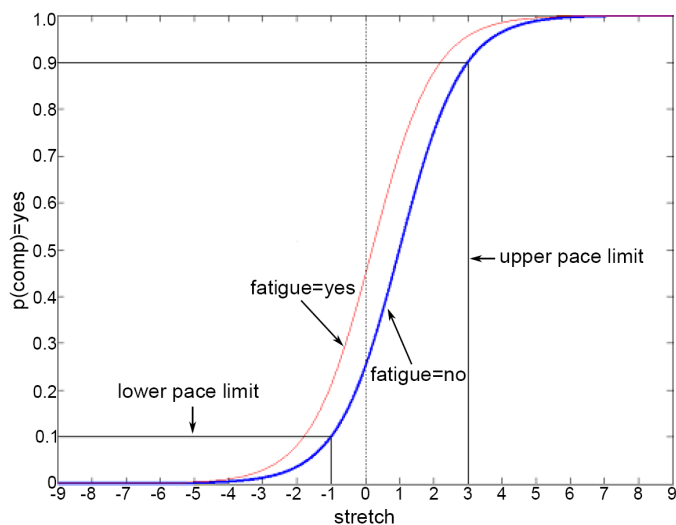


Figure 5. Pace function for  $comp=yes$

The detailed procedure to specifying  $m$ ,  $\sigma_s$ , and  $m(f)$  has been described in our previous work [13].

In the current model, the ranges were modeled separately. The dynamics for the ranges state that setting targets at or just above a user's range will cause his/her range to increase slowly, but less so if the user is fatigued. If a user's range is at  $d3$  for a particular resistance, then practicing at that distance and resistance will increase his/her range at the next higher resistance from  $none$  to  $d1$ . The dynamics also includes constraints to ensure that ranges at higher resistances are always less than or equal to those at lower resistances. Finally, the dynamics of range includes a dependency on the learning rate ( $learnrate$ ): higher learning rates cause the ranges to increase more quickly.

### C. Rewards and computation

The reward function was constructed to motivate the system to guide the user to exercise at maximum target distance and resistance level, while performing the task with maximum control and without compensation. Thus, the system was given a large reward for getting the user to reach the furthest target distance ( $d=d3$ ) at maximum resistance ( $r=max$ ). Smaller rewards were given when targets were set at

or above the user's current range (i.e. when  $stretch \geq 0$ ), and when the user was performing well (i.e.  $ttt=norm$ ,  $ctrl=max$ ,  $comp=no$ , and  $fatigue=no$ ). However, no reward was given when the user was fatigued, failed to reach the target, had no control, or showed signs of compensation during the exercise.

The POMDP model had 82,944 possible states. The size of this reaching rehabilitation model renders optimal solutions intractable, thus, an approximation method was used. This approximation technique exploits the structure of the large POMDP by first representing the model using algebraic decision diagrams (ADDs) and then employing a randomized point-based value iteration algorithm [15], which is based on the Perseus algorithm [16] with a bound on the size of the value function. The model was sampled with a set of 3,000 belief points that were generated through random simulation starting from 20 different initial belief states: one for every range possibility. The POMDP was solved on a dual AMD Opteron™ (2.4GHz) CPU using a bound of 150 linear value functions and 150 iterations in approximately 13.96 hours.

### D. Simulation

A simulation program was developed in MATLAB® (before user trials) to determine how well the model was performing in real-time. The simulation began with an initial belief state. The POMDP then decided on an action for the system to take, which was predetermined by the policy. Observation data were manually entered and a new belief state was computed. This cycle continued until the system stopped the exercise because the user was determined (by the system) to be fatigued. Before the next cycle occurred, the simulation program reset the fatigue variable (i.e. user is un-fatigued after resting) and the user ranges were carried over.

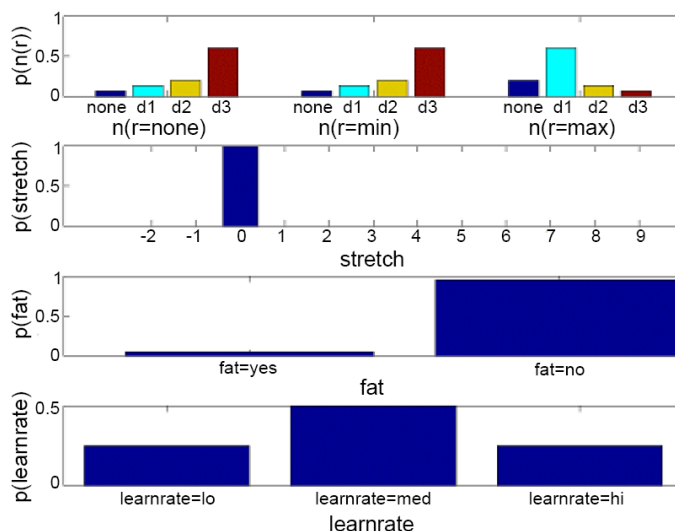


Figure 6. Initial POMDP belief state of example simulation

The simulation was continued for nine decision cycles before the stop decision was made. During simulation, the POMDP slowly increased the target distance and resistance level when the user successfully reached the target in normal time, had maximum control, and did not compensate. However, once the user started to lose control, compensated, or had trouble reaching the target, the POMDP increased its

belief that the user was fatigued and stopped the exercise to allow the user to rest.

An example simulation is described as follows with the assumption that the user is able to reach the maximum target ( $d=d3$ ) at the maximum resistance level ( $r=max$ ), but then slowly starts to compensate after several repetitions. The initial belief state (Fig. 6) assumes that the user’s range at both zero and minimum resistance (i.e.  $n(none)$  and  $n(min)$ ) is likely to be  $d3$ , and the user’s range at maximum resistance ( $n(max)$ ) is likely to be  $d1$ . In addition, the initial belief state assumes that the user is not fatigued with a 95% probability. From this belief state, the POMDP sets the first action to be  $d=d1$  and  $r=max$ . According to the assumption, the user successfully reaches this target in normal time, with maximum control, and with no compensation. In the next five time steps, the POMDP sets the target at  $d=d2$  and then increases it to  $d=d3$ , assuming the user successfully reaches each target with maximum control and no compensation. Here, the user’s fatigue level has increased slowly from approximately 5% to 20% due to repetition of the exercise. Now, during the next time step when the POMDP decides to set the target at  $d=d3$  again, the user compensates but is still able to reach the target with maximum control. In this case, the fatigue level jumped to about 40% due to user compensation. The POMDP sets the same target during the 9<sup>th</sup> time step and the user compensates once more. This time, the POMDP decides to stop the exercise because it believes the user is fatigued due to performing compensatory movements for two consecutive times.

#### IV. PILOT STUDY – EFFICACY OF POMDP

A pilot study was conducted with therapists and stroke patients to evaluate the efficacy of the POMDP agent—i.e. the correctness of its decisions.

##### A. Participants

Due to a delay in receiving ethics approval, only one therapist and one patient were recruited for the study. Therefore, the results of this section, being taken from a single patient-therapist dyad, should be taken only as a preliminary demonstration of the applicability of the system, and not a direct evaluation of the efficacy. The therapist was a physical therapist with more than nine years of experience in post-acute upper-limb stroke rehabilitation, and was fluent in English. The patient was right-side hemiparetic, had a stroke onset of 227 days (7 months and 14 days) before enrolment, scored 4 on the arm section of the Chedoke-McMaster Stroke Assessment (CMSA) Scale [17], was able to move to some degree but still had impaired movements as determined by her therapist, and could understand and respond to simple instructions. The patient continued with any outpatient therapies in which they were enrolled in at the time of study acceptance.

##### B. Methods

The patient participant was paired up with the therapist participant for the duration of the study. Each session lasted for approximately one hour and was completed three times a week for two weeks.

For each session, the therapist brought the patient to the testing room. The patient participant was seated on a regular, straight-back chair positioned to the left of the robotic device. The therapist was responsible for adjusting the position of the chair, placing the trunk sensors at the appropriate spots (lower back, lower left scapula, and lower right scapula), and adjusting the height of the robot to ensure that the end-effector was correctly positioned in the sagittal plane of the patient’s right shoulder. When both participants were ready to begin, the researcher powered on the robotic device and started the computer programs that controlled the POMDP agent, robotic device, and virtual environment.

The exercise was performed in three parts: (A) after the POMDP made a decision (i.e. to set the target position and resistance level, or to stop the exercise) the therapist either agreed or disagreed with the decision made; (B) the researcher had the device either execute the decision made by the POMDP if the therapist agreed or execute the decision made by the therapist if the therapist disagreed; and (C) the patient then performed the reaching exercise by trying to reach the target on the computer screen. These parts were repeated in the order of A-B-C until the end of the session.

Questions were asked at the end of each session and at the completion of the study for both participants. The questionnaire for the therapist participant was designed to focus on rating the decision-making strategy of the POMDP agent. For the patient participant, the questionnaire focused on gathering feedback with respect to her satisfaction in using such a robotic system. Both questionnaires consisted of quantitative and qualitative questions for statistical analysis and to provide insight into future design improvements, respectively. A four-point Likert scale was used for each quantitative question, with one representing complete disagreement and four representing complete agreement.

#### V. RESULTS AND DISCUSSION

##### A. Agreement of POMDP decisions

Every decision made by both the POMDP and therapist was decomposed into three separate decisions: 1) the distance to set the target, 2) the level to set the resistance, and 3) whether or not to stop the exercise. The level of agreement by the therapist to the decisions made by the POMDP was calculated based on the three separate decisions as described above. A point of agreement would be given if the therapist set the same target distance as the POMDP, set the same resistance level as the POMDP, or agreed with the POMDP to stop the exercise or not. Figure 9 shows the percentage of agreement over all sessions. Note that there were 636 state transitions (i.e. total number of trials) and 1,154 decisions made by the system during the study.

TABLE I. PERCENTAGE OF AGREEMENT

	Target	Resistance	Stop	Overall
Percentage of agreement (%)	94.21	97.30	43.08	65.25

The therapist agreed with both the target distance and resistance level decisions made by the POMDP approximately



94% and 90% of the time, respectively, during the study (shown in Table I). Most of this agreement was with the POMDP repeatedly setting the target distance at  $d3$  and the resistance at  $max$ . Since the patient was able to reach this setting within the first session with proper posture and control, the POMDP continued to make this decision as it was given large rewards for getting the user to reach the furthest target at maximum resistance. This indicates that the overall mapping from the POMDP resistance/distance values to the actual settings on the device was set improperly for the subject in question. This observation underlines the necessity of proper calibration by a therapist of the system prior to use. This calibration is part of our future work.

The therapist only agreed with the POMDP approximately 43% of the time for the stop decision. The POMDP wanted to stop the exercise to let the user take a break far more often than the therapist wanted. If the therapist did not see any signs of fatigue from the user, she would have the patient continue practicing the exercise for a longer period of time and not stop. Since the percentage of agreement for the stop decision was low, the overall therapist agreement with the POMDP decisions dropped to approximately 65%.

The therapist’s decisions alternated between having the patient work on muscle strengthening (by repeatedly setting the distance and resistance at the highest level) and on control (by randomizing the target distance and resistance levels). However, randomization was not part of the POMDP’s initial objective and thus, the POMDP would never make the decision to randomize the target distance and resistance levels.

### B. Questionnaire Data

Table II summarizes the therapist’s session responses, in terms of mean and standard deviation (SD), regarding the appropriateness of the decisions made during the exercise and whether the patient was given enough time to complete each exercise before the next decision was made.

TABLE II. THERAPIST EVALUATION ON POMDP DECISION

Rating on Likert scale (1-4)	Decision type			
	<i>The decision made during the exercise was appropriate</i>		<i>The patient was given an appropriate amount of time to complete each exercise before the next decision was made</i>	
	Mean	SD	Mean	SD
	2.8	0.41	3.2	0.41

The therapist’s rating on the appropriateness of the amount of time given to complete each exercise before the next decision was made was generally favorable with a mean score of 3.2 out of 4.0 on the Likert scale. However, the appropriateness of the decisions made by the POMDP during the sessions was less favorable with a mean score of 2.8 out of 4.0. Comments from the therapist suggested that randomizing the target distance and resistance level would be beneficial for the patient to work on control in addition to strengthening.

With the help of a translator, the patient was able to answer the final questionnaire at the end of the study, which consisted of eight quantitative four-point Likert scale questions and four qualitative questions. From the patient’s quantitative results, the patient found the quality of motion of the robotic device to

be very smooth with a score of 4.0 out of 4.0. The patient also felt that the resistance applied by the robotic device was too little, scoring 1.0 out of 4.0. Throughout the study, the patient repeatedly commented that the exercise was “too easy”, a reflection of the device’s resistance levels not being properly tuned to this particular user before the start of the trial. The patient was not able to feel the trunk sensors at all during the exercise, which suggests that trunk compensatory movements can be captured unobtrusively. The patient also felt that the bull’s eye game was somewhat interesting, scoring 3.0 out of 4.0. The patient felt that the exercise closely resembled the reaching motion and conventional upper-limb therapy, scoring 3.0 out of 4.0 for both. In addition, the patient believed he would use this robotic system as his primary therapy, scoring 4.0 out of 4.0. The patient did not elaborate on the qualitative questions, thus, feedback from this section of the questionnaire was discarded.

## VI. FUTURE WORK

The immediate future work of this project is to test the POMDP model with more participants in order to obtain significant results. Besides this, the results from the pilot study provide the following insight into the future development of the POMDP model and overall system.

- The effect of randomization of different target distances and resistance levels on control needs to be studied.
- The dynamics of the fatigue variable and the cost of stop action may need to be changed in order to stop the exercise less often. For example, Fig. 7(a) shows simulated exercise run lengths between stops for different cost of stop action. The higher costs of stop action generates, on average, longer runs.

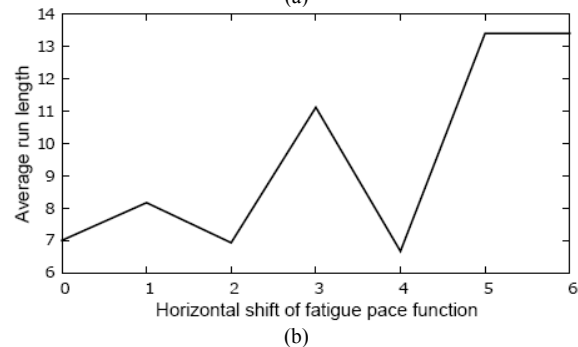
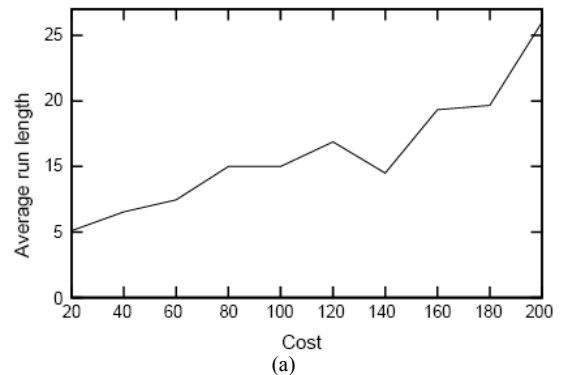


Figure 7. Average exercise run length between stops

Fig. 7(b) shows another simulated result for different horizontal shift of fatigue pace function. A horizontal shift of

the pace function changes the rate at which the user gets fatigued. Hence, higher shifts indicate that the user will be less likely to get fatigued with the same stretch. As a result, higher shifts lead to longer runs.

- The POMDP model needs to be expanded in order to include targets in 2D space. As a first step of this expansion, currently we are developing 2D virtual games that include target positions in 2D space. Fig. 8 shows an example where the target positions are set in a rectangular trajectory and the reaching task is to position the ball, which represents the end-effector of the robot, in the designated target position.

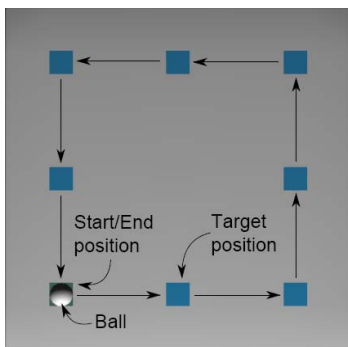


Figure 8. 2D trajectory

- The current robotic system only applies three discrete levels of resistance, which can be either increased or maintained at the same level during the exercise. The system will be more realistic if it is able to select varying levels of resistance that can be both increased and decreased to cope up with the need of an individual patient. Decreasing the resistance level may also result in lower fatigue probability and less frequent compensatory motion, which in turn may lead to longer duration of the exercise. To include these features into the current system, we are currently formulating a new probabilistic framework that models the users ability using Beta distributions [18] as a function of continuous resistance levels.

A Beta distribution is initially chosen since it is suitable for modeling success or failure in continuous space. Fig. 11 shows a simulated example with a range of continuous resistance levels from 0-20, where the probability of successfully finishing an exercise at a given resistance level is modeled with the following Beta distributions:  $\beta_n$  in case the person is not fatigued and  $\beta_y$  in case the person is fatigued. The total model is a weighted mixture of these two distributions, weighted according to the current belief that *fatigue=yes*. In this example, the posterior belief state assumes that probability of *fatigue=no* is 0.9 and probability of *fatigue=yes* is 0.1. The mixture model can be used to select the next resistance level for the exercise. In this example, the next resistance level 9.3 (shown in green circle in Fig. 11) is selected as the maximum resistance that produces  $\beta_{sum} \geq 0.5$ . Fig. 10(a) shows the next sequence where the distributions and the belief state are updated using the simulated observation that the person successfully completed the exercise (shown in red circle in Fig. 10(a)) at the resistance level 9.3.

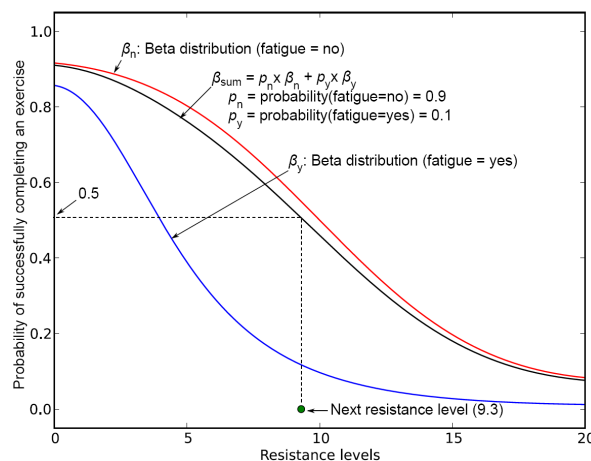


Figure 9. Continuous action using Beta distribution

The updated model is the posterior according to Bayes' rule. The next resistance level is set to 10.3 according to the updated  $\beta_{sum}$ .

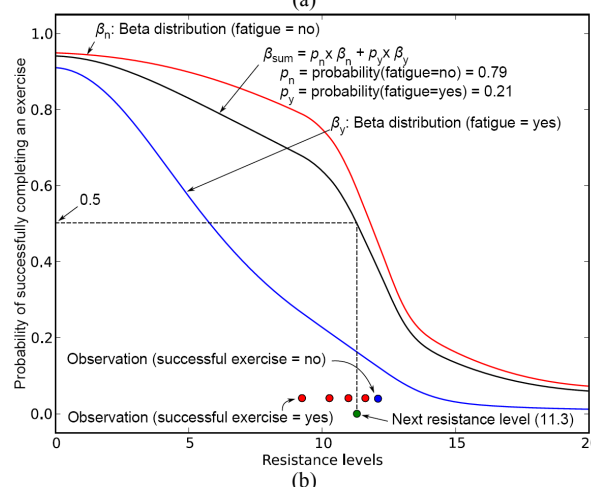
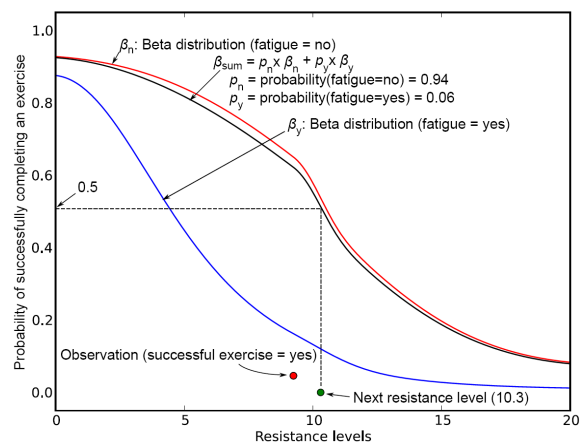


Figure 10. Updated distributions

Fig. 10(b) shows an instance where the distributions and belief state is updated after five observations. The first four observations are successful exercises (shown in red circle in Fig. 10(b)) and the last one is an unsuccessful exercise (the person did not reach the goal within acceptable time and

control or had to compensate too much - shown in blue circle in Fig. 10(b)). As a result, the next resistance level is set smaller compared to the current resistance. The exercise can be continued until the probability of *fatigue=yes* reaches a predefined threshold. Hence, this formulation - 1) is able to increase and decrease resistance levels in continuous space, and 2) is more adaptive to each individual patient's need since the distributions - the model of the person's abilities - are updated with the new observations. The initial shapes of the distributions can also be varied according to the condition of individual patient so that it produces appropriate resistance level while starting the exercise.

The same formulation can be applied to other state variables of the system. The preceding simulations are meant to demonstrate the feasibility of such a representation, and we are currently in the process of applying them to our rehabilitation device.

## VII. CONCLUSION

This paper presents a POMDP system that is designed for an upper-limb rehabilitation robotic device. A POMDP was chosen for this system because it has the ability to handle partial observability (e.g. user fatigue), adapt to users' needs, and operate autonomously. The goal of the POMDP agent is to help patients regain their maximum reaching distance at the most difficult level of resistance, while performing the exercises with control and proper posture. Computer simulations of the POMDP model showed that the POMDP was making decisions in alignment to those of conventional reaching rehabilitation, which was to gradually increase target distance first, then resistance level as the user performed well, and increase the rate of fatigue if the user was not performing well.

The performance of the system was also evaluated by comparing the decisions made by the system with those of a human therapist. A single patient participant was paired up with a therapist participant for the duration of the study. Overall, the therapist agreed with the system decisions approximately 65% of the time. In general, the therapist thought the system decisions were believable and could envision this system being used in both a clinical and home setting. The patient was satisfied with the system and would use this system as her primary method of rehabilitation. The data collected in this study can only be used to provide insight into the performance of the system since the sample size was limited.

The feedback from the therapist also suggests that the present system needs to include 2D target locations and varying levels of resistance. To include these features into the current system, we are currently developing virtual games with 2D target locations and a new probabilistic framework that expresses the probability of successfully completing an exercise using Beta distributions as a function of continuous resistance levels. The distributions are continuously updated with the new observations to reflect the performance of each individual patient. The system is also able to increase or decrease resistance levels according to the performance of a patient. The flexibility of decreasing resistance levels may also

result in lower fatigue probability and thus may prevent early stopping of the exercise.

## REFERENCES

- [1] Heart and Stroke Foundation of Canada: Stroke Statistics [<http://www.heartandstroke.com/site/c.iklQLcMWJtE/b.3483991/k.34A8/Statistics.htm#stroke>]
- [2] American Heart Association: Stroke Statistics [<http://www.americanheart.org/presenter.jhtml?identifier=4725>]
- [3] S. E. Fasoli, H. I. Krebs, and N. Hogan, "Robotic technology and stroke rehabilitation: Translating research into practice," *Topics in Stroke Rehabilitation*, vol. 11, no. 4, pp. 11-19, 2004.
- [4] L. R. Caplan, *Stroke*, New York: Demos Medical Publishing, 2006.
- [5] D. J. Reinkensmeyer, L. E. Kahn, M. Averbuch, A. McKenna-Cole, B. D. Schmit, and W. Z. Rymer, "Understanding and treating arm movement impairment after chronic brain injury," *Progress with the ARM guide, Journal of Rehabilitation Research and Development*, vol. 37, no. 6, pp. 653-662, 2000.
- [6] P. S. Lum, C. G. Burgar, P. C. Shor, M. Majmundar, and 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, no. 7, pp. 952-959, 2002.
- [7] 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, no. 4, pp. 1-16, 2007.
- [8] H. I. Krebs, N. Hogan, M. L. Aisen, and B. T. Volpe, "Robot-aided neurorehabilitation," *IEEE Transactions on Rehabilitation Engineering*, vol. 6, no. 1, pp. 75-87, 1998.
- [9] L. E. Sucar, R. Leder, D. Reinkensmeyer, J. Hernández, G. Azcárate, N. Castañeda, et al., "Gesture Therapy: A low-cost vision-based system for rehabilitation after stroke," In *Proceedings of the First International Conference on Health Informatics, Funchal, Madeira, Portugal, 28-31 January 2008*, vol. 2, pp. 107-111.
- [10] M. S. Ju, C. C. K. Lin, D. H. Lin, I. S. Hwang, and S. M. Chen, "A rehabilitation robot with force-position hybrid fuzzy controller: Hybrid fuzzy control of rehabilitation robot," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 3, pp. 349-358, 2005.
- [11] D. Erol, V. Mallapragada, N. Sarkar, G. Uswatte, and E. Taub, "Autonomously adapting robotic assistance for rehabilitation therapy," Paper presented at the *First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechanics, Pisa, Italy, 20-22 February 2006*.
- [12] M. Barnes, B. Dobkin, and J. Bogouslavsky, *Recovery after stroke*, United Kingdom: Cambridge University Press, 2005.
- [13] P. Kan, J. Hoey and A. Mihailidis, "Automated upper extremity rehabilitation for stroke patients using a partially observable Markov decision process," *AAAI Fall Symposium on AI in Eldercare*, 2008.
- [14] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, pp. 99-134, 1998.
- [15] P. Poupart, *Exploiting structure to efficiently solve large scale partially observable Markov decision processes*, PhD thesis, University of Toronto, Department of Computer Science, 2005.
- [16] M. T. J. Spaan and N. Vlassis, "Perseus: Randomized point-based value iteration for POMDPs," *Journal of Artificial Intelligence Research*, vol. 24, pp. 195-220, 2005.
- [17] C. Gowland, P. Stratford, M. Ward, J. Moreland, W. Torresin, S. Van Hullenaar, et al., "Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment," *Stroke*, vol. 24, no. 1, pp. 58-63, 1993.
- [18] A. C. G. V. Lazo and P. N. Rathie, "On the entropy of continuous probability distributions," *IEEE Transactions on Information Theory*, vol. IT-24, no. 24, pp. 120-122, 1978.