On the Relationship between Autonomy, Performance, and Satisfaction: Lessons from a Three-Week User Study with post-SCI Patients using a Smart 6DOF Assistive Robotic Manipulator

Dae-Jin Kim\(^1\), Rebekah Hazlett\(^2\), Heather Godfrey\(^3\), Greta Rucks\(^3\), David Portee\(^3\), John Bricout\(^2\), Tara Cunningham\(^3\), and Aman Behal\(^1\)

\(^1\)EECS & NSTC, University of Central Florida, Orlando, FL 32826 (abehal@mail.ucf.edu)\(^2\)School of Social Work, University of Central Florida, Orlando, FL 32826-3\(^3\)Orlando Health Rehabilitation Institute, Orlando, FL 32806.

Abstract: The UCF-MANUS, a vision-based 6DOF assistive robotic arm, has been designed to aid individuals with arm function limitations to complete tasks of daily living that they would otherwise be unable to complete themselves. This paper reports a small dual cohort pilot study with traumatic spinal cord injured (SCI) subjects designed to investigate the utility of the UCF-MANUS for these subjects. Pick-and-place ADL tasks were de-ned and users trained and tested with the system for three weeks during which they controlled the robot either through a manual or an autonomous (supervised) mode of operation. Baseline characteristics (pre-study), quantitative performance metrics (during study) and psychometrics (post-study) were obtained and statistically analyzed to test a set of hypotheses related to performance and satisfaction with the two control modes. It was seen that manual interaction showed more variability and ine-ciency in performance metrics as compared to autonomous operation. Suprisingly the latter mode, however, did not lead to better measures for user satisfaction. A dis-cussion is provided to explain the results. Based on qualitative feedback and quantitative results, possible directions for system design are presented in order to concurrently achieve better performance and satisfaction outcomes.

Index Terms: User Study, Spinal Cord Injury, Assistive Robotics

I. INTRODUCTION

In the United States, there are nearly 1 in 50 people (approximately 6 million) living with paralysis \(\mu\) about 23% of these cases result from Spinal Cord Injury (SCI) [1]. Many of these individuals with traumatic SCI may use a power wheelchair for mobility and have limited upper body mobility/strength which results in diminished function and thus requires caregiver assistance to perform activities of daily living (ADLs) [2]. A variety of assistive robotic devices/arms have been developed to improve functional ability and shown to assist with the performance of ADLs [3]. For instance, a desktop mounted robot, the Professional Vocational Assisitive Robot (ProVAR), was developed by researchers at the Palo Alto VA and Stanford University to assist individuals with high level SCI to independently perform ADLs and work activities in an of-ce environment [4][5]. A 5DOF robotic arm was developed by Farahmand et al. to help people on a wheelchair with mild to severe disabilities [6]. Most recently, Schrock et al. reported a new and light wheelchair-mounted robotic arm, WMRA-II, to meet the needs of mobility-impaired persons with limitations of upper extremities [7]. Yeong et al. presented an analysis of diverse pick-and-place motions of patients using a 3DOF robotic device [8]. A novel easy-to-use robotic device was developed by Kawashima et al. to prevent disuse syndrome and associated complications for patients post SCI [9].

While many assistive robotic devices have been developed over the years, it is clear from the latest IFR (International Federation of Robotics) report [10] that the market for robotic prostheses and assistive robots for the elderly and for users with disabilities is small at the current time compared with the leaps that have been made in other classes of service robots. It has been suggested that the development of robot autonomy has played a major part in the increased growth rate of robots that assist or entertain people in domestic settings or in recreational activities [11]. However, assistive technology adoption has lagged even though there has been a signi-cant increase in performance due to technical advances that have led to increased autonomy.

In this paper, we report the design and results of a three-week long comparative user study with two sets of patients post SCI using the UCF-MANUS [12], a 6DOF vision-guided assistive robotic arm. A group of patients was ini-tially recruited from spinal cord support groups and various baseline characteristics were measured to limit inclusion to an IRB approved subset of those recruited. Prior to the main study, a focus group was conducted to investigate user requirements and preferences for use of assistive technology. For the user study, two cohorts (formed by random selection process) employed two different control modes with differing amounts of autonomy and performed identical sets of tasks over a 3 week long period under the supervision of occupa-tional therapists (OTs). Based on the designed pick-and-place tasks, extensive quantitative and qualitative datasets were acquired and then statistically analyzed for signi-cance. The statistical analysis results were utilized to reject or validate a set of prior hypotheses related to performance, satisfaction, and their interrelationship.

This paper is organized as follows. Section II deals with the research objectives and hypothesis formulation. Technical details for the robotic device, inclusion/exclusion criteria for user selection as well as choice of outcome metrics and data
analysis techniques have been provided in Section III. An analysis and discussion of results from hypothesis testing is provided in Section IV. Finally, Section V concludes with an overview of the significance and outcomes from the study.

II. RESEARCH OBJECTIVE

The purpose of this study is to evaluate and compare two control modes with different levels of autonomy, namely, Auto (supervised autonomous operation) and Cartesian (i.e., manual operation) control modes, for the UCF-MANUS in terms of their effectiveness in improving independent function and quality of life for participants with traumatic SCI. Specifically, this study aims to provide sufficient quantitative and qualitative analysis to prove the utility of UCF-MANUS for a range of SCI users. Furthermore, the goal is to understand the differences in ability to complete tasks, rate of completion, and subject experience using the different control modes. The experimental design aims to investigate the following set of hypotheses about the user interaction with the developed UCF-MANUS system:

**Hypothesis 1 (H1)** Selection of a specific human computer interface modality does not detract from user performance in controlling the robot.

**Hypothesis 2 (H2)** ADL tasks can be classified into easy and hard categories based on initial relative pose between object and robot.

**Hypothesis 3 (H3)** Baseline characteristics of subjects are correlated with the quantitative performance metrics.

**Hypothesis 4 (H4)** User degree of satisfaction (as measured through psychometrics) is correlated with performance metrics.

III. STUDY DESIGN

A. Selection Criteria

In collaboration with Orlando Health Rehabilitation Institute (OHRI - a part of Orlando Health), subjects were recruited from central Florida spinal cord support groups through advertisement and screened for appropriateness according to the following exclusion and inclusion criteria. Inclusion criteria are individuals who are 21 years of age or older and at least 90 days post traumatic injury and suffer from C3 to C7 impairment. Candidates must be able to use a power wheelchair with joystick control (or other devices) as their primary means of mobility, and be cognitively able to perform tasks with a Mini Mental Status Exam (MMSE) score of 22 or greater. Additionally, participants must be willing to attend weekly videotaped training/testing sessions for a three-week consecutive time period. Duration of the weekly sessions will vary between one and two hours. Subjects are to be excluded if they can independently perform self care activities of daily living (ADLs) as evidenced by a score of 40 or greater on the self care subscale of the Functional Independence Measure™ [13]. Based upon an examination of medical records and test results conducted by clinicians at OHRI, qualifying subjects were asked to provide informed consent. Informed consent form was drawn from the research protocol approved by the Orlando Health and UCF institutional review boards.

B. Subject Grouping

Once enrolled, subjects were randomly divided into two cohorts to evaluate two control modes (Auto versus Cartesian) with the assistive robotic manipulator. Cartesian mode is where users command 3-D Cartesian position (forward, backward, left, right, up, and down) and 3-D orientation (yaw left, yaw right, pitch up, pitch down, roll clockwise, and roll counterclockwise) for the gripper at the end of the arm to enable interaction with the environment. On the other hand, Auto mode is where users click anywhere on the object of interest displayed on a screen as part of the scene captured by a gripper-mounted wide-angle video camera¹. After the click, the robotic system transports and steers its end-effector to appropriately grab the object autonomously. Because of the use of computer vision based generation of motion, Auto mode users are not required to command multiple translation and orientation velocities at the end-effector. The subjects utilizing the Cartesian mode were classified as Cohort C while the subjects utilizing the Auto mode were classified as Cohort A.

C. Robotic Platform

The base robotic manipulator used in this trial is the MANUS, which has been developed by Exact Dynamics based in The Netherlands. The MANUS is a commercially available robot with 6+1+1 degrees of freedom. Six rotational joints are used to generate 3-D position/orientation of the robot hand while two residual joints are used to control the opening/closing of the robot gripper and up/down motions of the base lift. The MANUS weighs approximately 33 lbs. The MANUS has a reach of 80 cm (31.5”) which can be extended another 25 cm using the optional lift. The payload capacity is 4.5 lbs while the maximum grip strength is 20N. It is typically controlled through interfaces such as a joystick, keypad, or other optional standard interface devices [14].

Since mid-2006, the UCF Assistive Robotics Laboratory has worked on developing the UCF-MANUS Arm through a series of sensory modulations to the gripper of the standard MANUS arm. Visual sensing is utilized to provide direct video feedback to the users as well as processed information to assist during autonomous operation. Pressure and touch sensing is utilized to avoid collisions with environment and safely grab objects through adaptive thresholding techniques. A multimodal user interface and computation algorithms have also been developed to incorporate autonomy and increase access to the device. The reader is referred to [12] and [15] for more details about the UCF-MANUS.

¹In some cases, users may be optionally required to pan/tilt the camera as needed if the object is not initially located in the scene captured by the gripper camera.
Fig. 1. Testing environment with 6DOF robotic arm with bi-level shelves and six ADL objects

D. Environmental Setup

Fig. 1 shows a simulated ADL task setup designed for user testing. Six ADL objects are placed on bi-level shelves and the user is asked to pick up each of the objects and bring them back to a specified spot. The different objects have varying shapes (curved and flat), sizes, and pose (upright versus laid down). These objects were chosen based on top ADL items reported by clinicians in a survey conducted by us. The two shelves represent different levels of task, namely, easy and hard. We hypothesize that a task is considered easy when the object to be picked up is placed on a normal countertop level shelf (30 cm height) while it is categorized as hard when the object to be picked up is placed at the floor level shelf (65 cm height). Subjects' wheelchairs and items used in the ADL tasks were positioned in predetermined spots that were marked on the table and floor to ensure a consistent starting and stopping point for each task and for each user. The robot was positioned next to the subjects and not mounted on their wheelchairs due to the different hardware requirements of the various wheelchairs and the time it would take to mount the robot.

E. Outcome Measures

In the above setup, two quantitative performance metrics were employed. From the starting point to the stop point, a time to task completion (TTC) metric was stored to represent task completion efficiency while a number of clicks (NOC) metric was used for effectiveness of robot commands. When the user operated the robot, these two factors were automatically processed and saved in each user’s database. Upon completion of the study, a modified Psychosocial Impact of Assistive Devices Scale (PIADS) [16] was administered for assessment of user satisfaction. The PIADS is designed to assess the impact of assistive technologies on functional independence, quality of life, and well-being. It contains three subscales: Competence, Adaptability, and Self-esteem. Scores on each subscale range from -3.0 (indicating maximum negative impact of the assistive technology) to +3.0 (indicating maximum positive impact of the assistive technology). Finally, semi-structured exit interviews were conducted post completion to gather qualitative feedback.

F. Study Protocol

Prior to and post completion of the study, subjects were medically evaluated by a physiatrist to ensure that they were not adversely affected by participation in the study. Optionally, all participants were given the opportunity to participate in a 90 minutes focus group during which assistive technology needs were discussed while no specific details of the ensuing user study were offered in order to avoid biasing the subset of subjects participating in the focus group. Prior to the three-week study, all subjects were evaluated by an OT to measure baseline characteristics as well as to assess the suitability of the various modalities of the multimodal user interface. During the study, all sessions were conducted and monitored by an occupational therapist (OT) to train users through demonstrations and prompting as well as to ensure the safety and well-being of the subjects. Each subject was scheduled to participate in one session lasting between one and two hours each week. During the first week, an OT directly assisted the subjects to become familiar with both the robotic device itself and the human computer interface, providing verbal feedback and physical assistance if necessary when performing different tasks with the robot. During the second and third weeks, subjects practiced using the robot independently but under direct supervision of the OT. Throughout the trials, a research assistant was in the same room for safety and in case subjects experienced any technical difficulty with the system. Sessions were video-taped to allow for review as necessary. At the end of the final evaluation, PIADS assessment was administered by the OT to understand impact of assistive technology as felt by the user.

G. Data Analysis Techniques

Due to the small sample size, a nonparametric test was adopted to analyze the data. Wilcoxon signed-rank test [17] was utilized to observe significant differences between the two cohorts using the TTC and NOC performance metrics. For testing the null hypothesis, alpha was set at 0.05, i.e., if Wilcoxon test for the given datasets resulted in \( p < alpha = 0.05 \), it can be concluded with more than 95% confidence that the datasets are significantly different. However, if the Wilcoxon test resulted in \( p > 0.05 \), it implies that the null hypothesis cannot be rejected, i.e., more data may need to be collected and analyzed to prove the alternate hypothesis. For the purposes of comparison between quantitative metrics (TTC and NOC) versus baseline characteristics or psychometrics (i.e., MMSE, MVPT-R, FIM, PIADS, etc.), we adopted Pearson product-moment correlation coefficient (PMCC) [18].
IV. RESULTS

A. Target Population

After the recruiting process described in Section III-A, ten individuals with SCI were chosen. Baseline characteristics of the target population were measured prior to the 3-week study. Mean age of the subjects was 41.1 (9.9 μ hereafter; this number in parentheses will denote standard deviation of a dataset) years and they were 16.7 (11.8) years past date of SCI. All of their SCIs occurred from traumatic causes such as automobile crashes and falls. Initial diagnosis of subjects was between C3 and C7 μ only one subject suffered from C7 impairment, however, the subject did not have the functionality of a typical C7 case. Mean MMSE score was 27.7 (1.64), and mean score on the self care subscale of the FIM\textsuperscript{TM} was 18.6 (9.50). Additionally, MVPT-R score was measured to assess the user’s visual perception independent of motor abilities μ mean MVPT-R score was 57.2 (5.01). Without the aid of the robotic assist device or help from other people, most of the subjects could not grab any of the six objects μ only one subject was able to grab all the objects on the high shelf. Most of the subjects had gross mobility in their lower arms, so, they were able to use a trackball and jelly switch combination as a suitable interface modality> however, one subject had severe upper extremity disability, therefore, they utilized a speech based interface.

B. Focus Group

Four subjects participated in the focus group prior to the start of the study. The focus group and its subsequent analysis were conducted to explore the process by which participants learn new assistive technology. Several thematic questions were asked during the focus group. The purpose of the questions was to gather information on the three main sensitizing concepts\textsuperscript{2} used by the participants, namely 1) Heuristics, 2) Technological adaptation, and 3) Social learning through networks.

Using Nvivo 8 software \cite{21}, a preliminary analysis of the focus group was conducted based on linguistically-coded transcripts of dialogues in the session. Four parent nodes were identified in the \textit{nal} coding tree which included (i) heuristics, (ii) interface, (iii) social learning, and (iv) training preferences. As shown in Fig. 2(A), participants reported using a variety of heuristic techniques which included practicing, observing others, and problem solving techniques. Child nodes under \textit{problem solving techniques} comprised of adaptation (tool and self adaptation), seeking information (client-to-supplier, peer to peer, and online), and trial and error (see Fig. 2(B)). The social networks in which participants interact to learn new assistive technology and solve concerns related to existing assistive technology were reported by participants as being primarily well established groups/networks whether they be face-to-face contacts or online. The interface preferences of participants yielded diverse and rich results (see Fig. 2(C)). The most important issues to this group of participants were \textit{simplicity, affordability, and reliability}. The initial results of training preferences yielded few responses (such as demonstration, verbal and visual instruction) and need to be revisited in future focus groups. For details on the development of the \textit{nal coding scheme and the various iterations that preceded it, the reader is referred to \cite{22}}.

C. Hypothesis Testing and Discussion

Given the experiment design, measurements, and statistical analysis techniques explained in Section III, the set of hypotheses provided in Section II was tested as follows:

\textbf{Hypothesis 1 (Choice of user interface):} We hypothesized that the choice of human computer interface doesn\textsuperscript{t} detract from the user\textsuperscript{s} performance in controlling the robot. Due to their limited functional capability, post-SCI users cannot effectively use (and therefore cannot comparatively test) the various modalities in the multimodal user interface. Hence, a group of able-bodied subjects was recruited and tested on their ef\textsuperscript{ciency at a task across different interfaces, namely, Touch Screen (TS), Trackball only (TO), Trackball and Jelly Switch (TJ), and Microphone and Jelly Switch (MJ)). To avoid bias, order of selection of interface was randomized for each user. Pairwise analysis was conducted to compare one interface against another using the Wilcoxon signed rank test. It was seen that TO performed signi\textsuperscript{cantly} poorly than TS in terms of time to completion \cite{22}. However, other interfaces had no significant difference with TS \textit{(TJ: Z=-0.0673, p>0.05 >MJ: Z=-0.6592, p>0.05)}. This implies that TO detracts from user performance and should not be selected. Next, voice command was not signi\textsuperscript{cantly different compared with other interfaces \textit{(TS: Z=-0.6592, p>0.05 >TO: Z=-0.6054, p>0.05 >and TJ: Z=-0.7399, p>0.05)}. This is important for post-SCI users because high injury (C3-C4) patients will only use MJ but they will not be disadvantaged by using that. Since most post-SCI users do not have signi\textsuperscript{cant movement dexterity}, TS was not considered usable by these users. Hence, our \textit{nal choice of two feasible human computer interfaces (TJ and MJ) for this study was validated by this analysis. For more details on the healthy user study, the reader is referred to \cite{15}}.

\textbf{Hypothesis 2 (Task categorization):} Our task discrimination into easy and hard levels seems appropriate in TTC: \textit{Z=-3.0854, p<0.05 and NOC: Z=-3.4327, p<0.05}. This implies that the combined subjects (across both groups) completed easy tasks quicker and with less effort than they did the hard tasks. Within Cohort C, a signi\textsuperscript{cant difference in TTC: Z=-2.8925, p<0.05}. This is important for post-SCI users because high injury (C3-C4) patients will only use MJ but they will not be disadvantaged by using that. Since most post-SCI users do not have signi\textsuperscript{cant movement dexterity}, TS was not considered usable by these users. Hence, our assignment of two feasible human computer interfaces (TJ and MJ) for this study was validated by this analysis. For more details on the healthy user study, the reader is referred to \cite{15}}.

\textsuperscript{2}Originally used by Blumer \cite{19}, sensitizing concepts are constructs or organizing ideas that guide the qualitative researchers in their analysis \cite{20}.
difference in both TTC: $Z=-1.4067$, $p>0.05$ and NOC: $Z=-0.0514$, $p>0.05$. Since the performance of Cohort A was mainly driven by the effectiveness of the system autonomy, this implies that the easy and hard categorization is not meaningful for computer-based control, i.e., initial pose between robot and object does not affect the autonomous system performance.

**Hypothesis 3 (Quantitative metrics versus Baseline characteristics):** As previously stated, various baseline characteristics (MMSE, MVPT-R, and FIM self care subscale) were measured before the start of the study. For both Auto and Cartesian mode users, we did not find any significant correlation between performance and FIM self care subscale or MMSE scores. As shown in Fig. 3, the Pearson correlation coefficient (PMCC) values between MVPT-R and TTC/NOC for Cohort A are low in value as well as inconsistent in the direction of the correlation $\mu$ this is expected because the autonomous mode relies heavily on the visual perceptual ability of the underlying computational algorithms and is not heavily user dependent. However, higher correlation values and consistent inverse correlation for TTC and NOC can be seen for Cohort C. Even though these values are not statistically significant enough to reject the null hypothesis and claim inverse correlation, the size and directional consistency of $\rho$ values indicates need for further data collection and analysis. It can be seen that the visual spatial relationship (which is a component of the MVPT-R score) understanding by the subject is poor. It was especially seen in multiple cases where subjects were confused between yaw and roll motions which resulted in an increase both in the TTC and NOC metrics.

**Hypothesis 4 (Quantitative metrics versus Psychometrics):** It is obvious that Cohort A performed given ADL tasks quickly and with less number of commands than Cohort C. Hence, by the statement of Hypothesis 4, we expected that the satisfaction degree of Cohort A would be higher than that of Cohort C. Even though it is seen that the overall impact of the assistive technology is highly positive for both Cohort A and Cohort C, it is clear to see from the last column in Table I that the mean PIADS score is higher for Cohort C than Cohort A, i.e., Cohort C is seen to be more satisfied with the limited assistance provided by the robot during manual operation. In other words, even though Auto mode is quicker and easier to use than the Cartesian mode, participants in Cohort C are more satisfied with what they can do independently with the help of UCF-MANUS. Thus, Hypothesis 4 is rejected in the sense that the increase in mean performance (as measured by TTC and NOC) does not lead to an increase in mean user satisfaction (as measured by PIADS).

When Pearson correlation coefficients were computed between Mean PIADS score and the quantitative performance metrics, it was seen that Cohort C showed little or no correlation, i.e., as can be seen in Fig. 4, the user satisfaction metric does not deteriorate much even when user performance declines substantially. On the other hand, user satisfaction is very sensitive to decline in performance of the autonomous system. As seen in Fig. 4, NOC is significantly inversely correlated with satisfaction while TTC shows a large value of $\rho$ even though it is not considered statistically significant. Thus, in an average sense, potential users of assistive technology have less tolerance for decline in autonomous system performance than they have for decline in their own performance. Furthermore, it can be seen from the intersection of the lines in both Fig. 4(A) and Fig. 4(B) that satisfaction is identical for Cohort A and Cohort C at near the peak of the autonomous system performance.

---

3Previously, a significantly stronger inverse correlation has been reported in a study with MS patients [23] when the mean MVPT score was lower.
TABLE I
MEAN AND STANDARD DEVIATION OF PIADS SCORES FOR COHORT A AND COHORT C
Competence Adaptability Self-esteem Mean
A 1.97 (0.60) 2.13 (0.81) 1.90 (0.77) 2.00 (0.67)
C 2.18 (0.45) 2.70 (0.22) 1.70 (0.29) 2.19 (0.22)

Fig. 4. PMCC analysis between TTC/NOC and mean PIADS scores: Blue-colored dots and solid lines denote data and fitting from Cohort A while red-colored dots and dashed lines denote data and fitting from Cohort C.

V. CONCLUDING REMARKS
In this paper, we have presented a three-week user study with UCF-MANUS to help individuals post-SCI to perform pick-and-place tasks with six items on bi-level shelves. Based on the pre-evaluation assessment by OTs, none of the ten subjects could perform grasping tasks without help from a caregiver. After a three-week training/testing period, subjects were able to manipulate the robotic arm to perform ADL tasks with speed and command efficacy. Compared with manual (Cartesian) control mode, Auto mode was seen to enable the users to perform the given tasks faster and with lesser effort, however, the manual mode operation was perceived to be better by the users. For both manual and auto modes, users felt that UCF-MANUS would improve their functional abilities, quality of life and overall well being. During semi-structured exit interviews, users felt that they could benefit from the robot’s autonomy, however, they indicated satisfaction with being in charge during the interactive manual mode operation. It can therefore be concluded that the autonomy provided by UCF-MANUS or any other assistive technology needs to be appropriately channelled so that user satisfaction can be enhanced at the same time as their objective performance. Since there is great variability in performance of populations with disabilities as compared to healthy individuals, flexible interfaces need to be designed that are capable of providing a tailored amount of feedback to the user based on an estimate of the specific bottlenecks in their performance.

ACKNOWLEDGEMENTS
The authors are grateful to the volunteer subjects, Joshua Treadway of UCF, Holly Yanco and Kate Tsui of U. Massachusetts Lowell, Robert G. Melia of OHRI, the staff at Exact Dynamics, and the management team at OHRI. This study was supported by NSF under grant IIS-0649736.

REFERENCES