Improving Robotic Stroke Rehabilitation by Incorporating Neural Intent Detection: Preliminary Results from a Clinical Trial

Jennifer L. Sullivan¹, Nikunj A. Bhagat², Nuray Yozbatiran³, Ruta Paranjape³, Colin G. Losey¹, Robert G. Grossman⁴, Jose L. Contreras-Vidal², Gerard E. Francisco³, Marcia K. O’Malley¹

Abstract—This paper presents the preliminary findings of a multi-year clinical study evaluating the effectiveness of adding a brain-machine interface (BMI) to the MAHI-Exo II, a robotic upper limb exoskeleton, for elbow flexion/extension rehabilitation in chronic stroke survivors. The BMI was used to trigger robot motion when movement intention was detected from subjects’ neural signals, thus requiring that subjects be mentally engaged during robotic therapy. The first six subjects to complete the program have shown improvements in both Fugl-Meyer Upper-Extremity scores as well as in kinematic movement quality measures that relate to movement planning, coordination, and control. These results are encouraging and suggest that increasing subject engagement during therapy through the addition of an intent-detecting BMI enhances the effectiveness of standard robotic rehabilitation.

I. INTRODUCTION

Stroke is one of the leading causes of permanent disability in the United States [1]. Fortunately, rehabilitation research has shown that it is possible for stroke survivors, even those who are well into the chronic stage, to make motor improvements with continued physical therapy [2], [3], [4], [5]. However, chronic-stage rehabilitation seems to be most effective in inducing lasting neuroplastic changes when it is intensive, comprising a high number of effortful repetitions at an appropriate level of difficulty [3], [5], [6], [7].

This insight has led to the increased use of robots as rehabilitation tools due to their suitability for high repetitions, precise measurement capabilities, and versatility in programming [5]. They can also provide large assistance forces (or resistance forces, depending on nature of the desired exercise), that would be physically burdensome for a human therapist. More recently, assist-as-needed controllers have been developed that can modulate the amount of assistance provided by the robot, in real time, depending on the patient’s physical capability [6], [8], [9], [10]. This design more closely mimics the nature of rehabilitation provided by human therapists, who have the pathological expertise and patient specific-knowledge to adjust the intensity of the therapy as appropriate [11]. However, many of these control algorithms are based on signals related to physical exertion, such as force, movement speed, or electromyography (EMG). The downside to relying solely on physical exertion signals is that they do not necessarily correlate with effort; while a decrease in the exertion signal can be indicative of fatigue and a need for additional assistance, it might also be an indication that the patient is “slacking,” i.e. relying too much on the robot’s assistance [10]. Furthermore, even if the patient is contributing an acceptable level of physical effort, they still might not be focused on the task or concentrating on the movement. This lack of mental engagement diminishes the effectiveness of the therapy exercise [5], [7], [12].

The most reliable way to ensure mental engagement is to measure neural signals directly. A number of research groups have explored the use of non-invasive electroencephalography (EEG) as a measure of subjects’ mental effort and have been successful in both detecting movement intention and distinguishing it from a rest state [13], [14], [15], [16], [17], [18], [19]. These intent-detection classifiers have been implemented in brain-machine interfaces (BMIs) as triggers for robot motion; that is, the robot is only activated once the user has made a conscious intention to move [16], [17], [18], [19], [20]. Although small-scale studies have been able to validate this concept by demonstrating high classification accuracy of the intent detection algorithms, more longitudinal clinical trials are needed to assess its actual efficacy for rehabilitating chronic stroke survivors (see [7] and [14] for more thorough reviews).

The study presented here is based on the framework proposed by Blank et al. [20], and the feasibility studies published by Bhagat et al. [16], [17]. In this paper, we present the preliminary findings of a 12-session clinical trial (NCT 01948739) evaluating the effectiveness of a BMI-exoskeleton system for elbow flexion/extension rehabilitation in chronic stroke survivors.

*This work is supported by NIH grant R01NS081854
¹J.L.S (jls3@rice.edu), C.G.L, and M.K.O are with the Dept of Mechanical Eng, Rice Univ.; ²N.A.B and J.L.C-V are with the Dept of Electrical Eng, Univ. of Houston; ³N.Y., R.P., and G.E.F. are with TIRR Memorial Hermann and Univ. of Texas Health Sciences Center; ⁴R.G.G is with the Houston Methodist Research Institute. All institutions in Houston, TX.

![BMI-Exo system setup.](image1)

![Target-hitting task displayed on monitor](image2)

Fig. 1. Left: BMI-Exo system setup. Right: Target-hitting task displayed on monitor (both targets shown for reference).
stroke survivors. The BMI employs a noninvasive, EEG-based, movement-intent detector that triggers an upper limb exoskeleton, the MAHI Exo-II, to guide the subject through a passive elbow movement. The close temporal proximity of conscious movement intention to sensory feedback associated with smooth, coordinated motion is intended to activate Hebbian mechanisms that strengthen the appropriate neural pathways and activation timings [21], [22]. Motor improvements were assessed in terms of both movement quality and functional ability using kinematic metrics and the Upper-Extremity portion of the Fugl-Meyer Assessment (FM-UE).

II. METHODS

A. System Description

The MAHI Exo-II system is a 5-DOF upper-limb robotic exoskeleton that provides motor-actuated movement in elbow flexion/extension, forearm pronation/supination, and wrist flexion/extension as well as radial/ulnar deviation. The design also allows for passive positioning of shoulder abduction angle. For this study, the wrist and forearm module was removed, and only elbow flexion/extension was trained (Fig. 1). The elbow joint has a range of motion of 60 degrees and is actuated by a brushed DC motor. A counterweight provides passive gravity compensation for the weight of the user’s arm. Detailed descriptions of the MAHI Exo-II are provided in [23] and [24].

Four usage modes are available on the MAHI Exo-II: user-passive, triggered, backdrive, and active constrained. In passive mode, the movement is entirely motor-actuated so that the exoskeleton guides the user through flexion/extension motion. In triggered mode, the user must provide an initial push to exceed a pre-set threshold before the motor takes over and guides the user through the rest of the motion. In backdrive mode, the motor is disabled and the user moves the exoskeleton arm without assistance. In active constrained mode, the motor provides a scalable resistance force modeled as a viscous force field. These four modes were designed to accommodate a broad range of impairment levels, from full paralysis to near-healthy function.

B. Movement Intent Detection

Intent detection was predicted with an EMG-gated BMI system. A 64-channel, noninvasive EEG cap (actiCAP system, Brain Products GmbH, Germany) was used to monitor neural signals — specifically, slow movement-related cortical potentials (MRCPs). MRCPs have been shown to be effective in detecting volitional movement in stroke subjects using the MAHI Exo-II with an accuracy of approximately 78% [16], [17], [25] as well as in other contexts [15], [26], [27]. Surface EMG was collected from the biceps brachii and triceps brachii of both arms. EMG signals from the affected arm were used as additional inputs to the intent-detection algorithm to minimize false positives of the BMI system [14], [17]. EEG and EMG were sampled at 500 Hz and synchronized. See [17] for a detailed description of the BMI system and intent-detection algorithm.

C. Task

A target-hitting task was displayed on a computer monitor positioned in front of the user. For elbow flexion/extension, targets were arranged vertically, as shown in Fig. 1, where the upper target corresponded to flexion and the lower to extension. Each trial required the subject to start from the midpoint of the pre-defined range of motion, move to whichever target was presented, and then return back to center. Equal numbers of flexion and extension trials, in random order, were performed in each block.

To enforce mental movement planning in backdrive mode, subjects were instructed to pause for a few moments and think about their arm motion before moving towards the displayed target. If they moved too soon, the robot would block their movement with a virtual wall, reset back to the center position, and the subject would redo the trial. Subjects were not told how long they needed to wait in between trials, but the software was programmed to choose a random value between 1.75 and 2.25 seconds [15].

D. Participants

Six chronic-stage hemiparetic stroke survivors have participated in the study. Inclusion criteria required that participants have suffered only a single stroke, have sufficient proprioception in the affected upper limb, and not be participating in any other therapy program. IRB approval was granted at all collaborating institutions, and all participants provided written consent. Individual subject information is provided in Table I.

E. Protocol

Before beginning therapy on the BMI-Exo system, each subject underwent two clinical baseline assessment sessions, approximately one month apart, with a physical therapist. Provided there was minimal change in FM-UE score (requirement: difference within ± 3 points; actual average: +0.3 points), robotic therapy commenced within a week of the second baseline session. After completing robotic therapy, subjects returned within one week for a post-treatment clinical assessment. Data were analyzed using the second baseline and one-week follow-up as pre- and post-treatment scores, respectively.

The five-week therapy program consisted of two calibration sessions followed by 12 training sessions. In each
TABLE II
AVERAGE (STANDARD DEVIATION) CLINICAL AND MOVEMENT QUALITY SCORE CHANGES. (+): IMPROVEMENT INDICATED BY SCORE INCREASE; (-): IMPROVEMENT INDICATED BY SCORE DECREASE. BOLD VALUES INDICATE STATISTICAL SIGNIFICANCE.

<table>
<thead>
<tr>
<th></th>
<th>Fugl-Meyer</th>
<th>Trial Duration (-)</th>
<th>Number of Peaks (-)</th>
<th>Time to First Peak (+)</th>
<th>MJ Smoothness (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Flex</td>
<td>Ext</td>
<td>Flex</td>
<td>Ext</td>
</tr>
<tr>
<td>Baseline</td>
<td>39.0</td>
<td>1.43</td>
<td>1.72</td>
<td>1.56</td>
<td>2.15</td>
</tr>
<tr>
<td></td>
<td>(12.4)</td>
<td>(0.36)</td>
<td>(0.54)</td>
<td>(0.29)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>End</td>
<td>42.3</td>
<td>1.20</td>
<td>1.06</td>
<td>1.38</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>(13.4)</td>
<td>(0.42)</td>
<td>(0.32)</td>
<td>(0.38)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>Change</td>
<td>3.3</td>
<td>-0.23</td>
<td>-0.66</td>
<td>-0.18</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td>(3.1)</td>
<td>(0.63)</td>
<td>(0.41)</td>
<td>(0.26)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>P-value</td>
<td>0.048</td>
<td>0.438</td>
<td>0.031</td>
<td>0.312</td>
<td>0.031</td>
</tr>
</tbody>
</table>

F. Data Analysis

Elbow angle position data were collected at 1000 Hz with a high-resolution encoder. Velocity was calculated and low-pass filtered at 50 Hz in real time from the position data. In post-processing, backdrive trials were segmented into flexion (center to upper target) and extension (center to lower target) movements, which were analyzed separately. Movements from the outer targets back to the center position were considered “reset” movements and excluded from the analysis. A Savitzky-Golay filter (3rd order polynomial, window size of 101) was applied to the velocity data to filter out high-frequency noise without removing the natural characteristics of the subjects’ movements. The smoothed velocity was then used to calculate four movement quality metrics:

**Trial Duration:** Time (in seconds) for the subject to move from center target to the outer target. Since the task in this study does not require high precision, a decrease in Trial Duration indicates improvement.

**Minimum-Jerk Smoothness:** The minimum-jerk (MJ) velocity profile is a smooth, symmetric, bell-shaped curve that closely matches the velocity profiles of healthy point-to-point movements [3], [28]. The MJ velocity profile is defined as

\[ v_{mj}(t) = \Delta \left( \frac{30t^4}{T^5} - \frac{60t^3}{T^4} + \frac{30t^2}{T^3} \right) \]  

where \( t \) is time, \( T \) is the duration of the movement, and \( \Delta \) is the total distance traveled. Its formulation is based on the assumption that the underlying objective of the neuro-motor control system is to minimize squared jerk, the time derivative of acceleration [29]. The metric is calculated as the correlation coefficient (\( \rho \)) between the subject’s velocity profile and the corresponding MJ velocity profile, given the same distance traveled (\( \Delta \)) and Trial Duration (\( T \)) [30], [31]. Values generally range from 0 to 1, where 0 is no correlation and 1 is perfect correlation. Negative values (0 to -1) occur when the subject’s velocity profile bears more resemblance to a concave-up 4th-order polynomial than concave-down (bell-shaped). An increase in MJ Smoothness corresponds to improvement.

**Number of Peaks:** Number of velocity peaks, where a “peak” needed to be at least (peak speed)/4 larger than the surrounding data to be counted [32]. Since velocity profiles for impaired movements are often fragmented with many peaks, a decrease in the Number of Peaks metric indicates improvement.

**Time to First Peak:** Time elapsed from movement start to the first velocity peak, as a percentage of total Trial Duration. As discussed, healthy velocity profiles are typically single-peaked and symmetric, and therefore have a Time to First Peak value of approximately 0.5 (50%). Impaired, fragmented velocity profiles often have multiple velocity peaks, where Time to First Peak is less than 0.5. Thus, an increase in Time to First Peak indicates improvement.

Statistical analysis was done on within-subject changes in FM score and in movement quality metrics from beginning to end of treatment. Baseline movement quality metrics for each subject were calculated by averaging movement quality...
scores from the two robotic calibration sessions. End-of-treatment scores were calculated by averaging the movement quality scores from the assessment blocks of the last two training sessions (11 & 12). Due to slight non-normality that was evident from quantile-quantile plots, Wilcoxon Signed-Rank tests were used on the paired differences for each movement quality metric in flexion and in extension. Pre- and post-treatment FM scores were compared using a paired t-test to evaluate clinical improvement. For all difference testing, $p$-values less than 0.05 were considered significant. Linear correlations were also calculated between pre/post FM scores and starting/ending movement quality scores (flexion and extension scores averaged within-subject).

III. RESULTS

Table II shows average score changes, with associated $p$-values, across all subjects for the FM assessment and for each movement quality metric.

A. Clinical Improvements

Functional gains from clinical assessments are shown in Fig. 2. FM score increases ranged from -1 to 8 out of 66, with an average of 3.3. This increase was statistically significant ($t(5) = 2.6$, $p = 0.048$). Moreover, there were some interesting trends in the specific sections and items where subjects showed improvement. Four out of six subjects (S1, S3, S5, S6) had lower impairment levels, with baseline FM scores above 40. These higher-functioning (HF) individuals tended to have score gains in section A: Shoulder/Elbow/Forearm, specifically in “volitional movement within synergies” and “volitional movement mixing synergies,” and in section B: Wrist. Overall, the HF subjects had larger functional score increases than the two lower-functioning (LF) individuals (S2 and S4; baseline FM scores less than 30): a 4.0-point increase for HF versus 2.0 points for LF. The two LF individuals only made improvements in section D: Coordination/Speed.

B. Movement Quality Improvements

Baseline and end-of-treatment movement quality scores are shown in Fig. 3. Score values tended to be lower for extension than for flexion, but score improvements were on average higher in extension than in flexion (see Table II). Results of the Wilcoxon Signed-Rank tests showed statistically significant improvement in all movement quality metrics in extension ($p = 0.03$), but not in flexion. However, group statistics were skewed by S4’s flexion scores, which was the only set of scores that got worse. Her extension scores, on the other hand, were comparable to the rest of the group’s. The other five subjects showed score improvements in all metrics, and in similar amounts between flexion and extension. S2, who had the lowest baseline movement quality scores, showed particularly large score improvements. Ceil-
ing effects were evident in Trial Duration, Number of Peaks, and Time to First Peak for exceptionally high-functioning subjects.

C. Correlations

Strong relationships were found between FM scores and MJ Smoothness, Time to First Peak, and Number of Peaks ($R^2 = 0.56, 0.54, 0.51$, Fig. 4). No relationship was found between Trial Duration and FM score. For Trial Duration, baseline scores were more strongly correlated with score improvement ($R^2 = 0.58$); that is, subjects with the largest (worst) baseline scores made the largest improvements.

IV. DISCUSSION

Subjects showed impressive functional and movement quality gains for a 5-week therapy program. Given the very low number of active repetitions (20 backdrive trials/session, 3 sessions/week), it was expected that subjects with higher impairment levels would benefit more from this treatment protocol than the subjects with lower impairment levels.

However, this did not seem to be the case for the clinical assessments, as FM score increases tended to be larger for the HF subjects than for the LF subjects (average increase of 3.8 versus 2.0 points, respectively). Furthermore, the HF subjects tended to improve in section A: Shoulder/Elbow/Forearm, and in section B: Wrist, where their baseline scores were already fairly high (24.8/36 and 7.5/10, respectively). While this seems logical given that the therapy involved elbow flexion/extension, 3 of the 4 HF subjects had already scored the maximum number of points for the elbow flexion/extension items at baseline, so their improvements were actually in items related to shoulder and forearm function. This trend is encouraging, suggesting that individuals can make functional gains even if they are not directly related to the specific rehabilitation task performed.

Interestingly, the item in which subjects made the most substantial improvements was dysmetria (section D: Coordination/Speed). Dysmetria refers to a movement-planning impairment demonstrated by the tendency to overshoot or undershoot a target in point-to-point movements. The FM dysmetria item is scored from 0 to 2, where 0 = “pronounced or unsystematic”, 1 = “slight and systematic”, and 2 = “no dysmetria.” Baseline dysmetria scores across all six subjects were low: four (S3, S5, S2, S4) had baseline scores of 0, and two (S1, S6) had baseline scores of 1. Post-treatment, however, all four of the individuals with baseline scores of 0 improved: two increased by 1 point (S2, S3), and two increased by 2 points (S5, S4). In other words, of the four subjects who started with “pronounced or unsystematic” dysmetria, after only 5 weeks, half of them improved to “slight and systematic” dysmetria, and the other half improved all the way to “no dysmetria.” Although our sample size is still quite small, this finding is encouraging in its implication of the effectiveness of the movement planning and mental engagement aspects of this therapy protocol.

The results of the movement quality measures do somewhat support the idea that this therapy was more effective for subjects with higher impairment levels. S2, who showed the largest movement quality score improvements overall, was also the individual with the lowest baseline movement quality scores, especially in extension. This is likely due, in part, to the fact that she had the most room for improvement, whereas ceiling effects were evident for especially high-functioning subjects (e.g., S1). This was not a limiting factor across the board, though, as correlations between baseline movement quality scores and score improvements were low for MJ Smoothness, Time to First Peak, and Number of Peaks.

Although their improvements were more moderate than for S2, our subjects overall showed movement quality gains in both flexion and extension. However, the score increases in extension were statistically significant and larger than the increases in flexion scores. As our subjects’ baseline movement quality scores were lower for extension than for flexion, this inconsistency seems to provide additional support for the idea that this therapy protocol is particularly effective for more highly-impaired movements. That said, the group statistics were strongly skewed by S4’s flexion scores, which were the only set of scores that got worse. These decreases could simply be anomalies attributable to fatigue or unusually high flexor muscle tone in the last two sessions, as S4 maintained her usual work schedule during the study and had a long commute to the research center. Removing S4 from the group statistics produces marginally significant p-values ($p = 0.06$).

Strong correlations were found between FM scores and average flexion/extension movement quality scores for Number
of Peaks, Time to First Peak, and MJ Smoothness. Celik et al. [30] also found a strong relationship between FM scores and MJ Smoothness, even with a slightly narrower range of FM scores (most greater than 35). It is unsurprising that similar trends would be found for MJ Smoothness, Time to First Peak and Number of Peaks, as they all assess the shape of the velocity profile but with varying degrees of coarseness. Currently, our data is very bimodal due to the large difference in scores between the HF and LF subjects, so it will be interesting to see how strongly that relationship persists as more subjects complete the therapy.

V. CONCLUSION

This paper presents the preliminary findings of a multi-year clinical study evaluating the effectiveness of a BMI-exoskeleton system for elbow flexion/extension rehabilitation in chronic stroke survivors. The first six subjects to complete the program have shown improvements in both FM-UE scores as well as in movement quality measures that relate to movement planning, coordination, and control. These improvements were made in spite of the fact that subjects only performed approximately 60 active movement repetitions per week, which would normally be considered nowhere near sufficient for inducing neuroplastic changes. These preliminary results are promising, and suggest that increasing subject engagement during therapy through the addition of an intent-detecting BMI can enhance the effectiveness of standard robotic rehabilitation.

REFERENCES
