Supplementary Material

Power Analysis

Power analysis was based on a pilot study with the MusicGlove, a wearable sensor that also allows people to train finger movement using a Guitar-Hero-like musical computer game [1]. In that study we measured an effect size of 1.2 with the Box and Blocks test after repetitive finger training compared to conventional tabletop training. Assuming a similar effect of robotic assistance, at a power of 0.9 with alpha = 0.05, the power analysis predicted 15 people per group (two-sided t-test).

Randomization

To randomize participants into the high and low assist groups we used a block allocation procedure which grouped participants according to age (50-60, 61-70, 71-80) and Box and Blocks score (3-10, 11-20, 21-60). The first two participants in each bin were randomly assigned to success levels by drawing a piece of paper from an envelope. Addition participants were assigned to maintain balance between the two groups within each bin.

Robotic training procedure

FINGER uses a pair of independently-actuated 8-bar mechanisms to assist the index and middle fingers of the hand in moving through a natural curling motion [2]. In this study we used the FINGER robot to assist participants in playing a musical game in the style of Guitar Hero. During each training session, participants played five songs two times each, for a total of 1065 possible movements per session. In each song, note objects corresponding to the beat and general inflexion of the melody scrolled across the screen in the direction of finger flexion (e.g. left to right for left hand impairments) (Figure 6). These notes were distributed across three rows and moved towards color-coordinated targets located on the opposite side of the screen. Notes in the top row were green and corresponded to flexion of the index (top) finger. Notes in the bottom row were yellow and corresponded to flexion of the middle (bottom) finger. Notes in the middle row were blue and corresponded to movement of both fingers together. In addition to the notes and the targets, each row also contained a ball indicating the position of the row's corresponding finger. Participants were instructed to try to hit each note by moving the corresponding ball so as to stop inside of the target at the same moment that the note passed through the target. The size of the target (referred to as the hit window) corresponded to a physical distance of 20% of the robot's range of motion, which was about 15% of the finger's range of motion. When participants hit a note correctly, the game played a rewarding animation of the note exploding to provide feedback to the participant. After attempting to hit a note, participants were instructed to move their fingers back behind a green line corresponding to 80% of full extension. To discourage participants from simply slamming their fingers into the hard-stop at end of the robot, the targets were placed at 80% of the robot's full range of motion. This placed the hard stop of the robot just outside of the target's hit window, thereby requiring participants to make controlled, accurate movements rather than slamming movements.

Finger movements were performed in the horizontal plane in a way consistent with the computer graphics. No rotational mapping was required. The index finger always controlled the top ball, and the middle finger always controlled the bottom ball. For participants with more severe impairments on the left side, the targets moved from right to left, and for participants with more severe impairments on the right side the targets moved from left to right. The monitor was placed relative to the robot to make the connection between finger movements and the screen as obvious as possible.

As participants played the game, the FINGER robot provided assistive forces that guided the fingers along desired spatiotemporal trajectories. However, the robot only provided these forces if the participants initiated the movements themselves. We detected intent to move by monitoring the resistive forces of our actuators. When the resistive forces exceeded 1.75 times their measured noise threshold (equivalent to a force of ~6N) the robot would generate an ideal trajectory and begin the new movement. This threshold was not so high as to be prohibitive, but was high enough to create a subtle “catch” that made it clear that the participants themselves had triggered the movements. As soon as a movement was triggered, the robot calculated a naturalistic trajectory using a minimum-jerk profile designed to move the finger from its starting position to the target position at the right time to hit the note. A mirrored trajectory was programmed to guide the finger back to full extension. If a participant tried to initiate a movement when there was no upcoming note corresponding to that movement, the robot did not change its desired position but rather resisted the erroneous movement. To determine if there was an upcoming note corresponding to an attempted movement and to prevent movements from overlapping, we restricted the maximum flexion duration to 0.8 seconds. To prevent the robot from moving at speeds that might surprise participants, we also defined a minimum flexion duration of 0.15 seconds.

Assistive forces guiding participants toward their desired trajectories were calculated using a proportional-derivative (PD) position controller, and the gains for the controller were selected adaptively to guide participants toward a desired success rate. The algorithm used to control success rate has been described in a previous publication [2], but briefly, it decreases the compliance of the robot following a “miss” and increases the compliance following a “hit”. Participants in the high assistance group were guided towards a desired success rate of 85%, and participants in the low assistance group were guided towards a success rate of 50%. The algorithm adapted gains separately for each finger and for both flexion and extension (i.e. four sets of gains). However, we only turned the adaptive gain selection on during songs played on the first visit of each week. During subsequent visits that week, the gains remained at the final value from the previous visit, giving the participant two sessions to improve their scores without the robot increasing or decreasing its assistance to compensate. Because the last song in the set was significantly more difficult than the other songs (0.9 notes per second for the last song versus an average of 0.5 notes per second for the other songs), the gain adaptation was not turned on for that song. Once per week the participants also played one song without assistance to quantify their ability to move their fingers without assistance.

Outcome Measures

A single, trained, blinded rehabilitation therapist performed all clinical assessments.. Since training focused on finger movement, we chose the Box and Blocks Test [3] as the primary outcome measure because it requires finger movement to pick up the blocks and is a measure of hand function. Another possibility was the Nine Hole Peg Test, but that test requires finer finger movement and thus has a greater floor effect. The secondary outcomes were the Fugl-Meyer test [4] and Action Research Arm Test [5]. In addition, we performed the Nine Hole Peg Test, Lateral and Three Jaw Pinch Strength measurements, the NIH Stroke Scale, Finger Tap Test, Visual Analog Pain Scale, Modified Ashworth Scale of Spasticity, Motor Activity Log, Geriatric Depression Scale, and Beck Hopelessness Scale.

The Intrinsic Motivation Inventory (IMI) is a multi-dimensional assessment designed to measure six different variables related to motivation [6]. After each robotic therapy session the participants completed a subset of the Intrinsic Motivation Index questionnaire chosen for their relevance. A list of the included statements and their corresponding subscales is given in Supplemental Table 1. Participants were instructed to evaluate the truth of each statement on a scale from 1 to 7 as pertaining to the training given by the FINGER robot. Scores were reversed for questions phrased as negatives.

At the end of each week of training, we also assessed the self-efficacy of the participants with regards to their hand function using a standardized procedure derived from the self-efficacy assessment literature [7,8]. Specifically, we measured the participant's self-efficacy with respect to the Box and Blocks test by asking them to report their confidence (on a scale from 0% to 100% in increments of 10%) that they could score at least 4, 6, 8, 10 13, 20, 30, 40, and 50 blocks.

We queried the participants weekly in an open-ended manner on whether they were able to achieve new tasks or execute existing tasks better with their impaired hand. They also rated their enjoyment and motivation for the robotic training on a 5 point Likert scale.

Finger Proprioception Measure

We measured finger proprioception at baseline, post therapy, and at the one month follow up using a novel procedure that we recently found to be sensitive to changes in proprioception with aging [9]. Briefly, FINGER slowly moved (13 deg/s) the index and middle fingers in a series of 12 finger-crossing movements through flexion/extension at the metacarpophalangeal (MCP) along separate trajectories until they crossed each other; the trajectories were randomized in direction and amplitude. With visual feedback of their hand occluded, participants were asked to press a key when they sensed their two fingers exactly overlapped one other in terms of degree of MCP flexion. Error was defined as the angular distance between the MCP joints once the key was pressed. Proprioceptive error reported here is the average, absolute error for each participant.

Data Analysis

We tested the significance of time, group (high assistance vs low assistance), and the timeXgroup interaction on all the outcomes using linear mixed-effect models (LME). The model allowed random intercepts for each participant to account for the fact that the participants spanned a wide range of impairment levels.

After fitting each model, we performed model comparison analyses by using the Bayesian Information Criterion (BIC) [10] to prune away terms that substantially reduced the fit of the model (as marked by an increase in the BIC score of 5 or more points). To verify that the normality requirements for the tests were being satisfied we used quantile-quantile (Q-Q) plots to graphically compare our model residuals to normally distributed data. Data for models that did not natively yield normally distributed residuals were transformed towards normality using Box-Cox power transformations. However, a few of the outcomes could not be corrected to yield normal residuals. For these outcome measures we used Friedman tests to evaluate the effect of time on the outcome. All statistical analyses were performed in R, and used two-tailed methods with alpha = 0.05.

To obtain a single, generic variable describing the participant’s motivation we first normalized the IMI data and then used principal component analysis to reduce the full IMI dataset to its first principal component. The effect of this procedure was to define a measure that exhibited the most variance across participants with respect to the individual IMI questions. We then used a linear mixed model to study the effects of the high and low assist training on the first principle component. In this case, in addition to time, group and the time-group interaction, we added baseline Box and Blocks score as a potential factor. Note that the IMI data was collected after training began. Thus, a significant effect on the group factor would indicate that training at high versus low assistance levels yielded a significant difference in the subscale being tested.

We quantified self-efficacy by fitting a sigmoidal curve to the confidence estimates for each participant regarding the weekly Box and Blocks test. Using this curve, we then estimated the number of blocks that each participant was 50% confident he or she could achieve, a procedure developed previously in other settings to obtain a single number quantifying self-efficacy [7,8].

Subjective Outcomes

On average, participants reported achieving three new tasks and performing three existing tasks better by the end of training, and rated the training highly. The reported tasks typically included holding and manipulating objects (e.g. during cooking, eating, and drinking), donning clothing (e.g. fastening buttons, shoe tying), and grooming (e.g. combing hair, brushing teeth, showering). They rated their enjoyment and motivation for the robotic training on a Likert scale at 4.8 and 4.7 out of 5, respectively. Participants with lower baseline Box and Blocks scores tended to rate their improvement in hand movement lower at the one month follow-up, but this trend did not reach significance (BB score < 20 mean rating = 3.4, BB score > 20 mean rating = 4.1 out of 5, t-test, $p = 0.08$). Participation in the high versus low assistance group did not significantly influence any of these ratings.

Change in finger proprioception over time

We used FINGER to measure proprioceptive error at baseline, after therapy, and at the one month follow up using the method described above. To determine whether exercising in the FINGER robot improved finger proprioception, we used a linear, mixed effect model to test for the effect of time, group, and the timeXgroup. Although the average proprioceptive error of the population did not change significantly over time (LME $F(1/54)=0.42$, $p=0.52$), the timeXgroup interaction trended towards significance (LME $F(1/54)=3.62$, $p=0.0623$) with more reduction in error in the low assistance group (FIG-S1).

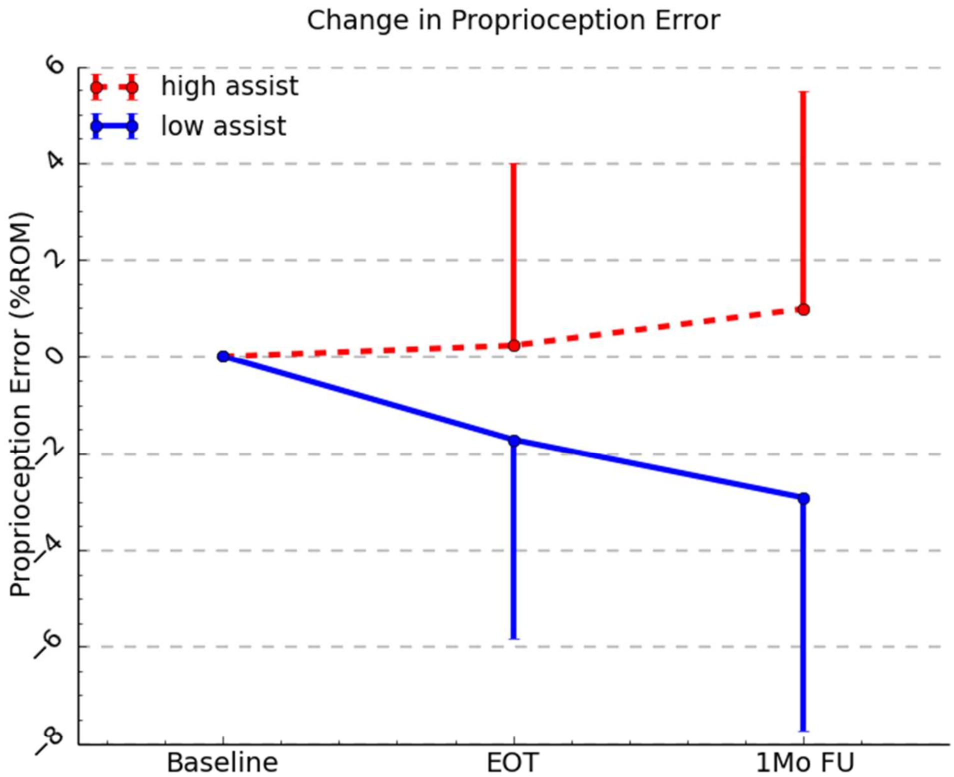

FIG-S 1 Change in proprioception error over the duration of the study. EOT represents the end of therapy time point, and 1Mo Fu represents the one month follow-up.

Although care should be taken in drawing conclusions off a non-significant trend, it is interesting that finger proprioception was the only outcome measure that favored the low-assistance group. At the very least, this makes it unlikely that the FM improvements of the high-assist group over the low-assist group were the result of greater improvements in their finger proprioception.

Could the greater FM improvement observed for participants with lower baseline FM be related to a ceiling effect?

To respond to this concern, we re-analyzed the change in Fugl-Meyer score at the post-therapy and one-month-follow-up evaluations with the participants near the celling removed (Fig-S 2). We set the cutoff for this analysis to be FM=58 (the max possible upper extremity FM score minus twice the average delta FM of the high assist group at follow-up). All the relationships that were significant before the participants near the ceiling removed were still significant afterwards (p=0.01 for high assist at PT, p=0.009 for low assist at PT, p=0.025 for high assist at 1MoFU, and p=0.186 for low assist at 1MoFU).

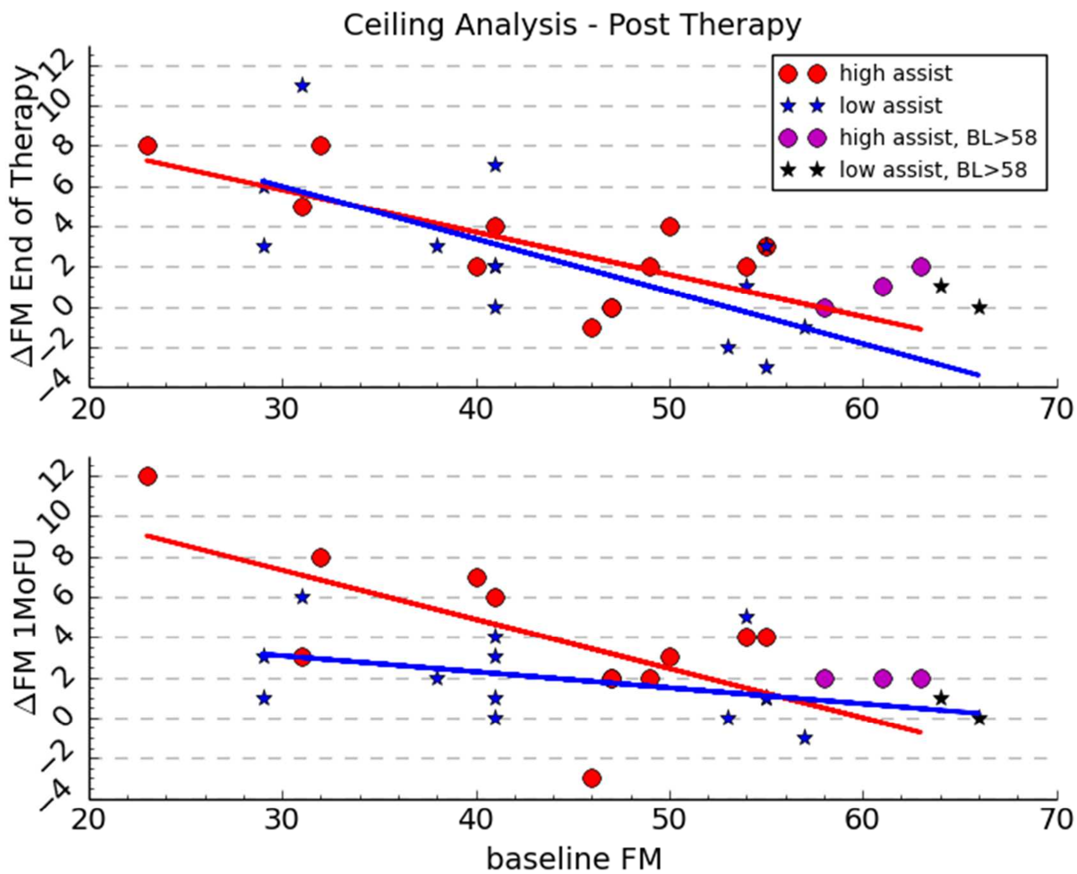

FIG-S 2 Change in Fugl-Meyer score versus baseline FM score. Regressions were performed without the five participants with the highest baseline FM scores (purple circles and black stars).

Fig-S 2 is a modified version of Figure 3 D-E in the main text, for which we built the regression lines using only the participants with baseline FM scores below 58. The figure still shows the participants with baseline scores greater than 58, but we plotted these participants in different colors to set them apart. Notice that none of the five participants are below the regression line. If a ceiling effect in the FM scale were preventing us from measuring recovery in these participants, and if that effect were the cause of the relationship that we observed, then one would expect these participants to be below the regression lines, artificially pulling the trend down. At the very least, one would expect these participants to remain at their baseline values. However, this was not the case.

Did proprioception at baseline influence improvement in the high-assistance group more than in the low-assistance group?

To respond to this question, we also ran a model that includes baseline proprioception, group, and the proprioceptionXgroup interaction as potential factors. Using this model, the effect due to proprioception was still significant (p=0.007), but the proprioceptionXgroup interaction was not (p=0.413).

References cited in the supplementary material

1. Taheri H, Rowe JB, Gardner D, et al. Design and preliminary evaluation of the FINGER rehabilitation robot: controlling challenge and quantifying finger individuation during musical computer game play. *J. Neuroeng. Rehabil.* 2014;11:10. doi:10.1186/1743-0003-11-10.

2. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the Box and Block Test of manual dexterity. *Am. J. Occup. Ther. Off. Publ. Am. Occup. Ther. Assoc.* 1985;39(6):386-91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3160243. Accessed March 21, 2012.

3. See J, Der-Yeghian L, Chou C, et al. A standardized approach to stroke outcome assessments. II. The Fugl-Meyer motor assessment. *Neurorehabilitation and Neural Repair,* 2013;27(8):732-741.

4. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair* 2008;22(1):78-79. doi:10.1177/1545968307305353.

5. Ryan RM. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory.

6. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change.

7. Bandura A. Guide For Constructing Self-Efficacy Scales. In: *Self Efficacy Beliefs of Adolescents*. Information Age Publishing; 2006:307-337. doi:10.1017/CBO9781107415324.004.

8. Ingemanson ML, Rowe JB, Chan V, Wolbrecht ET, Cramer SC, Reinkensmeyer DJ. Use of a robotic device to measure age-related decline in finger proprioception. *Exp. brain Res.* 2016;234(1):83-93. doi:10.1007/s00221-015-4440-4.

9. Schwarz G. Estimating the Dimension of a Model. *Ann. Stat.* 1978;6(2):461-464. Available at: http://projecteuclid.org/euclid.aos/1176344136. Accessed January 23, 2016.

Table S1 List of Songs the participants played twice per session

Song Name (Artist)	Number of notes	Duration (min)
Bad Moon Rising (Creedence Clearwater Revival)	48	2:21
Everyday (Buddy Holly)	58	2:09
Rocky Raccoon (The Beatles)	101	3:41
With or Without You (U2)	166	4:56
Joy to The World (Three Dog Night)	159	3:15

Table S2. A list of the Intrinsic Motivation Index (IMI) questions used for the FINGER study

Statement to be evaluated	Subscale	Reversed
I think that this activity was boring.	Interest	Yes
I enjoyed doing this activity very much.	Interest	No
I thought that this activity was quite enjoyable.	Interest	No
This activity did not hold my attention at all.	Interest	No
I would be willing to do this activity again because it has some value to me.	Value	No
I believe that doing this activity could be beneficial to me.	Value	No
I think this is an important activity.	Value	No
I felt very tense while doing this activity.	Pressure	No
I felt pressure while doing this activity.	Pressure	No
I was anxious while working on this activity.	Pressure	No
I was very relaxed in doing this task.	Pressure	Yes
I felt like I had to do this.	Choice	Yes
I did this activity because I wanted to.	Choice	No
I put a lot of effort into this.	Effort	No
It was important to me to do well at this task.	Effort	No
I did not try very hard to do well at this activity	Effort	Yes
I think I am pretty good at this activity.	Competence	No
After working on this activity for a while I felt pretty competent.	Competence	No
This was an activity that I could not do very well	Competence	Yes

Table S3. Statistical analysis of the change in the IMI sub-scales over the nine weeks of training. NI =Not Included (pruned based on BIC score)

IMI Sub-scale	Time	Group	Baseline	GrpXBaseline	High Assist	Low Assist
Pressure	P=0.172	P=0.123	NI	NI	2.2+/-1.3	3.1+/-1.6
Competence	P=0.833	P=0.235	P=0.054	NI	5.2+/-1.2	4.5+/-1.6
Value	P<0.001	P=0.358	P=0.811	P=0.503	6.6+/-0.8	6.3+/-1.1
Choice	P=0.197	P=0.122	NI	NI	6.1+/-1.4	5.3+/-1.5
Interest	P=0.052	P=0.851	P=0.152	P=0.605	6.1+/-1.1	5.9+/-1.2
Effort	P=0.006	P=0.002	P=0.99	NI	6.7+/-0.6	5.7+/-1.3
Motivation	P=0.027	P<0.001			-0.5+/-1.3	0.5+/-1.7