

Investigating the feasibility and acceptability of real-time visual feedback in reducing compensatory motions during self-administered stroke rehabilitation exercises: A pilot study with chronic stroke survivors

Shayne Lin¹ , Jotvarinder Mann^{2,3}, Avril Mansfield^{2,4,5}, Rosalie H Wang^{2,6}, Jocelyn E Harris⁷ and Babak Taati^{2,8,9}

Abstract

Introduction: Homework-based rehabilitation programs can help stroke survivors restore upper extremity function. However, compensatory motions can develop without therapist supervision, leading to sub-optimal recovery. We developed a visual feedback system using a live video feed or an avatar reflecting users' movements so users are aware of compensations. This pilot study aimed to evaluate validity (how well the avatar characterizes different types of compensations) and acceptability of the system.

Methods: Ten participants with chronic stroke performed upper-extremity exercises under three feedback conditions: none, video, and avatar. Validity was evaluated by comparing agreement on compensations annotated using video and avatar images. A usability survey was administered to participants after the experiment to obtain information on acceptability.

Results: There was substantial agreement between video and avatar images for shoulder elevation and hip extension (Cohen's κ : 0.6–0.8) and almost perfect agreement for trunk rotation and flexion (κ : 0.80–1). Acceptability was low due to lack of corrective prompts and occasional noise with the avatar display. Most participants suggested that an automatic compensation detection feature with visual and auditory cuing would improve the system.

Conclusion: The avatar characterized four types of compensations well. Future work will involve increasing sensitivity for shoulder elevation and implementing a method to detect compensations.

Keywords

Stroke rehabilitation, visual feedback, upper limb exercises, self-administered rehabilitation exercise, Graded Repetitive Arm Supplementary Program

Date received: 8 October 2018; accepted: 22 January 2019

Background

Up to 80% of strokes result in hemiparesis and 65% of people with chronic stroke have long-term motor

¹Division of Engineering Science, University of Toronto, Toronto, Canada

²Toronto Rehabilitation Institute, University Health Network, Toronto, Canada

³Department of Kinesiology, University of Waterloo, Waterloo, Canada

⁴Department of Physical Therapy, University of Toronto, Toronto, Canada

⁵Evaluative Clinical Sciences, Hurvitz Brain Sciences Research Program, Sunnybrook Research Institute, Toronto, Canada

⁶Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, Canada

⁷School of Rehabilitation Sciences, McMaster University, Hamilton, Canada

⁸Department of Computer Science, University of Toronto, Toronto, Canada

⁹Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada

Corresponding author:

Babak Taati, Toronto Rehabilitation Institute, University Health Network, 550 University Avenue, Toronto, ON M5G 2A2, Canada.
Email: babak.taati@uhn.ca



impairments that interfere with daily, vocational, social, and leisure activities.^{1,2} In particular, 34–38% of stroke population with upper limb impairment achieved some dexterity after six months and only 12% achieved full functional recovery.^{3,4} Various motor interventions have been developed to support recovery and facilitate use of the upper limb post-stroke.^{5,6} The Graded Repetitive Arm Supplementary Program is a set of upper limb exercises intended for people with a stroke to perform independently.⁷ These exercises improve upper limb function, as measured by the Chedoke Arm and Hand Activity Inventory and grip strength among sub-acute patients.⁸ In addition, the Graded Repetitive Arm Supplementary Program is included in the Canadian Stroke Best Practice Recommendations,^{9,10} and is frequently used in clinical practice.⁸ In studies examining the Graded Repetitive Arm Supplementary Program, participants have been supervised by therapists while performing exercises. Ultimately, it will be beneficial if users of the Graded Repetitive Arm Supplementary Program can achieve similar levels of functional recovery without supervision, reducing reliance on therapists, and thereby supporting self-management. However, one potential concern with this program and other similar therapies is that instructions alone cannot prevent compensatory motions (e.g. trunk rotation or flexion), which are often associated with poor functional recovery.^{11,12}

To address this issue, biofeedback is sometimes recommended in stroke therapies and has been shown to reduce compensatory motions.^{13–20} Biofeedbacks can be provided in the form of visual (virtual reality, looking at a mirror or screen), auditory (speech or sounds delivered by a therapist or device), or haptic feedbacks (force or tactile information). Visual feedback is highly customizable²¹ and can be conveniently setup on a TV or computer screen. However, visual feedback may distract users from performing exercises that required high level of attention. Auditory feedback allows parallel processing of information along with visual signal and has been shown to increase attention.^{22,23} Finally, haptic feedback can provide direct intervention of physical movements, but usually require expensive and sophisticated setup (e.g. robot-assisted devices).²⁴ In general, biofeedback can be in the form of concurrent or terminal feedback and there is no consensus on the optimal timing of feedback in the context of stroke rehabilitation.²⁵ While some studies suggest that very frequent concurrent feedback may lead to dependence on feedback, the exact relationship between feedback type, feedback frequency, and task complexity remains undetermined.^{13–17,25,26} In addition, the type of information displayed could also affect how users interpret the feedback and whether they correct their movements accordingly. Therefore, a new feedback method that

can provide either concurrent or terminal feedback and allow customization in its content and frequency is desirable.

In this pilot study, we explored the potential to provide visual feedback using a markerless motion tracking system when users perform exercises from the Graded Repetitive Arm Supplementary Program. Visual feedback is chosen here as some of these exercises are quite complex and visual display allows users to be aware their postures; by observing their postures, users can maintain or correct their movements and achieve self-learning.^{25,27} In this study, a Microsoft Kinect v2 sensor was used for this purpose, but video-based tracking²⁸ is also possible. The Kinect sensor tracks participants' motions by estimating their joint positions in real time. Data can be displayed in real time or saved and replayed after the exercises, thereby allowing either concurrent or terminal feedback with customized frequency, timing, and displayed content. Two types of feedback content were included in the study: video display, which was analogous to a mirror, and avatar display, which was an animated figure mimicking a user's body movements. The rationale for using an avatar display was that it removed irrelevant background information so that user attention could be directed towards body movements. The avatar display was created using 3D graphics to make out-of-plane rotations more noticeable.

Our primary objectives were to examine: (1) avatar validity; and (2) acceptability of the visual feedback system. Validity of the avatar display indicated how well the avatar depicts and approximates real body movements, and was determined by whether trained health science researchers could identify the same compensations from body movements shown with the avatar and video images. Validity of avatar display has not been investigated extensively in the context of compensatory motions. However, one study showed that the majority of joint position offsets for the Kinect v2 sensor were between 50 and 100 mm for upper limb and body joints, which was relatively small compared to the range of motion in the exercises.²⁹ Therefore, we expected a substantial agreement between video and avatar poses (Cohen's $\kappa > 0.6$).

Acceptability measures users' opinion on enjoyment, satisfaction, motivation, and level of effort after interacting with a rehabilitative intervention.^{30,31} Evaluating acceptability is crucial as it allows researchers to design and modify the intervention to meet users' needs and increase users' desire to adopt this intervention as part of their stroke rehabilitation. Acceptability is usually conducted through an interview with the participants and the aforementioned measures can be evaluated quantitatively using a Likert scale question qualitatively through open-ended questions.^{30,31} In this

study, acceptability of the visual feedback system was determined from a usability survey administered to participants immediately after the study. We hypothesized that the participants would prefer the avatar display over video display due to background removal and 3D representation of the avatar figure. The two secondary objectives were to determine: (1) participants' attention towards visual feedback; and (2) the effect of visual feedback on compensatory behavior. Attention towards visual feedback, inferred from observations of participants directing their faces toward the screen, was analyzed through annotating video footage of participants doing exercises. As mentioned earlier, visual feedback may distract users from performing exercises that required high level of attention. Therefore, analyzing user's attention can provide insight on the type of exercises that may distract user from viewing visual feedback and determine whether concurrent visual feedback is a feasible method for all exercises performed in this study. Our hypothesis was that participants would look at the visual feedback more on exercises that require less hand-eye coordination. The effect of visual feedback was evaluated by comparing the rate of compensations between feedback and no-feedback conditions. We hypothesized that the rate of compensation will be lower, though not substantially, in the presence of visual feedback due to raised self-awareness.

Methods

Participants

Participants were recruited from a research volunteer database maintained by Toronto Rehabilitation Institute. Inclusion criteria were: age >18 years, confirmed diagnosis of stroke at least six months^{32,33} prior to the study to ensure chronicity, and a Fugl-Meyer Assessment Upper Extremity score between 10 and 57 (out of 66).³⁴ Participants were excluded if they could not understand verbal and written instructions in English; had severe visuo-spatial neglect (i.e. score <44 on the Star Cancellation Test);³⁵ and/or had any neurological condition other than stroke, or an orthopedic condition that would affect their ability to perform exercises from the Graded Repetitive Arm Supplementary Program. To confirm eligibility, an initial telephone screening was conducted, followed by in-person screening where an occupational therapist administered the Fugl-Meyer Assessment – Upper Extremity and Star Cancellation Test. Data were collected on 10 eligible volunteers with at least two participants assigned to each level of the Graded Repetitive Arm Supplementary Program. The study protocol was approved by the University Health Network (UHN)

Research Ethics Board (Approval # 15-9704). All participants provided written informed consent.

Experimental protocol

Experimental setup. Participants sat on a chair with the Graded Repetitive Arm Supplementary Program manual placed on a table in front of or beside them. All exercises were performed while sitting on a chair with arm rests. A custom software application was used to display visual feedback on a television monitor 2 m in front of the participant. The Kinect sensor was mounted on a tripod to the left side of the monitor (Figure 1). For exercises that required a table, the table was placed in front of the participants. Prior to the experiment, a researcher demonstrated the protocol for the Graded Repetitive Arm Supplementary Program to each participant and confirmed they were able to perform the assigned exercises independently. Videos and joint positions of participants were tracked and recorded at 30 frames per second throughout all exercises.

Cross-over study design. Each participant was asked to perform the exercises in front of a computer monitor under three different conditions in three separate visits approximately one week apart. The three conditions were:

- Phase A (no feedback): The monitor was turned off and participants were asked to only focus on the exercises. The Kinect sensor was used for recording purposes only (Figure 1(a)).
- Phase B (video feedback): The live video of the participants doing exercises recorded from the Kinect camera was displayed on the monitor in real time. Participants were asked to pay attention to the screen during each exercise whenever possible (Figure 1(b)).
- Phase C (avatar feedback): C was similar to B, except that participants' movements were displayed as an avatar instead of a video (Figure 1(c)).

The order of the phases varied among the participants to include all permutations; e.g. ABC, ABC, BAC, etc.

Graded Repetitive Arm Supplementary Program

The program consists of three levels, each developed to reflect the severity of upper limb motor impairment as defined by the Fugl-Meyer Assessment – Upper Extremity scores (Level 1 = 10–25, severe; Level 2 = 26–45, moderate, Level 3 = 46–58, mild). Participants were assigned exercises from the level that corresponded

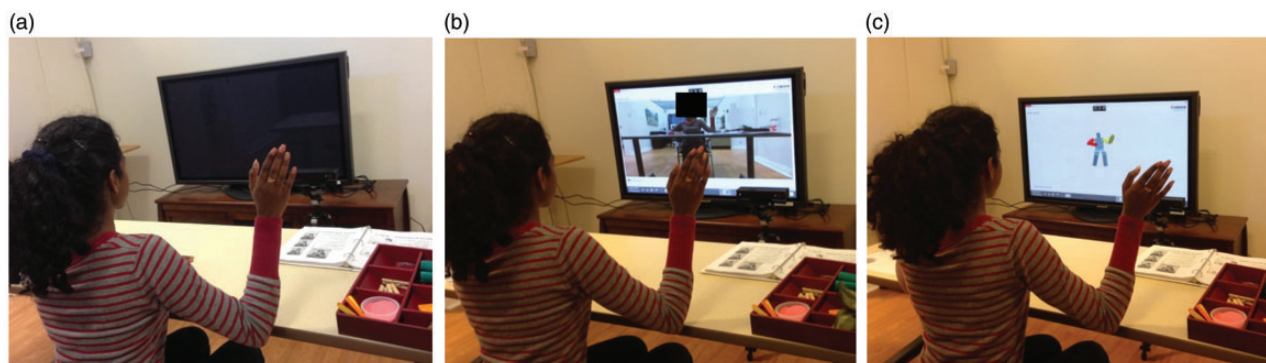


Figure 1. Experimental setup of performing exercises in (a) no feedback, (b) video feedback, and (c) avatar feedback conditions. For anonymity, the face of the volunteer (not a participant) illustrating the setups is blocked off in (b).

to their Fugl-Meyer scores. The manual³⁶ specifies certain exercises to be performed in 2–3 sets of 3–5 repetitions. However, since the focus of the study is on the feasibility of a visual feedback system rather than the therapeutic effect of these exercises, participants were asked to perform each exercise in one set of three repetitions. The manual consists of 17, 31, and 31 exercises for level 1, 2, and 3, respectively, so participants were asked to perform 51, 93, and 93 repetitions for each visit. Participants were also informed to stop doing an exercise if they felt any pain.

Motion tracking

Kinect v2 is a marker-less motion tracking device that uses a video camera and a depth sensor to detect and track people's major body joints ("skeleton") within 0.5–4.5 m from the sensor at 30 frames per second.³⁷

Video and avatar displays

Two forms of visual display were presented to the participants in this experiment: video and avatar. Video display showed the images captured by the video camera, which was used to mimic the effect of looking at a mirror (i.e. the limb on the left side of the image reflects the movement of participant's left limb).

The avatar was constructed by representing the body as a collection of rectangular blocks. Rectangular blocks were used instead of cylinders so that rotations around the axis of the limb or trunk were more noticeable. In addition, joints at the extremities (i.e. hands and feet) were excluded since they produce excess noise and have minimal impact on representing trunk and shoulder compensatory movements. Specifically, only 12 of the 25 joints were used to construct the head, the trunk, and four limbs (Figure 2). Hands were constructed such that they were always aligned in the same orientation as their respective upper limbs. Different colors were used in

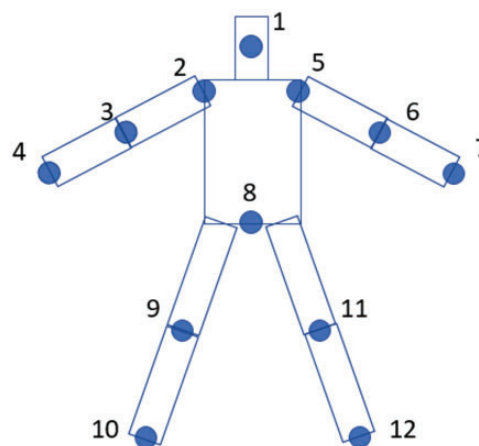


Figure 2. Joints tracked by Kinect sensor and used to construct the avatar. 1: Head, 2: Shoulder-left, 3: Elbow-left, 4: Wrist-left, 5: Shoulder-right, 6: Elbow-right, 7: Wrist-right, 8: Spine-base, 9: Knee-left, 10: Ankle-left, 11: Knee-right, 12: ankle-right.

both upper limbs and the trunk to provide contrast while performing exercises from the Graded Repetitive Arm Supplementary Program. Sample avatar images for different types of compensations are shown in Figure 3.

Assessment of acceptability

At the end of Phases B and C, participants completed a usability survey (see Appendix 1) to assess their perceptions of the visual feedback system. The survey consisted of 10 Likert items followed by seven open-ended questions. The 10 Likert items were based on the System Usability Scale, a scale commonly used in the industry for evaluating usability, which reflects the comfort level, amount of training required, and perceived complexity of a product.^{38,39} The Likert items used a 5-point scale from strongly disagree to strongly agree with each statement. The statements were designed such that odd items were positive comments

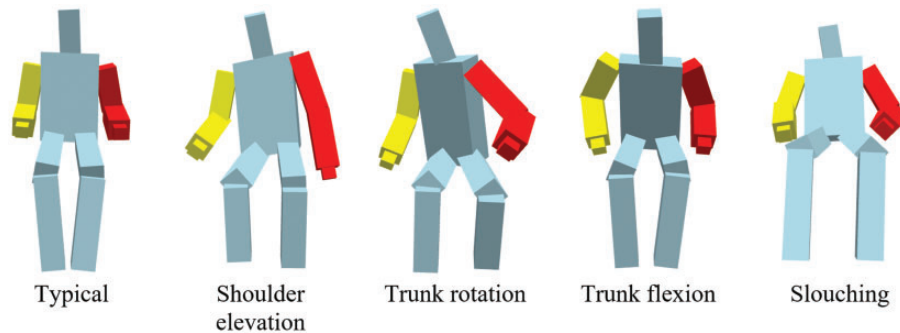


Figure 3. Avatar images of bodies displaying different types of compensations. Red arm indicates affected side.

about the usability of the product and even questions were negative comments. Two System Usability Scale items that were not relevant to the visual feedback system were replaced by two context-specific items that were more suitable for this study. The seven open-ended questions were designed to gain insight on the following information:

- Perceived accuracy of the system in terms of tracking participants' movements;
- Perceived usefulness of this feedback system during the exercises;
- Preference between video and avatar display on participants' ability to concentrate on their movements; and
- The user-friendliness of the interface.

Data analysis

Data annotation

Two members of the research team were trained by an occupational therapist to identify compensations in recorded video images of participants performing exercises. For each participant, the annotation of recorded videos was viewed frame by frame, and the annotators individually noted if any compensatory motions were observed. Three common types of compensation were identified: shoulder elevation, trunk rotation, and trunk flexion. In addition, slouching was also considered as a form of poor posture. Trunk flexion refers to leaning the upper body forward while slouching refers to a posture with hip extension (Figure 3). Up to two types of compensations were annotated for each frame. Overall, approximately 700 exercises and 13.4h of data were collected, as performed by the 10 participants across all three phases. The primary annotator annotated all exercises from all participants in all three phases using video images. To evaluate the validity of the avatar display, Phase C

exercises were annotated using the exact same avatar images that were displayed during the experiment. These avatar images were annotated separately from the video images so the annotators could not infer any compensatory movements from the corresponding video images. To calculate inter-rater agreement, the secondary annotator annotated a randomly selected subset of exercises in all three phases using video images as well as a subset of Phase C exercises using avatar images. The subset of exercises accounted for approximately 20% of the entire data set and was chosen to include different exercises from all participants and all three phases. To estimate attention toward visual feedback, the primary annotator also looked at all videos frame by frame to determine when the participant looked at the screen. Frames in which participants were not performing exercises (e.g. resting) were excluded from the analyses.

Primary outcome measures

Validity of avatar feedback. For each type of compensation, Cohen's kappa was used as an agreement measure to quantify how well the annotated avatar images characterized different types of compensation compared to the corresponding video images.^{40,41} To calculate the agreement, an exercise was labeled with a type of compensation if that particular compensation was identified in at least 5% of frames of the exercise. Each exercise usually lasted between 30s to 2min, so 5% of frames corresponded to approximately 2 to 6s. Kappa was then calculated based on how each exercise was labeled using video and avatar images. In this study, the interpretation of kappa, as defined by Landis and Koch, was as follows: 0–0.20 (slight agreement), 0.21–0.40 (fair agreement), 0.41–0.60 (moderate agreement), 0.61–0.80 (substantial agreement), and 0.81–1 (almost perfect agreement).⁴² In addition, validity of avatar feedback was verified by evaluating inter-rater agreement between the two annotators,

which was also quantified using Cohen's kappa for each type of compensation. The inter-rater agreement was evaluated for both video and avatar images. The former provided insight on annotators' interpretation of compensations and the latter showed how well the avatar resembles participants' movements during exercises.

Acceptability of visual feedback system. The score from the System Usability Scale was calculated as follows:³⁸

- (1) For odd items, a score of 0 to 4 was assigned from strongly disagree to strongly agree;
- (2) For even items, score of 4 to 0 was assigned from strongly disagree to strongly agree.
- (3) Scores for all 10 items were summed and multiplied by 2.5 to normalize the scale from 0 to 40 to 0 to 100.

Two System Usability Scale scores were calculated for each participant to reflect their perception of usability of both video and avatar displays. The responses from the seven open-ended questions of the survey were categorized and summarized to reflect what participants liked or disliked about the system and how the system could be improved.

Secondary outcome measures

Attention to visual feedback. Attention was estimated as the percentage of time a participant looked at the screen during an exercise. For each participant, attention was calculated separately for exercises performed in video and avatar feedback phases.

Since certain exercises required more hand-eye coordination than others, the relationship between attention and the nature of the exercises was also examined. Specifically, exercises from all three levels of the Graded Repetitive Arm Supplementary Program manual were classified into three categories (see Appendix 2) based on the extent in which the nature of the exercise would prevent participants from looking at the screen:

- Category 1: Minimal hand-eye coordination required and participant can face the screen throughout the exercise (e.g. shoulder shrug).
- Category 2: Exercise requires body movements that would impede participants' attempt to look at the screen or part of the exercise requires visually focusing on the task (e.g. bending forward).
- Category 3: Exercise requires constant visual focus and would be difficult to execute otherwise even for individuals without an upper limb motor impairment (e.g. pouring water).

Differences between video and avatar feedback were also compared to determine whether attention was dependent on the type of visual feedback.

Effect of visual feedback on compensation. The effect of visual feedback on mitigation of compensatory motions was evaluated by comparing the percentage of frames with compensation in each phase. Each type of compensation was quantified as the percentage of frames in which that type of compensation was observed. The overall compensation rate was quantified as the percentage of frames in which at least one type of compensation was observed. Compensation rates were calculated in phase A, B, and C and for each participant.

Statistical analysis

Shapiro-Wilk test was applied on all data to determine whether each sample was normally distributed. Non-parametric statistical tests were performed for all non-normal samples. A paired-sample *t*-test was used to determine if there was a significant difference in (1) acceptability and (2) overall attention between video and avatar feedback. Kruskal-Wallis one-way analysis of variance (ANOVA) was used to determine if attention from the three categories of exercises were significantly different. Post hoc analysis was performed using Bonferroni correction to determine if differences between each two categories were significant. Repeated-measures one-way ANOVA was used to determine whether compensation rate for each type of compensation (except slouching) and the overall rate were significantly different in phase A, B, and C. Mauchly's test for sphericity was also calculated to ensure that assumptions of repeated-measures one-way ANOVA were not violated. Friedman's test was used to determine whether slouching rates were significantly different in phase A, B, and C.

Results

Three participants were classified as level 1, four as level 2, and three as level 3 of the Graded Repetitive Arm Supplementary Program. The participants' characteristics are summarized in Table 1. All participants performed the three phases of the experiment at least one week apart except for participant 8 (7 days and 4 days apart) and participant 10 (6 days and 1 day apart). Most participants took short frequent breaks throughout the exercises, but fatigue and pain were not reported. In all three phases combined, all three level 1 participants completed all 51 repetitions; five participants did not complete between 1 and 5 out of 93 (level 2) or 84 (level 3) repetitions; one participant did not

Table 1. Participant characteristics. Range of Fugl-Meyer Assessment – Upper Extremity (FMA-UE) score for each level of the Graded Repetitive Arm Supplementary Program: Level 1 = 10–25, severe; Level 2 = 26–45, moderate, Level 3 = 46–58, mild.

Participant	Age (years)	Sex	Time post-stroke (years)	FMA-UE score	Level	Star-cancellation score	More affected side
1	43	F	7	46	3	55	Right
2	55	M	5	32	2	56	Right
3	69	F	2	50	3	56	Right
4	69	M	2	42	2	56	Right
5	79	M	6	43	2	43 ^a	Right
6	67	M	9 months	14	1	56	Left
7	59	F	2	26	1 ^b	56	Right
8	65	M	4	18	1	56	Left
9	59	M	3	35	2	56	Left
10	60	M	13	48	3	56	Right

F: female; M: male.

^aParticipant included since the star-cancellation score was one fewer than the cut-off score and the stars that were marked were evenly distributed across the test sheet, thus showing little sign of unilateral spatial neglect.

^bParticipant had a Fugl-Meyer Assessment – Upper Extremity score of 26 but elected to perform level 1 exercises based on her physical condition.

complete 6 out of 84 repetitions, and one participant did not complete 11 out of 93 repetitions. The reasons for not completing all repetitions include: (1) participants thought they had performed three repetitions for an exercise when only 1 or 2 were performed, and (2) participants found certain exercises too difficult and decided to move on without completing three repetitions.

Validity of avatar feedback

Cohen's kappa between annotation using video and avatar images during the avatar-feedback session were 0.61 (0.51–0.71), 0.84 (0.76–0.91), 0.80 (0.72–0.89), 0.73 (0.53–0.94) for shoulder elevation, trunk rotation, trunk flexion, and slouching, respectively (value in parentheses shows 95% confidence interval⁴³). For inter-rater agreement, Cohen's kappa for shoulder elevation, trunk rotation, trunk flexion, and slouching were respectively 0.41 (0.21–0.62), 0.59 (0.38–0.81), 0.31 (0.0–0.71), 0.55 (0.06–1.0) for annotation using video images. The corresponding values for annotation using avatar images were 0.41 (0.15–0.67), 0.74 (0.55–0.94), 0.50 (0.15–0.84), 0.0 (0.0–0.94) (Figure 4). In the subset of data chosen for avatar annotation, one of the annotators only found four instances of slouching while the other did not identify this movement in any of the exercises, resulting in zero agreement.

Acceptability of visual feedback system

System Usability Scale scores were calculated for both video feedback and avatar feedback systems.

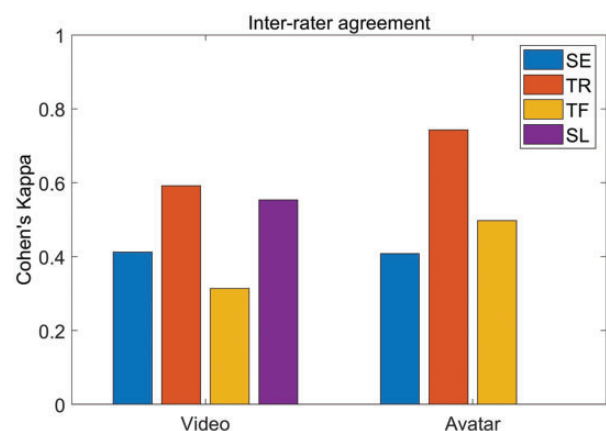


Figure 4. Inter-rater agreement for annotating compensation using video and avatar images. Agreement values for video feedback are 0.41 (0.21–0.62), 0.59 (0.38–0.81), 0.31 (0.0–0.71), 0.55 (0.06–1.0) for shoulder elevation, trunk rotation, trunk flexion, and slouching, respectively. Agreement values for avatar feedback are 0.41 (0.15–0.67), 0.74 (0.55–0.94), 0.50 (0.15–0.84), 0.0 (0.0–0.94) for shoulder elevation, trunk rotation, trunk flexion, and slouching, respectively. Values in parentheses represent 95% CI.

SE: shoulder elevation; TR: trunk rotation; TF: trunk flexion; SL: slouching.

The average System Usability Scale scores for video and avatar feedback were 39.3 ± 29.5 and 28 ± 18.8 , respectively out of a total of 100 points. There was no significant difference for System Usability Scale scores between video and avatar feedback ($t_9 = 1.02$, $p = 0.33$).

Table 2 highlights the important findings from the open-ended questions of the usability survey.

Table 2. Participant feedback on the visual feedback system.

	Video feedback	Avatar feedback
Positive	<ul style="list-style-type: none"> Helped see things of which they were not aware Liked the idea of doing stroke rehab exercises at home Would encourage them to do exercises 	<ul style="list-style-type: none"> Provided information on posture during exercise Helped see things of which they were not aware More comfortable than looking at oneself
Negative	<ul style="list-style-type: none"> Did not warn users about compensation and provide corrective motion Participants were required to actively look at the screen 	<ul style="list-style-type: none"> Did not warn users about compensation and provide corrective motion Participants were required to actively look at the screen Avatar was not accurately tracked occasionally

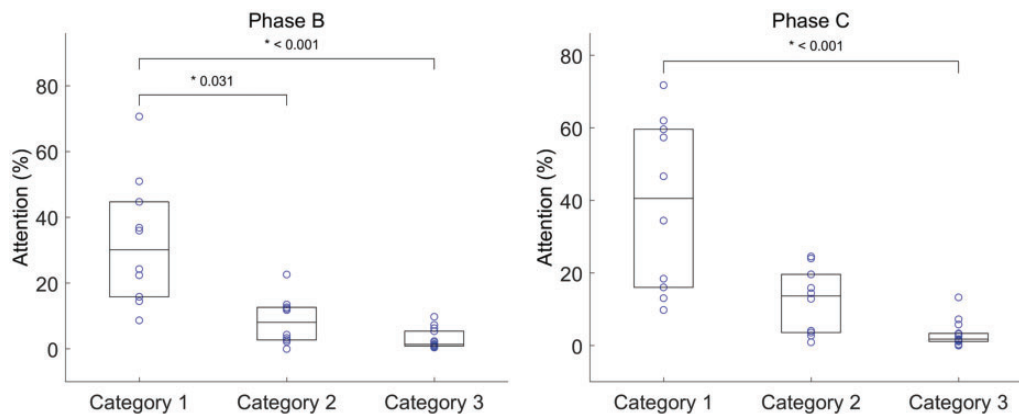


Figure 5. Participants' attention towards visual feedback during phase B (left) and phase C (right) exercises. Average attention in phase B for categories 1–3 are 28.6%, 17.2%, and 9.8%, respectively. Average attention in phase C for categories 1–3 are 27.6%, 18.9%, and 9.3%, respectively. Each point represents the mean attention of all participants who performed that exercises in either phase B or C. Boxes represent mean and upper and lower quantiles. Category 1: minimum hand–eye coordination required; Category 2: hand–eye coordination required during initial setup before each repetition of the exercise. Category 3: hand–eye coordination was crucial throughout the execution of the tasks specified in the exercises.

Of the ten participants, three preferred video display, three preferred avatar, and four were either neutral or had no preference. One participant stated that video display was more realistic than avatar. Others stated that the avatar was easier to comprehend and was less distracting than seeing oneself in the feedback. For both video and avatar displays, six out of ten participants would use the system if it were available to them and only two participants found the system setup technically challenging. When asked about implementing features to automatically detect compensations and alert the person, seven participants supported the idea of having both visual and auditory cues while two suggested that auditory cues would be sufficient. In addition, six participants preferred displaying instructions on the same screen as the visual feedback to avoid constantly alternating between looking at the exercise manual and paying attention to the visual feedback.

Attention to visual feedback

Participants appeared to be looking at the video and avatar feedback in $10.9 \pm 8.6\%$ and $12.4 \pm 6.9\%$ (mean \pm standard deviation) of frames, respectively. There was no significant difference in attention between the two feedback conditions ($t_9 = 0.66$, $p = 0.53$). In both video and avatar feedback sessions, there were significant differences between category 1, 2, and 3 exercises ($\chi^2(2) = 20.84$, $p < 0.001$ for video feedback and $\chi^2(2) = 20.01$, $p < 0.001$ for avatar feedback). Specifically, in video feedback sessions, attention during category 1 exercises was significantly higher than that of categories 2 ($p = 0.03$) and 3 ($p < 0.001$). In avatar feedback sessions, category 1 was significantly higher than category 3 ($p < 0.001$) but not significantly higher than category 2 ($p = 0.06$) (Figure 5). Participants looked at the screen more in category 2 exercises than category 3 exercises; however, the

difference was not statistically significant ($p=0.22$ and $p=0.15$ for video and avatar feedback, respectively).

Effect of visual feedback on compensation

Sphericity assumption was not violated for shoulder elevation ($\chi^2(2)=0.807$, $p=0.668$), trunk rotation ($\chi^2(2)=0.235$, $p=0.889$), trunk flexion ($\chi^2(2)=5.112$, $p=0.078$), and the overall compensation rate ($\chi^2(2)=0.521$, $p=0.771$). The compensation rate for each type of compensation, as well as the overall compensation rate within in each phase are shown in Figure 6. There were no statistically significant differences between each type of compensation, as well as overall compensation rate across different conditions.

Discussion

Validity of avatar feedback

The purpose of this study was to characterize avatar display and acceptability of a home-based visual feedback system for preventing compensations during performance of self-directed upper extremity exercises among individuals with chronic stroke. Based on the result, shoulder elevation and slouching showed substantial agreement while trunk rotation and trunk flexion showed almost perfect agreement. Michaelsen et al. reported that trunk rotation among 28 participants

with stroke was $12^\circ \pm 3^\circ$ while other had shown that trunk flexion ranged between 67 and 178 mm⁴⁴⁻⁴⁶ for functional reaching and grasping tasks. The magnitudes of these trunk movements exceeded the joint position and angle error of the Kinect sensor,²⁹ which justified that trunk compensations were detectable in most exercises from the Graded Repetitive Arm Supplementary Program. On the other hand, shoulder elevation had the highest discrepancy between avatar and video images since the movements were more subtle compared to other types of compensation. Therefore, the movement was sometimes undetected in the avatar images. The result agreed with Brokaw et al.,⁴⁷ who reported that Kinect was not sensitive enough to detect shoulder elevation. Chang et al. and Weiss et al. also reported that Microsoft Kinect was not very sensitive to shoulder joint movement in the vertical direction.^{48,49} In the future, it may be beneficial to make the avatar more sensitive to shoulder elevation. One possible way to improve the sensitivity is to place light reflective markers or LED markers on the shoulders and use custom software or image processing algorithms to detect shoulder elevation.^{50,51}

Inter-rater agreement was analyzed to verify compensation identification from the primary annotator. For annotation using video images, shoulder elevation, trunk rotation, and slouching showed moderate agreement while trunk flexion showed fair agreement. For avatar images, only trunk rotation showed substantial agreement. The zero agreement for slouching was partially due to the rare occurrence of slouching (1.5% of frames from the overall experiment), which was only labeled four times by both annotators combined in the subset of data that was used for inter-rater agreement. In both avatar and video annotations, trunk rotation had the highest agreement. One possible explanation was that trunk rotation involved out-of-plane rotation with respect to the image plane, which was more noticeable compared to other types of compensations. Due to the large discrepancy between the two annotators, it was not possible to conclude whether low inter-rater agreement for avatar images was due to interpretation of compensations or inaccurate representation of compensatory movements. Nevertheless, the result supported our hypothesis that avatar images were reasonably representative of video images in the context of characterizing compensatory movements.

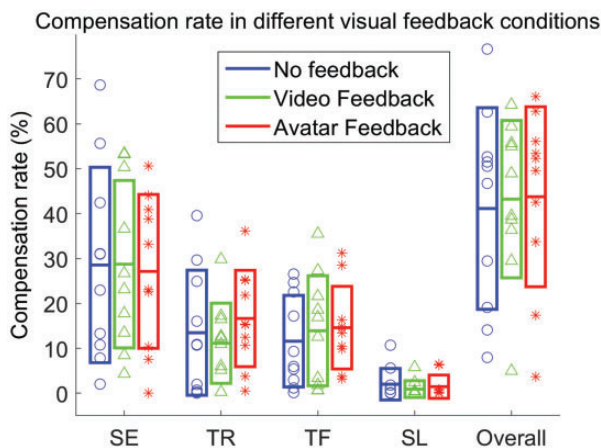


Figure 6. Compensation rates for shoulder elevation ($p=0.88$, $F_{2,7}=0.12$), trunk rotation ($p=0.29$, $F_{2,7}=1.32$), trunk flexion ($p=0.75$, $F_{2,7}=0.29$), slouching ($p=0.15$, $\chi^2(2)=3.8$) and the overall compensation rate ($p=0.75$, $F_{2,7}=0.29$) in different feedback conditions. Boxes represent mean and standard deviation of the sample, respectively. p -Values and F -statistics are for the within-subject effect across no-feedback, video-feedback, and avatar feedback conditions.

SE: shoulder elevation; TR: trunk rotation; TF: trunk flexion; SL: slouching.

Acceptability of visual feedback system

Overall, there was a large discrepancy between the Likert questions and the open-ended questions. The System Usability Scale score was low due to the passive nature of this visual feedback system and

occasional tracking errors that affected the avatar display. Most participants stated the system did not achieve more than a mirror since it could not detect compensations. Therefore, the system appeared to be unnecessarily complex and cumbersome to use compared to a mirror. However, the comments from the open-ended questions was generally positive and justified that the visual feedback system would be helpful if automatic compensation detection is implemented through audio and/or visual cues.

There was no clear preference between video and avatar display, which did not support our hypothesis that 3D representation of the arm and trunk movements were more interpretable compared to video display. To our knowledge, no previous studies have compared the effect of video and avatar feedback for upper limb stroke exercises. However, Petzold et al.⁵² conducted a study on a simple clockwork assembly task and showed that participants had faster completion time and exerted less force to complete the tasks when viewing a virtual 3D model of the apparatus compared to video stream. The study suggested that avatar display could direct participants' attention to relevant information and result in faster reaction time and less effort to perform accurate movements. Therefore, more participants need to be recruited to determine whether video and avatar feedback is more helpful for arm and hand stroke rehabilitation exercises.

Attention to visual feedback

The analysis on attention towards visual feedback showed a high variability among category 1 exercises while attention towards category 2 and 3 exercises was consistently low. The result supported our hypothesis that participants would look at the visual feedback more frequently in exercises that required less hand-eye coordination. For more complicated tasks, participants seemed more occupied with executing the tasks rather than looking at the visual feedback likely to see if they were compensating. It is worth noting that most category 3 exercises required precise arm movement and dexterity. If participants had inadequate motor control in arms and hands, these complex exercises may lead to muscle fatigue and increase the likelihood of compensatory behavior.⁵³ Therefore, the result suggested that participants directed less focus to visual feedback in exercises that were more prone to compensatory behaviors. This finding further motivates the need for an automated method to detect compensation and only draw user's attention when compensation is detected. Overall, there was no significant difference between amount of time looking at the video and avatar feedback among all participants, which supported our speculation that how much the participants

look at the screen was mostly dependent on the nature of the tasks they were performing.

Comparison with home-based stroke therapies

Currently, research on home-based, upper-limb stroke therapies with external feedback is mostly focused on tele-rehabilitation, exergaming (e.g. Kinect and Wii games), and home-based stroke robots.^{54–56} However, tele-rehabilitation requires supervision from a therapist while home-based stroke robots are limited to simple functionalities for affordability.^{54,57} Therefore, exergaming has emerged as a popular home-based therapy that uses low-cost sensors to provide visual, audio, and haptic feedback on accuracy and reachability of arm and hand movements.^{58,59} Some groups have also developed novel therapies that are feasible in a home setting. For example, Zhang et al.⁶⁰ created a virtual piano apparatus with visual and audio feedback as a home-based therapy for people with upper limb stroke, though it has not been tested on people with stroke.

While several exergaming systems have shown promising results, there is no consensus on which types of in-game arm and hand movements are best for stroke recovery and whether playing games with commercial game consoles are more effective than conventional stroke therapies.⁶¹ In addition, many studies that incorporated commercially available game consoles used existing games that were not designed for rehabilitation and there was little evidence on whether achieving a high score in these games translates to improvement in daily tasks.^{58,61}

In some respects, our system is similar to existing exergaming systems, although there are no gaming elements in the Graded Repetitive Arm Supplementary Program. Our study addressed some of the limitations of existing exergaming systems by using a clinically proven, self-administered protocol, where visual feedback was used to significantly reduce therapist supervision while achieving the same level of effectiveness. In addition, the Graded Repetitive Arm Supplementary Program focuses on a variety of functional tasks, which are directly translatable to daily activities.⁷ Therefore, this visual feedback system can increase the awareness of users' arm movements without compromising the therapeutic effect of an existing recommended therapy. Furthermore, the timing and frequency of visual feedback can be highly customizable for our system whereas most exergaming therapies require users to watch a display while performing an exercise. As a result, these tools can only be delivered through concurrent feedback. One drawback of the visual feedback system, however, is that this therapy might not appear as intriguing as game-based therapies that participants might be more motivated to complete.

Effect of visual feedback on compensatory movements

The compensation rate for each type of compensation as well as the overall compensation rate was very similar across all three phases. Contrary to our hypothesis, visual feedback had minimal impact on mitigation of compensation. This appeared to be because participants tended to focus more on the exercises than on the visual feedback if they had more difficulty with the tasks. Other possible reasons include: (1) participants were not aware of what constitutes a compensation; (2) avatar was not accurate enough to depict a compensatory motion; and/or (3) participants were aware of compensations, but they did not know how to or were not physically able to correct their movements.

Limitations

Skeletal tracking by the Kinect sensor could potentially be affected by two aspects of our experimental setup: (1) the table was in front of the participants in certain exercises, which could lead to joint occlusion and (2) the Kinect sensor was placed at the distance that optimally tracks the upper body of the participants, so their lower legs could sometimes be outside the field of view of the Kinect sensor and displayed incorrectly. To ensure that skeletal tracking was working properly, our setup was designed such that the participants were always facing the Kinect sensor throughout the exercises, as shown in Figure 1. In addition, the Kinect sensor was placed at the same height as the table top to minimize joint occlusion. Therefore, skeletal tracking was smooth when testing the setup and filtering was deemed unnecessary. To address the second concern for future studies, the lower body could simply be fixed or removed since it does not provide additional information on compensatory motions in the upper body.

Another limitation regarding the experimental setup was that the television monitor was placed in front of participants for all three phases. Therefore, in the baseline condition where the television is off, participants could have potentially looked at the reflection from the monitor during the experiment.

In addition, one participant had kyphosis and the Kinect sensor could not track the joint positions for up to 36% of the frames during a few exercises. Fortunately, the average tracking rate was $95.1 \pm 4.8\%$ among the other nine participants, so the missing frames should not have a significant impact on the validity and inter-rater agreement analysis.

With regards to data analysis, slouching was only labeled in 1.5% of all the frames in the entire dataset. Even though Cohen's kappa was used to account for chance agreement in the validity analysis, a bigger sample is required to further support the claims

regarding slouching. In addition, the cut-off score for substantial agreement was not universally accepted, so the result might differ if a different benchmark scale was used. In feasibility analysis, two nonrelevant questions from the System Usability Scale survey were replaced by context-specific questions. Therefore, the System Usability Scale score could not be used to directly compare our system with other products that used the same scale. Instead, these questions were used to determine participants' preference between different types of feedback and reflect their opinions on the usability of the visual feedback system. Regarding the open-ended questions in the usability survey, content analysis was not used since there were only 10 participants and we were able to incorporate all their comments in the discussion section. Furthermore, attention was estimated by looking at the video of the participants performing exercises; even if the participants were looking at the screen, we could not measure what information was gathered.

Future work

In the future, the first step will be to improve the functionality of the visual feedback system. This involves implementing a method for automatic compensation detection using the joints tracked by the Kinect sensor. If shoulder movements are too subtle to be reflected by the joint data, extra markers will be used instead for detecting shoulder elevation. Since lower legs are sometimes outside the field of view of the Kinect sensor, we would also consider fixing the position of the lower legs when displaying avatar images to avoid excess noise. To better understand how much participants observe the feedback during the exercises, gaze tracking could be used to more accurately measure visual attention.

Next, additional features can be explored to optimize the therapeutic effect and acceptability to the user. For example, visual and auditory cues can be implemented to determine whether different types of feedback or combination of feedbacks can affect the therapeutic outcome. More participants can be recruited to determine if there is a preference between video or avatar display. In addition, the screen display can also be modified to show the exercise instructions beside the video or avatar images. Upon validation, the visual feedback system can be customized further to provide concurrent or terminal feedback and display quantifiable measures to reflect the extent of compensatory movements. Ultimately, we aim to develop an effective and affordable visual feedback system for chronic stroke survivors with upper limb impairment to perform stroke recovery exercises at home with minimal supervision.

Conclusions

This pilot study aimed to examine the feasibility and acceptability of a visual feedback system for home-based stroke rehabilitation exercises. There was substantial agreement between video and avatar display for shoulder elevation and slouching, and almost perfect agreement for trunk rotation and flexion. The usability survey and attention analysis highlighted the need to implement automatic compensation detection as well as alerting features (e.g. visual or auditory cues) when compensation movements were detected. Most participants agreed that the visual feedback system can be built upon its existing functionalities and turn into a valuable tool for stroke patients performing rehabilitation exercises without assistance from a therapist.

Consent for use of personal materials in paper

The individual in Figure 1 (not a participant) has given written consent to the inclusion of material pertaining to her and acknowledges that she cannot be identified via the paper. The face in the figure has been blocked off for anonymity.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was partially funded by a research grant from the Promobilia Foundation. Equipment and space were funded with grants from the Canada Foundation for Innovation, Ontario Innovation Trust, and the Ministry of Research and Innovation. AM holds a New Investigator Award from the Canadian Institutes of Health Research (MSH-141983). The funding sources did not have any role in the experimental process or in the preparation of the manuscript, and the views expressed do not necessarily reflect those of the funders.

Guarantor

BT

Contributorship

SL implemented the visual feedback software, performed data analysis and drafted the manuscript. JM performed data collection and data processing. AM and BT helped with data interpretation. BT supervised the study progress. AM, RW, BT and JH were involved in the study design. All authors helped revise the manuscript and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge Cynthia Ho, who conducted Fugl-Meyer assessments on all participants, Laura Wu and Brandon Malamis, who assisted with recruiting participants, collecting, and processing data, and Marge

Coahran, who provided technical support on data processing and video annotation.

ORCID iD

Shayne Lin  <http://orcid.org/0000-0003-3716-5063>

References

- Langhorne P, Coupar F and Pollock A. Motor recovery after stroke: a systematic review. *Lancet Neurol* 2009; 8: 741–754.
- National Stroke Foundation. *The economic impact of stroke in Australia*. Report, National Stroke Foundation, Australia, 2013.
- Kwakkel G, Kollen BJ, van der Grond J, et al. Probability of regaining dexterity in the flaccid upper limb. *Stroke* 2003; 34: 2181–2186.
- Nijland RHM, van Wegen EEH, Harmeling-van der Wel BC, et al. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery. *Stroke* 2010; 41: 745–750.
- Takeuchi N and Izumi S-I. Rehabilitation with post-stroke motor recovery: a review with a focus on neural plasticity. *Stroke Res Treat* 2013; 2013: 128641.
- Hatem SM, Saussez G, Della Faille M, et al. Rehabilitation of motor function after stroke: a multiple systematic review focused on techniques to stimulate upper extremity recovery. *Front Hum Neurosci* 2016; 10: 442.
- Harris JE, Eng JJ, Miller WC, et al. A self-administered Graded Repetitive Arm Supplementary Program (GRASP) improves arm function during inpatient stroke rehabilitation: a multi-site randomized controlled trial. *Stroke* 2009; 40: 2123–2128.
- Connell LA, McMahon NE, Watkins CL, et al. Therapists' use of the Graded Repetitive Arm Supplementary Program (GRASP) intervention: a practice implementation survey study. *Phys Ther* 2014; 94: 632–643.
- Dawson A, Knox J, McClure A, et al. *Canadian best practice recommendations for stroke care*. Report, Heart and Stroke Foundation, Canada, 2013.
- Hebert D, Lindsay MP, McIntyre A, et al. Canadian stroke best practice recommendations: stroke rehabilitation practice guidelines, update 2015. *Int J Stroke* 2016; 11: 459–484.
- Jones TA. Motor compensation and its effects on neural reorganization after stroke. *Nat Publ Gr* 18. Epub ahead of print 2017. DOI: 10.1038/nrn.2017.26.
- Levin MF, Kleim JA and Wolf SL. What do motor 'recovery' and 'compensation' mean in patients following stroke? *Neurorehabil Neural Repair* 2008; 23: 313–319.
- Sigrist R, Rauter G, Riener R, et al. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon Bull Rev* 2013; 20: 21–53.
- Levin MF, Weiss PL and Keshner EA. Emergence of virtual reality as a tool for upper limb rehabilitation: incorporation of motor control and motor learning principles. *Phys Ther* 2015; 95: 415–425.
- Duff M, Yinpeng Chen Y, Attygalle S, et al. An adaptive mixed reality training system for stroke rehabilitation. *IEEE Trans Neural Syst Rehabil Eng* 2010; 18: 531–541.

16. Tseklevs E, Paraskevopoulos IT, Warland A, et al. Development and preliminary evaluation of a novel low cost VR-based upper limb stroke rehabilitation platform using Wii technology. *Disabil Rehabil Assist Technol* 2014; 11: 1–10.
17. Cacchio A, De Blasis E, De Blasis V, et al. Mirror therapy in complex regional pain syndrome type 1 of the upper limb in stroke patients. *Neurorehabil Neural Repair* 2009; 23: 792–799.
18. Thielman G. Rehabilitation of reaching poststroke. *J Neurol Phys Ther* 2010; 34: 138–144.
19. Valdés BA and Van der Loos HFM. Biofeedback vs. game scores for reducing trunk compensation after stroke: a randomized crossover trial. *Top Stroke Rehabil* 2018; 25: 96–113.
20. Alankus G and Kelleher C. Reducing compensatory motions in video games for stroke rehabilitation. In: *Proceedings of the 2012 ACM annual conference on human factors in computing systems - CHI '12*. New York: ACM Press, 2012, p. 2049.
21. Proffitt R and Lange B. Considerations in the efficacy and effectiveness of virtual reality interventions for stroke rehabilitation: moving the field forward. *Phys Ther* 2015; 95: 441–448.
22. Secoli R, Milot M-H, Rosati G, et al. Effect of visual distraction and auditory feedback on patient effort during robot-assisted movement training after stroke. *J Neuroeng Rehabil* 2011; 8: 21.
23. van Vugt FT, Kafczyk T, Kuhn W, et al. The role of auditory feedback in music-supported stroke rehabilitation: A single-blinded randomised controlled intervention. *Restor Neurol Neurosci* 2016; 34: 297–311.
24. Yeh S-C, Lee S-H, Chan R-C, et al. The efficacy of a haptic-enhanced virtual reality system for precision grasp acquisition in stroke rehabilitation. *J Healthc Eng* 2017; 2017: 1–9.
25. Molier BI, Van Asseldonk EHF, Hermens HJ, et al. Nature, timing, frequency and type of augmented feedback; does it influence motor relearning of the hemiparetic arm after stroke? A systematic review. *Disabil Rehabil* 2010; 32: 1799–1809.
26. van Vliet PM and Wulf G. Extrinsic feedback for motor learning after stroke: what is the evidence? *Disabil Rehabil* 2006; 28: 831–840.
27. Srivastava A, Taly AB, Gupta A, et al. Post-stroke balance training: Role of force platform with visual feedback technique. *J Neurol Sci* 2009; 287: 89–93.
28. Wei S-E, Ramakrishna V, Kanade T, et al. Convolutional pose machines. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 4724–4732. New York: IEEE.
29. Wang Q, Kurillo G, Ofli F, et al. Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect. In: *2015 international conference on healthcare informatics*, 2015, pp. 380–389. New York: IEEE.
30. Morone G, Pisotta I, Pichiorri F, et al. Proof of principle of a brain-computer interface approach to support post-stroke arm rehabilitation in hospitalized patients: design, acceptability, and usability. *Arch Phys Med Rehabil* 2015; 96: S71–S78.
31. Warland A, Paraskevopoulos I, Tseklevs E, et al. The feasibility, acceptability and preliminary efficacy of a low-cost, virtual-reality based, upper-limb stroke rehabilitation device: a mixed methods study. *Disabil Rehabil* 2018; 12: 1–16.
32. Wolf SL, Winstein CJ, Miller JP, et al. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke. *JAMA* 2006; 296: 2095.
33. Lum PS, Burgar CG, Shor PC, et al. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Arch Phys Med Rehabil* 2002; 83: 952–959.
34. Fugl-Meyer Assessment of Sensorimotor Recovery After Stroke (FMA). Canadian Partnership for Stroke Recovery, <https://www.stroking.ca/en/assess/fma/> (2016, accessed 1 February 2016).
35. Star Cancellation Test. Canadian Partnership for Stroke Recovery, <https://www.stroking.ca/en/assess/sct/> (2016, accessed 1 February 2016).
36. Neurorehabilitation Research Program F of M. GRASP Manuals and Resources | Neurorehabilitation Research Program. University of British Columbia, 2017.
37. Microsoft. Kinect hardware. Microsoft, <https://www.microsoft.com/kinect> (2016, accessed 1 February 2016).
38. Affairs AS for P. System Usability Scale (SUS), <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (2013, accessed 18 November 2018).
39. Brooke J. SUS - A quick and dirty usability scale, Report, Redhatch Consulting Ltd., UK.
40. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; XX: 37–46.
41. Karros DJ. Statistical methodology: II. Reliability and validity assessment in study design, Part B. *Acad Emerg Med* 1997; 4: 144–147.
42. McColl D and Nejat G. Affect detection from body language during social HRI. In: *2012 IEEE RO-MAN: the 21st IEEE international symposium on robot and human interactive communication*, 2012, pp. 1013–1018. New York: IEEE.
43. McHugh ML. Interrater reliability: the kappa statistic. *Biochem medica* 2012; 22: 276–82.
44. Michaelsen SM and Levin MF. Short-term effects of practice with trunk restraint on reaching movements in patients with chronic stroke: a controlled trial. *Stroke* 2004; 35: 1914–1919.
45. Wu C, Chen Y, Chen H, et al. Pilot trial of distributed constraint-induced therapy with trunk restraint to improve poststroke reach to grasp and trunk kinematics. *Neurorehabil Neural Repair* 2012; 26: 247–255.
46. Alt Murphy M, Willén C and Sunnerhagen KS. Responsiveness of upper extremity kinematic measures and clinical improvement during the first three months after stroke. *Neurorehabil Neural Repair* 2013; 27: 844–853.
47. Brokaw EB and Brewer BR. Development of the home arm movement stroke training environment for rehabilitation (HAMSTER) and evaluation by clinicians. In: Shumaker R (ed.) *Virtual, augmented and mixed*

- reality. *systems and applications*. Berlin, Heidelberg: Springer, 2013, pp.22–31.
48. Institute for Computer Sciences S-I. In: *2012 6th international conference on pervasive computing technologies for healthcare and workshops*, San Diego, CA, USA, 21–24 May 2012.
 49. Weiss PL, Kizony R, Elion O, et al. Development and validation of tele-health system for stroke rehabilitation. In: *Proceedings of 9th international conference on disability, virtual reality and associated technologies*, Laval, France, 10–12 September 2012.
 50. van Andel CJ, Wolterbeek N, Doorenbosch CAM, et al. Complete 3D kinematics of upper extremity functional tasks. *Gait Posture* 2008; 27: 120–127.
 51. Butler EE, Ladd AL, Louie SA, et al. Three-dimensional kinematics of the upper limb during a Reach and Grasp Cycle for children. *Gait Posture* 2010; 32: 72–77.
 52. Petzold B, Zaeh MF, Faerber B, et al. A study on visual, auditory, and haptic feedback for assembly tasks. *Presence Teleoperators Virtual Environ* 2004; 13: 16–21.
 53. Cirstea MC and Levin MF. Compensatory strategies for reaching in stroke. *Brain* 2000; 123: 940–953.
 54. Loureiro RCV, Harwin WS, Nagai K, et al. Advances in upper limb stroke rehabilitation: a technology push. *Med Biol Eng Comput* 2011; 49: 1103–1118.
 55. Coupar F, Pollock a, La L, et al. Home-based therapy programmes for upper limb functional recovery following stroke (Review) Home-based therapy programmes for upper limb functional recovery following stroke. *Cochrane Database Syst Rev* 2012; 5: 5–7.
 56. Pollock A, Farmer SE, Brady MC, et al. Interventions for improving upper limb function after stroke. In: Pollock A (ed.) *Cochrane database of systematic reviews*. Chichester, UK: John Wiley & Sons, Ltd, 2014, p.CD010820.
 57. Laver KE, Schoene D, Crotty M, et al. Telerehabilitation services for stroke. In: Laver KE (ed.) *Cochrane database of systematic reviews*. Chichester, UK: John Wiley & Sons, Ltd, 2013.
 58. Liston RAL, Brouwer BJ, Colomer C, et al. Reliability and validity of measures obtained from stroke patients using the balance master. *Arch Phys Med Rehabil* 1996; 77: 425–430.
 59. Saposnik G and Levin M. Virtual reality in stroke rehabilitation: a meta-analysis and implications for clinicians. *Stroke* 2011; 42: 1380–1386.
 60. Zhang D, Shen Y, Ong SK, et al. An affordable augmented reality based rehabilitation system for hand motions. In: *2010 international conference on cyberworlds*, 2010, pp. 346–353. New York: IEEE.
 61. Laver K, George S, Thomas S, et al. Virtual reality for stroke rehabilitation: an abridged version of a Cochrane review. *Eur J Phys Rehabil Med* 2015; 51: 497–506.

Appendix I

	Strongly disagree	Strongly agree										
1. I think that I would like to use this system frequently.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
2. I found the system unnecessarily complex.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
3. I thought the system was easy to use.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
4. I think that I would need the support of a technical person to be able to use this system.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
5. I would imagine that most people would learn to use this system very quickly.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
6. I found the system very cumbersome to use.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
7. I felt very confident using the system.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
8. I needed to learn a lot of things before I could get going with this system.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
9. I feel that the system can accurately track my body.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								
10. I do not think that this system will help me complete my exercises better.	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 20%; height: 20px;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td style="text-align: center;">1</td> <td style="text-align: center;">2</td> <td style="text-align: center;">3</td> <td style="text-align: center;">4</td> <td style="text-align: center;">5</td> </tr> </table>							1	2	3	4	5
1	2	3	4	5								

Open ended questions:

- What do you like/dislike about the system?

- If the system is available at home today, would you use it? Why?

- Do you feel any technical barriers that make the system more difficult to use?

- Do you like the first (camera) or second (stickman) system better? Why?

- Would you like to be reminded of (through video or audio cue) when compensations are detected?

- Instead of asking you to read this instruction manual, do you prefer the idea of recording the correct way of doing exercises and display it side by side with participant's arm movement?

- Other comments about the system:

Appendix 2

Category assignment for exercises from the Graded Repetitive Arm Supplementary Program

Category 1 (minimal interference with looking at the screen). These exercises involve arm and hand movements that could be achieved without observing themselves and their body is always facing forward (towards the screen).

- Shoulder shrug
- The twist (arms)
- Shoulder exercise (arm to the front and side)
- Elbow exercises
- Wrist exercises
- Hand the wrist stretch
- Grip power
- Squeeze

Category 2 (partial interference with looking at the screen)

- (1) Exercises that require focusing on the task during initial setup in each repetition of the exercises but not during the exercises:
 - Finger power: requires rolling the putty into a rope at the beginning of each repetition
 - Wash cloth: requires folding towel initially
 - The twist (putty): same as “Finger power”
 - Finger strength: requires rolling the putty into a ball at the beginning of each repetition
 - Jars: requires putting the lid back on before starting the next repetition
 - Drying off: requires folding towels first
- (2) Exercises that do not require hand–eye coordination but involve body movements that would

prevent subjects from looking at the screen during the exercise:

- Total arm stretch: Requires bending forward 90 degrees (head would be facing down)
- Push-up: requires leaning chest forward, which makes it more difficult to look at the screen
- Chair-ups: people tend to naturally lean forward while lifting their own bodies off the chair, which makes it difficult to look at the screen in that posture

Category 3 (substantial interference with looking at the screen). These exercises often involve aiming at a target, which require constant focus to achieve the task.

- Waiter cup: requires placing cup in circles on the target board
- Waiter-bean bag: requires placing bean-bag in circles on the target board
- Advanced waiter: similar to waiter exercises above
- Button: require putting button through holes
- Cutting: does not involve aiming but is highly unusual to not focus on the task
- Waiter-ball: requires place ball in circles on the target board
- Pouring
- Laundry: entire exercise involves folding towels
- Hanging up the clothes: requires clipping cloth pins on the edge of a cup
- Lego: requires stacking Lego pieces
- Block towers: requires stacking blocks without making it fall
- Pick up sticks: requires picking up sticks and placing it in the cup
- Paper clips: requires stringing clips together
- Poker chips: requires flipping poker chips