Towards standardized processes for physical therapists to quantify patient rehabilitation

Aisling Kelliher
Virginia Tech
Blacksburg, VA
aislingk@vt.edu

Setor Zilevu
Virginia Tech
Blacksburg, VA
zilevus@vt.edu

Thanassis Rikakis
Virginia Tech
Blacksburg, VA
rikakis@vt.edu

Tamim Ahmed
Virginia Tech
Blacksburg, VA
tamimahmed@vt.edu

Yen Truong
Virginia Tech
Blacksburg, VA
yen1027@vt.edu

Steven L Wolf
Emory University
Atlanta, Georgia
swolf@emory.edu

ABSTRACT
Physical rehabilitation typically requires therapists to make judgements about patient movement and functional improvement using subjective observation. This process makes it challenging to quantitatively track, compute and predict long-term patient improvement. We therefore propose a novel methodical approach to the standardized and interpretable quantification of patient movement during rehabilitation. We describe the expert-led development of a movement assessment rubric and an accompanying quantitative rating system. We present our movement capture and annotation computational tools designed to implement the rubric and assist therapists in the quantitative documentation and assessment of rehabilitation. We describe results from a movement capture study of the tool with nine stroke survivors and a movement rating study with four therapists. Findings from these studies highlight potential optimal methodical process paths for individuals engaged in capturing, understanding and predicting human movement performance.

Author Keywords
Home based rehabilitation therapy; stroke rehabilitation; human movement capture; human movement assessment

CSS Concepts
Human-centered computing~Interaction design process and methods ~*Interface Design Prototyping*

INTRODUCTION
With the aging of the global population [7], there is an increasing need for effective and accessible rehabilitation services for debilitating illnesses and injury such as stroke [18]. Stroke is the leading cause of long-term disability in the United States, with approximately 800,000 people experiencing a stroke each year [8]. Mortality rates have decreased in the past two decades as a result of better control of vascular risk factors [28, 38]. This in turn means that there is a substantial increase in the number of stroke survivors, making stroke management a significant therapeutic, financial, and social challenge. Large-scale studies point to the effectiveness of long-term rehabilitation therapy in facilitating recovery [22, 53]. While the majority of recovery is reported to take place in the first three months after stroke [51], there is a growing body of research that indicates that recovery, particularly of the upper extremity, can also occur many years after stroke [21], particularly within the context of learning adaptation strategies [1, 33]. However, the costs, availability of facilities and experts, as well as transportation to clinical facilities on a regular basis over an extended time period, limit the amount of supervised therapy that stroke survivors can receive in the clinic [15].

CHI '20, April 25–30, 2020, Honolulu, HI, USA

ACM ISBN 978-1-4503-6708-0/20/04…$15.00
https://doi.org/10.1145/3313831.3376706

Technology assisted rehabilitation therapy is proposed as a potentially key cost-effective avenue for improving health and wellness outcomes, as a supplement to, or even as the primary mode, of therapy delivery [3, 2, 36]. Precise movement data acquisition and analysis with computational tools could assist therapists in the automated assessment and evidence-based customization of rehabilitation. In addition, this approach could support the efficient semi-automated delivery of treatment at the home at scale [6]. However, there are several key challenges that need to be addressed in order to realize these promising outcomes.

Clinical rehabilitation measurement tests such as Fugl-Meyer [18], Action Research Arm Test [37], or the Wolf Motor Function Test [53] are largely observational in nature [39]. These scaled assessment tests focus on function, which is understandable given that both providers and payers (insurance companies and Medicare/Medicaid Centers) are primarily interested in recovery of function. Therapists cannot provide detailed interpretations of their ratings when using these clinical tests. The therapist provides an overall rating of function based on their observation and informed by their prior experience, but they cannot directly indicate how function changes and why. Even highly trained clinicians cannot attend to the totality of complex pathological movement or compare such observations to a normative value, which creates significant variation in rating assessments [12]. This means that there is a lack of detailed, standardized and interpretable assessment data that can support evidence-based adaptation of therapy and the training of semi-automated therapy systems.

Our key contribution is the process developed by our interdisciplinary team of designers, engineers, and clinicians in crafting a series of methodical steps aimed at moving towards realizing standardized and quantified movement assessment during therapy. Working closely with therapists, rehabilitation experts, and stroke survivors, we began by defining a set of *standardized rehabilitation activities* that are translatable to activities of daily living (ADL). These activities are designed to be easy to document and analyze, and significantly extendable to support long-term adaptive training in the clinic or the home. We then developed *a simple and robust capture tool for these rehabilitation activities* that is straightforward for therapists and patients to use in the clinic. The data captured by that tool provided the basis for expert therapists to then develop a *standardized and interpretable assessment rubric* to provide the processes and operational definitions of terms used to evaluate movement quality and inform rating assessment. The rating rubric is based on a *segmentation vocabulary* encapsulating all of the activities. This segmentation vocabulary comprises the operational definitions of each movement segment and the most significant movement features associated with each segment. To bring these assessment ideas into practice, we finally developed *an annotation tool* to assist therapists in the interpretation and standardized quantified rating of the movement through the application of the assessment rubric.

We briefly describe results from a study with nine stroke survivors and a team of therapists in evaluating the set of standardized rehabilitation activities, and in capturing data of upper extremity movement. Following this, we detail a sequence of in-situ workshops and online meeting sessions where we worked with five therapists on the development of the assessment rubric. We also describe a study with four occupational and physical therapists evaluating our assessment and annotation tool through rating sessions of captured data.

Findings from these studies highlight the importance of extensive rater experience in developing standardized approaches and the sensitizing effect of exposure to the support tool in reducing assessment bias and increasing standardization of assessment by therapists. Our findings also lead us to suggest design guidelines for developing interfaces requiring assessment precision, and possible best practices for surfacing hidden assessment intuitions that can lead to interpretable movement ratings. Our work here applies to a specific rehabilitation context, but we believe that it can also apply to other health contexts such as knee replacement (Total Knee Arthroplasty (TKA)). With TKA patients for example, the lower extremity compensatory patterns that are used to address muscle weakness and residual discomfort result in highly variable movement patterns and therefore present similar challenges for therapeutic interpretation and intervention [48, 56]. Standardized quantitative methods, processes and tools could assist in this rehabilitation context and also other areas where the robust automated capture and evaluation of human performance is important including education, safety and security, mediated arts performance, and the human control of complex operations.

PRIOR WORK

Upper extremity stroke rehabilitation

Almost 80% of stroke survivors experience some form of upper extremity weakness after acute stroke and functional recovery can be limited due to the complexity of limb management post stroke [32]. Physical and occupational therapy training built on motor learning principles (e.g. distributed practice of exercises with repetition and variability) has demonstrated increased likelihood of recovery [29]. Here, the therapist is the daily driver of the clinical recovery plan, beginning with a detailed clinical assessment of the patient's functional capacity using a toolbox of measures [18, 37, 53]. Following assessment, a therapy plan is created that typically includes a variety of physical activities such as strength training, range of motion exercises, and dexterity and bimanual training.

Technology assisted rehabilitation therapy has the potential to greatly expand the ability of the therapist to impact patients. There are a diverse variety of promising technology assisted systems currently in development, including virtual and augmented reality systems [16, 41], robotic assist systems [43], tangible systems [4], wearable sensors [17, 47], and computer vision techniques [44, 49]. Several of these systems seek to address key issues including attempting to replicate the functions of the therapist in an automated system or implementing adaptive rehabilitation through low cost systems. However, the lack of standardized understanding of how and why function changes limits the effectiveness any approach in adaptive long-term therapy at scale [5, 42]. The complexity of the assessment problem indicates that technology alone cannot solve it - rather it requires a socio-technical approach driven by iterative, human-centered design.

Rehabilitation movement capture and analysis

Prior work shows that tracking and analyzing low-level kinematics during rehabilitation training in the clinic can result in an automated assessment of performance that is highly correlated with expert assessment [10, 46, 50]. However, detailed tracking of movement through marker-based capture [19] or complex exoskeletons [43] is costly, complex, and obtrusive. For example, data gloves may not fit those with hand contractures or joint inflammation, especially among stroke survivors, and markers on hands can hinder performance of detailed functional activities [23]. Furthermore, different low-level features are known to work for different types of movement quality assessment and a combination of different sensors may be needed to acquire the right features. Finally, complex technological systems are challenging for the therapists to use and are rarely part of their educational curriculum, thus they are not widely adopted [11].

For several years, the relatively inexpensive Kinect camera tracking system supported multiple promising rehabilitation systems [24, 44, 52] but with manufacturing of the Kinect discontinued, together with other use-case criticisms [49], alternative low-cost movement capture approaches are needed. Tracking of performance through a small array of video cameras (up to 2 cameras) is low cost, low effort, and unobtrusive, but does not produce reliable tracking of all necessary low- level kinematics [45]. Similar issues with tracking reliability are noted also in some tangible systems where the unique movement profile of each patient creates challenges in terms of detecting movement and counting task repetitions [34]. Tradeoffs are therefore required in terms of developing a movement capture system that is effective, relatively inexpensive, and straightforward to install and operate.

Rehabilitation movement assessment

A key stroke rehabilitation challenge is deciding *what* the therapist and/or the support system should monitor and pay attention to during therapy activities. The performance of an activity is influenced by a large number of parameters including physical ability [20], cognitive state [12], neurological function, and personal and environmental factors. As it is practically impossible to measure and attend to all such factors, simplified classification approaches are required for discriminating and assessing activity behaviors.

Despite the existence of validated clinical measures such as the Wolf Motor Function Test [53] and the Fugl-Meyer [18], a lack of consensus remains among physical therapists regarding the standardized, quantitative and interpretable evaluation of movement quality components and the influence of such components on overall functional ability [29, 36]. In practice, therapists typically select which components to focus on based on their individual experience [25] and training, rather than a standardized ontology of component level labels for movement quality [54]. These factors contribute to the lack of large scale standardized interpretable assessment data that can be used for the evidence based adaptation of therapy or for training semi-automated rehabilitation systems to reproduce both a complex therapy experience and a reliable approach for movement quality assessment.

METHOD

We present here our multistep iterative approach in developing a methodical process for standardizing i) a movement assessment rubric, ii) a rubric implementation protocol, and iii) a therapist support tool to assist with the quantified annotation of movement videos. Our methodical approach originates deep within the therapeutic practice we aim to support. This approach is human-centered in that it places the rehabilitation experts and the particularities of their practice in the foreground, in order to ascertain how and when a computational support tool might be helpful.

Standardizing rehabilitation activities and objects

Our team has conducted embedded ethnographic work with rehabilitation experts specializing in the therapy of the upper extremity of stroke survivors for over 12 years [5, 9, 14, 26, 35]. This observational work, conducted in hospitals, clinics, and research labs across multiple different projects, revealed that therapists use a limited set of training exercises and a limited set of training objects during therapy (e.g. reach for a cup, fold a towel, brush hair etc.). These exercises and objects are typically structured so as to generalize to many activities of daily living (ADLs). The therapists explained that this optimizes the limited training time available for each patient (maximum two hourly sessions a week during recovery) by constraining the training while still having generalizable outcomes. While the sets of objects we observed were limited, the process for selection and inclusion of an object in a training session did not follow a standardized process.

Given the critical focus on translatable learning during rehabilitation training [30, 42], there is a need for a set of objects to systematically relate to as many types of ADLs as possible. This creates efficiencies in terms of cost and production, while also opening up a potential creative problem-solving space with regards to how the objects might relate and integrate with one another [55]. The set of activities performed with the objects need to scale in complexity, whereby different subsets can be used for different levels of impairment or different stages of training. The set of objects also need to be limited so as to facilitate standardized assessment, but also extendable so they can support long-term therapy and be potentially acceptable to the wider community of therapists and patients [26].

In addition, the movement of these objects, and the relationship between the patient and the moving objects need to be easily and consistently documented through a user-friendly capture interface (i.e. 1-2 low cost video cameras connected to a tablet computer). Here a constrained and standardized set of activities and objects that are easily captured can help reduce the human and computational challenge for assessment (and potentially facilitate low-cost semi-automated therapy).

Designing a set of standardized activities and objects

The rehabilitation experts and industrial designers on our team worked together on this standardization challenge. Elements of this work are presented in a TEI 2019 work in progress paper [26], which just focused on the iterative design of the objects themselves and the response of the therapists and patients (the same participants as in this study) to the objects. Here we extend our brief description of that work by introducing the design of the capture tool, the movement activities and relationship of the objects to the therapy activities, and the assessment rubric and annotation tool.

The rehabilitation team proposed 12 movement activities ranging from reach an object; to reach and grasp an object; to reach and grasp and transport an object(s); and finally to reach and grasp and manipulate an object(s). The 12 activities were designed to map to important activities of daily living (ADLs) including eating, drinking, putting on clothes, personal hygiene, and typical household chores. Table 1 lists the ADL mappings for each of the training activities, which increase in complexity from 1 – 12.

1. Press a button	7. Write with a pen
2. Grasp a handle	8. Zip up a jacket
3. Take a drink	9. Put arm in a jacket sleeve
4. Pick up a fork	10. Close a lid
5. Pull a door handle	11. Screw the lid on a jar
6. Tidy objects by stacking	12. Fold a towel

Table 1. The ADLs mapped to the 12 training activities.

The next challenge was designing a modular and easily capturable set of objects for realizing these activities. Building on the set of commonalities identified across different movement related gold standards (described in our prior publication [25]) and on insights derived from other tangible systems [4, 13, 26], the team produced a series of six objects. The objects, depicted in Figure 1, are intended to be perceived as having a variety of affordances [40], meaning each can be grasped and manipulated in a variety of ways. They aim to approximate objects encountered in the home (e.g. the purple object could be used as an iron or perhaps as the handle on a fridge), yet are strange enough that they may require some cognitive effort to work out how to approach and manipulate them. The objects can be stacked or screwed into each other, and are 3D printed using a combination of hard PLA and soft NinjaFlex material.

The training activities make use of the objects to approximate the ADLs in Table 1. For example, from screwing the screw top into the base object ("Screw the lid on a jar"; to moving the hourglass object in sweeping circles ("Write with a pen"); to extending an arm in between two objects "Put arm in a jacket sleeve").

Performing and capturing the standardized activities

Working with the therapists, we developed a simple two camera set-up connected to a tablet computer that capture activities from angles representing where the therapists on our team typically like to stand or sit during therapy. One video camera positioned beside the participant captures a sagittal side-view of the shoulders and torso, while a table-

Figure 1. Three base objects, one flat top, one screw top, and one soft top.

mounted video camera focuses on the wrist and fingers during manipulation and transportation activities.

The tablet is placed on a custom designed mat, with etchings on it denoting the range of the activity space (full capture setup is shown in Figure 3). We created an easy-to-use capture tool to deliver the activity protocol on the tablet and initiate recording on the cameras (Figure 2 depicts our current version of the capture application interface). The patient manages the "running" of the training, using their unaffected limb to navigate through the protocol on the tablet. This setup requires minimal assistance from the therapist in the clinic, which can be further refined for unassisted use in the home.

The capture tool first presents an introductory video explaining the activities to be performed and the related object setup (Fig. 2a). The patient is then invited to examine the objects and "rehearse" a number of the assigned activities as a warm up exercise. By allowing them to rehearse with the system, the patient can become more familiar with the interface features in a relaxed manner. Once the activity study begins, the patient is instructed on the object selection and object placement for the first activity (Fig. 2b). The patient then watches a video of the assigned activity being performed. When they are ready, they are first reminded to sit back in their chair with feet on the ground (Fig. 2c), then they "tap" the yellow screen and begin the activity (Fig. 2d). When they complete the activity, they tap the complete activity screen. For each of the 12 activities, they are asked to perform them four times to the best of their ability. For the purposes of this pilot study, at the end of each set of four, the patient is presented with a short multiple choice question which asks if the activity reminds them of any of three activities (e.g. press a button, pick up a phone, close a container etc.), which includes the intended mapped activity, an activity from the MAL, and an unrelated general activity.

Pilot Study and Data Capture

We conducted a pilot study to evaluate the activities, objects and capture system, and to also capture video data of the selected activities. The study took place at the Emory

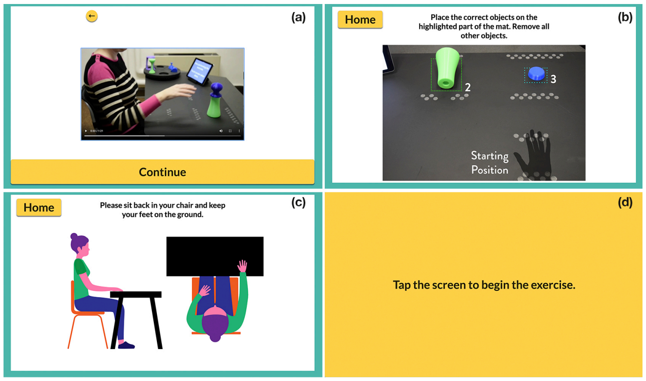

Figure 2. Screenshots of the current capture interface showing the introductory video screen, object selection and setup screen, a correct posture reminder screen, and the screen to start the activity and begin recording

Rehabilitation Hospital in Atlanta, Georgia and involved nine stroke survivors. Two women and seven men participated in the collection process; two with moderate impairment and seven with mild to moderate impairment after stroke. The definition of mild stroke impairment for patient inclusion in the study was an upper extremity Fugl-Meyer score of greater than 55, while a Fugl-Meyer score between 30-55 determined moderate stroke impairment. Three of the patients attended three capture sessions each over a two-month period, while the remaining six patients completed one session. The data collection process was approved by our Emory Hospital Institute Review Board and the patients were compensated for their travel expenses to the hospital and for their time ($50 per trial session). The patients were informed that this was not a clinical trial or a formal rehabilitation session, but rather a data capture session with an experimental setup. The data collection sessions were supervised directly by a physical therapist, and observed by the rehabilitation expert and the development team.

Study Results

Six of the participants were able to complete all four repetitions of each of the twelve activities, while three of the more moderately impaired participants struggled to complete the more complex movement activities. Different patient profiles (e.g. level of impaired sensation, increased spasticity, limitations in shoulder range of motion etc.) influenced the abilities of the participants to engage with some of the objects. In those cases, the therapist intervened in the recording session, and used the menu feature in the application to move to a different activity. Three sessions were also not fully completed because the capture process took longer than anticipated and the participants had to leave. Overall, we recorded 618 distinct activity attempts by the nine participants.

The patients had different understandings about the possible connections between the 12 activities and their intended related activity of daily living, but overall, for at least ten of the activities, the majority of the patients recognized and agreed with the intended ADL mapping [26].The patients also agreed on the "ramping" up of the activities in terms of complexity as they moved through the activities. For participants with less impairment, the rising complexity occurred not as much in the physical domain, but in the cognitive. For example, P7 (mild impairment) stated in the debrief interview: "*It didn't challenge my strength but it did challenge my pea brain (laughs). Remembering what to do was - the further we got into it, the more I had to think, which is real good for me. It's the deeper I got into it, the more it went from the physical to the brain.*"

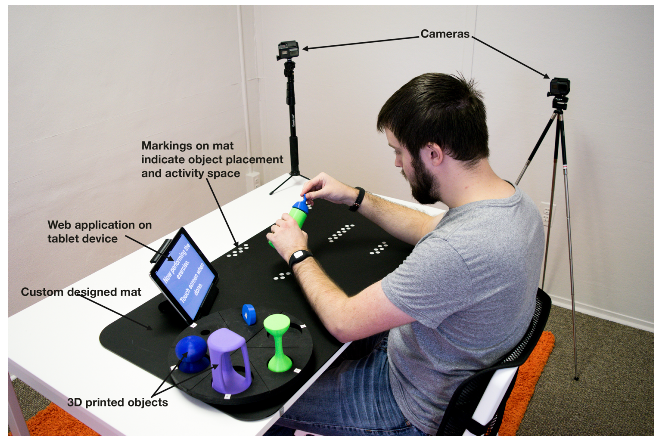

Figure 3. A team member combines two objects to screw them together during system testing

In another of the daily debrief sessions after data collection, the supervising therapist noted that because each of the patients had different grasp abilities, they might agree that the exercise mapped to some activity, but it was possibly very different from what our development team were imagining. This presents an opportunity for working and co-designing with patients in customizing objects based on their abilities, perceptions, and goals. The participants in our study offered many suggestions for additional activities and objects to include such as drying with a towel, opening a lock with a key, and cutting with a knife.

We wanted to build on this feedback and also further probe the limitations of our standardization approach. We created an online survey and recruited three other therapists (and the original supervising therapist) to participate. In the interactive survey, the therapists watched the 12 activity instruction videos and were asked the same multiple choice question as the participants in the study. They were also asked to suggest any important activities that our set of 12 was missing. The majority of the therapists linked the intended ADL to the depicted one for nine of the activity videos. In addition, they also suggested including a cutting/eating activity, buttoning a shirt, and opening a lock as additional activities to include. In response we created an interlocking chopstick-like object, and a lock/key mechanism for consideration in the next stage of our work (see Figure 4). We are also creating a hairbrush type object in response to P1 who wondered why we didn't include any above-the-shoulder activities. This brings our current set of objects to 9 with 15 accompanying activities.

Figure 4. Current set of modular rehabilitation objects including three base objects, two top objects, one button, and two object sets for bimanual manipulation.

With a captured dataset in hand from the Emory study, we next progressed to the vital stage of developing a standardized interpretation approach for assessing upper extremity movement. This necessitated both defining a segmentation methodology for the activity components, and an assessment rubric for evaluating the overall activity, together with the activity segments. We began drafting our approach during the pilot study based initially on our observations of the therapists and extensive discussions with the rehabilitation expert on our team.

Standardized segmentation vocabulary

Our segmentation approach grew from an observation about a key generalizing technique used by expert rehabilitation therapists. To achieve the generalizability of a limited set of activities to many more activities of daily living, they use an implicit segmentation of the activities into types of segments that hold true across all the activities. For example, they look at the characteristics of initiation or termination of movement across all different types of activities. Expressing all of the rehabilitation activities as permutations of a limited set of segment types can potentially assist the therapist in more concise measurement and assessment of the activities. We worked closely with the rehabilitation experts on our team to try and take an implicit (and noisy) segment codification, and turn it into an explicit and standardized representation that can be captured using minimal low cost computing infrastructure and can be manipulated by both humans and computers. This resulted in representing the standardized segment vocabulary through a state machine that can express all activities through different paths between the states.

Figure 5 depicts the five movement segments and potential paths in between that can represent our 15 current activities (and which can easily be extended). The first three movement stages (Initiation, Progression, and Termination) are interconnected parts of every reach and touch/grasp activity and are thus grouped together.

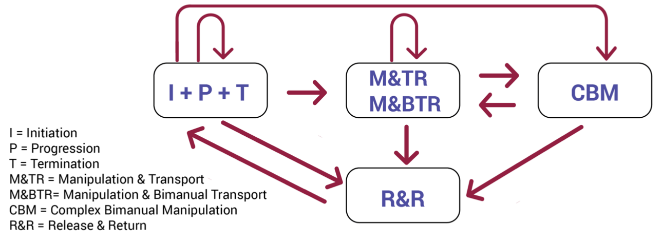

Figure 5. Diagram depicting the five potential movement stages for the training activities in our system

After the termination movement, there are multiple possible movement paths depending on the exercise. For example, the sixth exercise in our sequence involves stacking an object on top of another object, pausing, and then removing it again. It is represented by the sequence: Initiation, Progression, Termination/IPT (reach out and grasp object A); Manipulate &Transport/M&TR (pick up object A and place it on object B); Release & Return/R&R (let go of object A and return hand to the starting position).

Standardized assessment rubric for interpretable ratings

The expert therapists on our team created a preliminary draft of the standardized assessment rubric. Table 3 depicts the most significant movement features for each segment identified in the segmentation vocabulary (first two rows in Table 3), The rubric provides rules for assessing each activity segment based on the evaluation of the key movement features per segment and rules for assessing the performance of the complete activity (i.e. was the assigned transportation of the object completed). It also provides rules for combining the scores of the segments and the assessment of the performance of the activity into an overall score of 1, 2, or 3. Limiting the available ratings to 3-4 numbers (0-3 or 1-3) is a well-established practice in stroke rehabilitation movement assessment tools (i.e. WMFT, ARAT) as it facilitates training of therapists in using the tool and efficient implementation of the rating. The main difference however with our rubric is that every overall score can be connected to detailed elements of movement function.

The rubric also establishes four operational definitions of terms used to evaluate movement quality and inform rating. *Appropriate* is defined as “the range, direction, and timing of the movement component for the training activity compared to that expected for the less impaired upper extremity.” *Digit positioning* is defined by “the volitional placement of relevant digits is representative of what would normally be expected for the activity.” *Trunk sway* means “the forward (translational) movement of the torso is appropriate for the activity. Finally, *aperture* means “the positioning of separation between the thumb pad and index finger pad is what is normally expected for the activity.”

<table>
<tr><td>INITIATION
Appropriate shoulder elevation
Appropriate shoulder flexion or abduction
Appropriate trunk stabilization</td><td>PROGRESSION
Appropriate range of motion for elbow
Appropriate trunk sway, flexion and rotation
Appropriate pronation/supination</td><td>TERMINATION
Wrist position aligned to task
Appropriate hand aperture
Appropriate digit positioning</td></tr>
<tr><td>MANIPULATION & (BIMANUAL) TRANSPORT
Appropriate finger positioning and orientation
Limb trajectory with appropriate smoothness and accuracy
Limb orientation (supination/pronation)
Appropriate trunk movement and position</td><td>COMPLEX BIMANUAL MANIPULATION
Appropriate finger positioning
Appropriate finger motion after positioning
Appropriate limb motion after positioning
Limb trajectory with appropriate accuracy</td><td>RELEASE AND RETURN
Appropriate digit movement during release
Elbow flexion timing reasonable
Appropriate trunk sway/rotation</td></tr>
<tr><td colspan="3">OVERALL ACTIVITY RATING SCALE
3: Activity was completed (fully with multiple attempts as needed) within reasonable time and with minimal movement impairment
2: Activity was either fully completed with minimal movement impairment but the length of execution was unreasonably long OR Activity was fully completed but the upper extremity collective movement had ONE element showing significant impairment
1: Activity was either not completed or activity was fully completed but the upper extremity collective movement had TWO or MORE elements showing significant impairment
0: Movement was not attempted</td></tr>
<tr><td colspan="3">ACTIVITY SEGMENT RATING SCALE
3: Segment was completed (fully with multiple attempts as needed) within reasonable time and all three or four movement quality elements important to the execution of the segment showed no compensation or minimal compensation
2: Segment was completed (fully with multiple attempts as needed) within reasonable time but ONE movement quality element from the three or four movement elements important to the execution of the segment showed significant impairment
1: Segment was either not completed or segment was completed but TWO or more movement quality elements from the three or four movement quality elements important to the execution of the segment showed significant impairment
0: Movement was not attempted</td></tr>
</table>

Table 3. The assessment rubric including important movement quality elements and the overall and the segment rating scales.

The development team created a simple annotation tool to present the captured activity videos to therapists for assessment using the standardized rating rubric. The prototype tool was developed in preparation for a weekend workshop with the visiting rehabilitation expert team. The tool has a web application fronted developed using a Node.js framework and a MongoDB backend database. Our current version of the tool interface is depicted in Figure 6. In the tool, footage from both camera angles is displayed to the therapists to allow observation of torso/shoulder movement and arm/hand/digit movement. The navigation buttons at the top of the screen allow the rater to move linearly through the segments in an activity, with the significant segment features presented below for labeling. The annotation tool provides the rater with the functionality to assign each complete activity, and each activity segment, a score of 1, 2, or 3.

Rating Workshop

The two-day workshop took place at Virginia Tech with the goal of solidifying our thinking on the rubric and assessment approach. The primary intent of this initiative was to evaluate and refine the preliminary segmentation vocabulary and assessment rubric and also to assess the efficacy of the rating interface. We invited four rehabilitation therapists (two occupational, two physical) to participate in the workshop, in addition to our rehabilitation medicine expert who participated in a consultancy role. Two of the therapists were approximately mid-career, while the other two therapists had considerably more experience.

Our team segmented the captured video data for uploading into the rating interface, using the segmentation vocabulary developed by the expert therapists. The experts also proposed a coverage plan for selecting videos of specific participants (variety of impairment profiles) and presenting them in a particular order to facilitate training at the workshop. This segmentation was done by hand, with team members watching the videos and then inputting the start and stop timecodes for each segment into the backend database of the rating tool. In total, 66 complete training activities and their individual segments were prepared for the workshop.

The development team organized the videos into three rating "sessions" (Session I, II, III) and also created a practice session to assist with training. This practice session comprised six videos of three different participants attempting the same two activities (the simple activity 2 and the more challenging activity 11). The purpose of this set was to provide an opportunity for the therapists to collaboratively rate the videos as a group and in so doing, familiarize themselves with the rubric, the rating interface, and the activities depicted in the videos. Rating Session I contained 12 videos of one mildly impaired patient completing the 12 activities in ascending order of challenge, and 12 videos of one mild-moderate patient also completing the 12 activities in order. The purpose here was to provide the therapist raters with an overall sense of the progression of the activities as completed by patients with different levels of impairment.

Rating Session II contained 24 different videos sampled across all of the nine participants and with the videos again presented in order from activity 1 to activity 12 (twice). The purpose here was to see the potential ordering influence of the videos across and between participants. Rating Session III contained 24 different videos sampled across all the participants, and with the videos presented in randomized activity order.

Workshop Outcomes

Most of the in-situ workshop time ultimately ended up dedicated to detailed and extensive discussion about the assessment rubric, and in particular, the most significant movement features to consider for each activity segment. This long discussion was nonetheless essential as it reinforced the complexity of the problem at the level of movement quality interpretation. Ultimately, this meant that the therapists only managed to complete the group practice session and Rating Session I before the workshop ended.

Based on review of the videos and extensive discussion, the expert therapists were satisfied with the segmentation approach. However, the assessment rubric was more extensively refined, in response to the review of the first 18 activity videos and their segments by the experts. A careful balance needed to be orchestrated between the level of detail (number of movement feature labels) and the desire not to overwhelm the rating therapists with options. Over the course of the two days, the set of significant features was focused and constrained, with no more than four possible movement features assigned per movement segment. The annotation tool managed to provide the therapists with a concrete way of "practicing" the assessment process in a discursive and reflective forum. Moving forward, our team wanted to preserve this reflective feedback look between the expert therapists and the tool developers.

Refining the assessment process

Following the workshop, our team created a more refined version of the annotation tool (see Fig. 6). This online version allows therapists in remote locations to view and rate the videos. This more robust web application included additional features including presenting the activity instruction video alongside the two camera angles of the participant videos; adding the ability for therapists to add comments and flag technical problems; providing a tabbed navigation scheme to move through the video segments; and adding a progress bar so the therapists would have a better sense of their progress status in a particular rating session.

Over the course of almost four weeks, the four therapists logged repeatedly onto the secure online application hosted at our institution. Once logged in, they were directed to begin rating the videos in Rating Session II. The videos themselves were stored on encrypted external hard drives that were sent by registered mail to each participating therapist. The annotation pulled the stored videos into the browser for viewing, but only the rating data was transferred back to our host server. This was done in order to protect the privacy and security of the patient data as per the permissions in our IRB protocol. Using this version of the tool the therapists were assigned to rate 72 patient activities for a total of 240 ratings.

Rating results

Our approach in analyzing the assessment process began at the level of examining the overall activity score, followed by examining the individual activity segment scores.

Figure 6. Screengrab of the online annotation tool showing the interface for rating the "total" score for an activity.

In both Session I and II, the four therapists were in full agreement for seven of the 24 rated activities in each session (for a total of 14 full agreements across 48 ratings). In Session I, three therapists agreed on one rating for 15 activities and the remaining therapist provided a +/-1 rating from the rating of the other three. In session I there were two instances where the four therapists split evenly across two ratings. In Session II, there were seven instances where three therapists agreed on one rating and the remaining therapist provided a +/-1 rating. There were also seven instances where the four therapists split evenly across two ratings. Finally, there were three instances where all three ratings were given by different therapists.

We observed several items of note in the rating data across Sessions I and II . It is possible to discern the signature styles of the therapists from the rating data. T1 rates the movement quality in a relatively lenient way compared to the other therapists. The ratings by T1 are higher than the mode rating for an activity in nine of the 48 instances. T2, the therapist with the most experience, is very consistent in their rating and disagreed with the mode in only two of the 48 activities. Similarly, T3, the next most experienced therapist only disagreed with the mode in four of the 48 instances. T4 was the youngest therapist and their ratings are more mixed, demonstrating variance from the mode both over and under on multiple occasions.

In Session II, the therapists rated training activities from all nine different participants. For the study patients rated highest and lowest on the Fugl-Meyer measure, there was greater agreement among the therapist raters. However for patients scoring more in the middle of the Fugl-Meyer range, we observed much greater variety in rating scores across therapists. We also note that rating disagreements become even more apparent at the level of the segment. For example, inter-rater consistency calculated using Krippendorf Alpha [31] is α=.608 for the overall activity scores, but drops to α=.45 for the individual segment scores.

In the spirit of fostering a reflective forum for our team of experts, we therefore wanted to surface the reasons *why* the therapists were in disagreement on certain ratings and determine if the problem was being generated by the therapist, the study participants, the rated activity, or the annotation tool itself. We needed their expert input again to help direct the next stage of iteration.

Rating Data Interpretation

We created a multi-part online survey for the four therapists and our rehabilitation expert, with sections on rating patient/activity order, impact of patient impairment profiles, interpretation of the disagreement in rating scores, and impact of therapist training and experience. The survey also presented the therapists with examples of the split (2X2) and 1, 2, 3 inter-rater disagreements for comment. The therapists were asked if they expected or were surprised about the increased level of disagreement when rating segments and when rating different participants, and asked for suggestions as to how to optimize the order and presentation of the activity videos for improving interrater reliability.

T1 noted *"It was difficult and somewhat unnatural as a therapist to break down movement to the degree that we were, especially between the initiation of movement to the progression of movement. It is very difficult to determine when initiation stops and progression begins, so it would not surprise me to have much variability there."* The challenge of rating segments was corroborated by similar comments by the other therapists. This expert input, along with the observed drop in interrater consistency at the segment level, confirms a core assessment challenge presented in our introduction. Even expert therapists are not able to consistently observe all detailed movement parameters of an activity performance [54].

T4 pointed to an order effect for the drop of interrater consistency. Rating all activities of one patient as a block (as in Session I) is easier that rating individual activities of different patients in mixed order (as in Session II). T4 also noted that *"mixing the order may still help generalize rating approaches"* and added that handling mixed order with greater consistency would *"require further clarification and training for rater."* The suggested requirement of additional training along with enhanced instructions was also repeated by several of the therapists in responses in multiple sections. All therapists concurred that greater exposure to rating offline using repeated views of videos (as compared to therapists rating in real time using only their observation) would improve the accuracy of the rating performance.

In the section on patient impairment profiles, the respondents all agreed that the participants in the mild to moderate impairment range (as opposed to only mild or only moderate) were the most challenging to evaluate, with T1 noting that *"the patients in the middle are always the hardest to judge because their movement patterns are likely more varied. In other words, a person may be considered moderately impaired overall, but will have elements of mild and severe movements intermixed."* These comments conform with patterns observed in health related ratings and expert ratings in other fields. "Easy" instances are the ones at the edges of a continuum (not impaired, very impaired) and thus further away from decision boundaries [27, 29, 33].

We asked the therapists if there was anything we might need to change in the rubric, or standardize in the assessment instructions to achieve greater rater agreement. T4 called for additional clarification in the rubric stating: "*Clarification may remove more of the subjective rating tendencies and give a clearer picture of what exactly is "abnormal" or "non-functional" about the movements. I think the issue is that abnormal and non-functional are not synonymous and individuals can use a wide array of movements to accomplish the same functional task.*

DISCUSSION

Our ultimate end goal is to create robust, scalable and affordable semi-automated systems to realize data supported rehabilitation in the home. Over several years, through extensive interdisciplinary collaboration and with much trial and error, we established a standardized set of human-centered processes for making as much of the expert therapist patient movement assessment process explicit, standardized and quantifiable. We embedded this codification within an interactive camera based system. Expert therapists are able to use this system for adaptive training of patients with different levels of impairment and rate the resulting videos of patients performing the exercises with good consistency. The codification and rating processes are also helping the participating therapists understand nuances about their own approach to therapy which hopefully will continue to improve their emerging ability to make standardized evidence-based decisions. Key outcomes and recommendations for other researchers engaged in the computer assisted evaluation of human movement include:

Constraining the observed movement space

Automating the capture and assessment of complex human movement requires careful constraint of the observed set of movement features. System designers need to engage in lengthy ethnographic and observational work within different movement contexts to assist in the generation of a limited set of artifacts, environments and activities that are extensible, standardized, and will support therapy at scale. Moving from subjective and intuitive observations towards a computable index is likely to be more successful with expert input leading to the development of an initial state-machine representation of the selected activities and their segment vocabulary.

Capturing movement data

Trade-offs are required when proposing a movement capture approach in terms of cost, reliability, accuracy, appeal, and usability. Technical infrastructure needs to be routinely powered/charged, and calibrated and positioned consistently, while interactive software needs to be reasonably intuitive to use and suitable for the target participant demographic (patients and therapists). Careful selection of captured participants is required to ensure diversity across a distribution of humans, with considerable coverage required for "those in the middle" who may present a more complex set of movements to assess.

Assessing movement data
A maximum of 3 – 4 assessment features (as in Table 3) per movement segment is recommended in order to facilitate standardized observation and training of experts in using movement rubrics consistently. Tightly coupling the movement features with a simple rating approach and providing standardized operational definitions for translating observation to rating can also add a strong interpretive understanding to each of the quantified scores. Instrumenting digital rating tools to best support in the correct implementation of the scoring rubric can potentially increase inter-rater reliability. It is recommended that the expert designers of the assessment approach also be the first users of the rating tool as they are most familiar with the intent of the measurement approach.

Iterative in-situ and online discussion
We recommend in-person expert-led workshops where discussion can involve participants enacting observed movements from captured videos and soliciting comments on their own performance. Within these workshops, the captured video artifacts can serve to focus analysis on the particularities of the captured individuals, as opposed to the memories of the movement experts. In addition, collaborative online surveys and video sessions can help continuously reveal insights from the assessed data back to the raters which may have a calibrating effect.

Mitigating potential issues
We observed that the therapists in our study did not always consistently remember to use the checklist of movement features as a way to document their rating interpretation. We have subsequently inverted the rating process in our current version of the annotation tool. In each segment, the therapist raters are given the list of movement features to focus on and must select, by checking a box, features they consider to be impaired in that segment. Their interpretation of movement function then automatically generates the score following the rules of the rubric (no checked movement features gives a rating of 3, one checked feature a 2 and two or more a 1). This tighter tool/rubric alignment should enable greater inter-rater reliability and consistency.

From an assessment perspective, we tried opening up the annotation system to 10 additional raters (with no experience of the project) and even with several hours of collective system training, the inter-rater correlation became even lower. In comparison, the two most experienced therapists in the rating studies (T2 and T3) provided the most internally consistent ratings that were also closest each time to the mode. One reason for this may be that their years of experience help make them cognitively more aware of their biases or signature style. Our approach moving forward is have two dedicated physical therapists working on the project one day a week to label videos throughout 2020.

The experience of developing our standardization process was lengthy, highly iterative, and time-consuming. For example, we made initial assumptions about the length of time it might take for the rehabilitation experts to agree on the movement rubric that became challenging to negotiate and manage after the initial in-situ workshops. This required us to speed up development of our online version of the tool in order to facilitate continued rating of videos while the ideas were still "fresh" in the minds of the raters. This led to considerable technical issues dealing with different operating systems, hospital firewalls, and the diversity of general technical knowledge and support expectations across the rating participants. To address this, we propose more extended training sessions using encrypted researcher controlled machines in the formative stages of data collection and annotation.

FUTURE WORK
We are installing our system in three major rehabilitation clinics in the United States with the goal of capturing patients performing the 15 activities described in this paper. The tool will also be adapted to capture patients performing the activities of a widely used clinical tool for the functional assessment of stroke survivors, the Action Research Arm Test (ARAT) [37]. Over 200 patients will be captured performing the two sets of activities and the videos will be rated by expert therapists across the three clinics. The captured data will then be used to train machine learning algorithms for the automated segmentation and assessment of the movement performance. The algorithms will support the automated assessment of the ARAT in the clinic and the semi-automated implementation of long-term therapy in the home. The availability of highly standardized and interpretable video and rating datasets will make feasible the development of machine learning algorithms even though the dataset is relatively small and noisy. This approach therefore can address a long term problem of applying machine learning to complex health contexts, where the capture and analysis of data is challenging, costly, and requires significant effort.

CONCLUSION
Achieving semi-automated rehabilitation or developing standardized assistive tools for movement capture and assessment requires a carefully measured process to ensure the full scope and intricacy of the problem space is revealed. We present a series of methodical steps demonstrating our approach within the context of upper extremity stroke rehabilitation. The strategic workflow of these steps can be generalized as follows: helping experts reveal, articulate, and standardize their process; creating tools to capture data and reflect on the process; and creating annotation assessment tools that assist in the efficacy and efficiency of the entire process. This methodical approach can potentially be used in many other situations that involve the capture and assessment of complex human performance.

ACKNOWLEDGMENTS
We thank the physiotherapists, the occupational therapists and the nine stroke survivors and their families for their gracious participation in our studies.

REFERENCES

[1] Gazihan Alankus, Rachel Proffitt, Caitlin Kelleher, and Jack Engsberg, *Stroke Therapy through Motion-Based Games: A Case Study.* ACM Trans. Access. Comput., 2011. **4**(1): p. 1-35.

[2] Craig Anderson, et al., *Home or hospital for stroke rehabilitation? results of a randomized controlled trial : I: health outcomes at 6 months.* Stroke, 2000. **31**(5): p. 1024-31.

[3] Craig Anderson, Cliona Ni Mhurchu, PM Brown, and K Carter, *Stroke rehabilitation services to accelerate hospital discharge and provide home-based care: an overview and cost analysis.* Pharmacoeconomics, 2002. **20**(8): p. 537-52.

[4] Madeline Balaam, et al., *Motivating mobility: designing for lived motivation in stroke rehabilitation*, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011, ACM: Vancouver, BC, Canada. p. 3073-3082.

[5] Michael Baran, et al., *Design of a home-based adaptive mixed reality rehabilitation system for stroke survivors.* Conf Proc IEEE Eng Med Biol Soc, 2011. **2011**: p. 7602-5.

[6] Michael Baran, et al., *Interdisciplinary concepts for design and implementation of mixed reality interactive neurorehabilitation systems for stroke.* Phys Ther, 2015. **95**(3): p. 449-60.

[7] John Beard, et al., *The World report on ageing and health: a policy framework for healthy ageing.* Lancet, 2016. **387**(10033): p. 2145-2154.

[8] Emeilia Benjamin, et al., *Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association.* Circulation, 2017. **135**(10): p. e146-e603.

[9] Yinpeng Chen, Michael Baran, Hari Sundaram, and Thanassis Rikakis, *A low cost, adaptive mixed reality system for home-based stroke rehabilitation.* Conf Proc IEEE Eng Med Biol Soc, 2011. **2011**: p. 1827-30.

[10] Yinpeng Chen, et al., *A Computational Framework for Quantitative Evaluation of Movement during Rehabilitation*, in *AIP Conferernce*. 2011. p. 317-326.

[11] Karen Chua and Christopher Kuah, *Innovating With Rehabilitation Technology in the Real World: Promises, Potentials, and Perspectives.* Am J Phys Med Rehabil, 2017. **96**(10 Suppl 1): p. S150-S156.

[12] Carmen Cirstea and Mindy Levin, *Improvement of arm movement patterns and endpoint control depends on type of feedback during practice in stroke survivors.* Neurorehabil Neural Repair, 2007. **21**(5): p. 398-411.

[13] Margaret Duff, et al., *An adaptive mixed reality training system for stroke rehabilitation.* IEEE Trans Neural Syst Rehabil Eng, 2010. **18**(5): p. 531-41.

[14] Margaret Duff, et al., *Adaptive mixed reality rehabilitation improves quality of reaching movements more than traditional reaching therapy following stroke.* Neurorehabil Neural Repair, 2013. **27**(4): p. 306-15.

[15] Pamela Duncan, et al., *Adherence to Postacute Rehabilitation Guidelines Is Associated With Functional Recovery in Stroke.* Stroke. **33**(1): p. 167-178.

[16] Aviv Elor, Mircea Teodorescu, and Sri Kurniawan, *Project Star Catcher: A Novel Immersive Virtual Reality Experience for Upper Limb Rehabilitation.* ACM Trans. Access. Comput., 2018. **11**(4): p. 1-25.

[17] Elisa Ferrara, Sonia Nardotto, Serena Ponte, and Silvana G. Dellepiane, *Infrastructure for data management and user centered rehabilitation in Rehab@Home project*, in *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*. 2014, ACM: Rhodes, Greece. p. 1-8.

[18] Axel Fugl-Meyer, Lisbeth Jaasko, Ingegerd Leyman, Sigyn Olsson, and Solveig Steglind, *The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance.* Scand J Rehabil Med, 1975. **7**(1): p. 13-31.

[19] Brian Hanley, Catherine Tucker, and Athanassios Bissas, *Differences between motion capture and video analysis systems in calculating knee angles in elite-standard race walking.* J Sports Sci., 2018. **36**(11): p. 1250 - 1255.

[20] Jocelyn Harris and Janice Eng, *Paretic upper-limb strength best explains arm activity in people with stroke. .* Phys Ther, 2007. **87**: p. 88-97.

[21] Samar Hatem, 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**: p. 442.

[22] Sarah Housman, Kelly Scott, and David Reinkensmeyer, *A randomized controlled trial of gravity-supported, computer-enhanced arm exercise for individuals with severe hemiparesis.* Neurorehabil Neural Repair, 2009. **23**(5): p. 505-14.

[23] Huey Huang, et al., *Novel design of interactive multimodal biofeedback system for neurorehabilitation.* Conf Proc IEEE Eng Med Biol Soc, 2006. **1**: p. 4925-8.

[24] Aisling Kelliher, Jinwoo Choi, Jia-Bin Huang, Thanassis Rikakis, and Kris Kitani, *HOMER: An Interactive System for Home Based Stroke Rehabilitation*, in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 2017, ACM: Baltimore, Maryland, USA. p. 379-380.

[25] Aisling Kelliher and Barbara Barry, *Designing Therapeutic Experiences with AI in Mind*, in *AAAI 2018 Symposium, Designing the User Experience of Artificial Intelligence*. 2018: Stanford, CA.

[26] Aisling Kelliher, Andrew Gibson, Eric Bottelsen, and Edward Coe, *Designing Modular Rehabilitation Objects for Interactive Therapy in the Home*, in *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 2019, ACM: Tempe, Arizona, USA. p. 251-257.

[27] Faisal Khan, Bilge Mutlu, and Zhu Jerry. *How Do Humans Teach: On Curriculum Learning and Teaching Dimension*. in *Advances in Neural Information Processing Systems 2011*. 2011.

[28] Silvia Koton, et al., *Stroke incidence and mortality trends in US communities, 1987 to 2011.* JAMA, 2014. **312**(3): p. 259-68.

[29] John Krakauer, *Arm function after stroke: from physiology to recovery.* Semin Neurol, 2005. **25**(4): p. 384-95.

[30] John Krakauer, *Motor learning: its relevance to stroke recovery and neurorehabilitation.* Curr Opin Neurol, 2006. **19**(1): p. 84-90.

[31] Klaus Krippendorff *Computing Krippendorff's alpha reliability.* 2007. **43**.

[32] Gert Kwakkel, Boudewijn Kollen, Jay van der Grond, and Arie Prevo, *Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke.* Stroke, 2003. **34**(9): p. 2181-6.

[33] Gert Kwakkel, Boudewijn Kollen, and Eline Lindeman, *Understanding the pattern of functional recovery after stroke: facts and theories.* Restor Neurol Neurosci, 2004. **22**(3-5): p. 281-99.

[34] Mikko Kytö, Laura Maye, and David McGookin, *Using Both Hands: Tangibles for Stroke Rehabilitation in the Home*, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, ACM: Glasgow, Scotland Uk. p. 1-14.

[35] N. Lehrer, S. Attygalle, S. L. Wolf, and T. Rikakis, *Exploring the bases for a mixed reality stroke rehabilitation system, part I: a unified approach for representing action, quantitative evaluation, and interactive feedback.* J Neuroeng Rehabil, 2011. **8**: p. 51.

[36] Mindy Levin, Jeff Kleim, and Steven Wolf, *What do motor recovery and compensation mean in patients following stroke?* Neurorehabilitation and neural repair, 2008.

[37] Michelle McDonnell, *Action research arm test.* Aust J Physiother, 2008. **54**(3): p. 220.

[38] Lewis Morgenstern, et al., *Persistent ischemic stroke disparities despite declining incidence in Mexican Americans.* Ann Neurol, 2013. **74**(6): p. 778-85.

[39] A. Nordin, M. Alt Murphy, and A. Danielsson, *Intra-rater and inter-rater reliability at the item level of the Action Research Arm Test for patients with stroke.* J Rehabil Med, 2014. **46**(8): p. 738-45.

[40] Donald A. Norman, *The design of everyday things*. 1st Doubleday/Currency ed. 1990, New York: Doubleday. xv, 257 p.

[41] Cristina Ramírez-Fernández, Alberto L. Morán, Eloísa García-Canseco, and Felipe Orihuela-Espina, *Design Factors of Virtual Environments for Upper Limb Motor Rehabilitation of Stroke Patients*, in *Proceedings of the 5th Mexican Conference on Human-Computer Interaction*. 2014, ACM: Oaxaca, Mexico, Mexico. p. 22-25.

[42] D. J. Reinkensmeyer, et al., *Computational neurorehabilitation: modeling plasticity and learning to predict recovery.* J Neuroeng Rehabil, 2016. **13**(1): p. 42.

[43] David Reinkensmeyer and Michael Boninger, *Technologies and combination therapies for enhancing movement training for people with a disability.* J Neuroeng Rehabil, 2012. **9**: p. 17.

[44] Thanassis Rikakis, et al., *Semi-automated home-based therapy for the upper extremity of stroke survivors*, in *ACM PErvasive Technologies Related to Assistive Environments (PETRA '18)*. 2018: Corfu.

[45] Thanassis Rikakis, et al., *Semi-automated home-based therapy for the upper extremity of stroke survivors*, in *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. 2018, ACM: Corfu, Greece. p. 249-256.

[46] A. Schwarz, C. M. Kanzler, O. Lambercy, A. R. Luft, and J. M. Veerbeek, *Systematic Review on Kinematic Assessments of Upper Limb Movements After Stroke.* Stroke, 2019. **50**(3): p. 718-727.

[47] D. Son, et al., *Multifunctional wearable devices for diagnosis and therapy of movement disorders.* Nat Nanotechnol, 2014. **9**(5): p. 397-404.

[48] Standifird TW, et al., *Influence of Total Knee Arthroplasty on Gait Mechanics of the Replaced and Non-Replaced Limb During Stair Negotiation.* J Arthroplasty, 2016. **31**(1): p. 278-283.

[49] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang, *Physio@Home: Exploring Visual Guidance and Feedback Techniques for Physiotherapy Exercises*, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015, ACM: Seoul, Republic of Korea. p. 4123-4132.

[50] Vinay Venkataraman, et al., *Decision support for stroke rehabilitation therapy via describable attribute-based decision trees.* Conf Proc IEEE Eng Med Biol Soc, 2014. **2014**: p. 3154-9.

[51] David Wade, R. Langton-Hewer, V. A. Wood, Clive Skilbeck, and H. M. Ismail, *The hemiplegic arm after stroke: measurement and recovery.* J Neurol Neurosurg Psychiatry, 1983. **46**(6): p. 521-4.

[52] David Webster and Ozkan Celik, *Systematic review of Kinect applications in elderly care and stroke rehabilitation.* J Neuroeng Rehabil, 2014. **11**: p. 108.

[53] Steven Wolf, et al., *Assessing wolf motor function test as outcome measure for research in patients after stroke.* . Stroke, 2001. **32**(7): p. 1635-1639.

[54] Steven Wolf, et al., *Effect of constraint- induced movement therapy on upper extremity function 3 to 9 months after stroke: the excite randomized clinical trial. J.* Jama, 2006. **296**(17): p. 2095-2104.

[55] Ching-Yi Wu, Catherine Trombly, Keh-Chung Lin, and Linda Tickle-Degnen, *Effects of object affordances on reaching performance in persons with and without cerebrovascular accident.* Am J Occup Ther, 1998. **52**(6): p. 447-56.

[56] Y. Yoshida, J. Zeni, and L. Snyder-Mackler, *Do Patients Achieve Normal Gait Patterns 3 Years After Total Knee Arthroplasty?* Journal of orthopaedic & sports physical therapy, 2012. **42**(12): p. 1039-1049.