Machine Perception for Occupational Therapy

Sonya J. Allin

CMU-HCII-07-103
August 2007

Human Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Christopher G. Atkeson, Chair
Nancy Pollard
Jie Yang
Nancy Baker, University of Pittsburgh

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This work was supported by in part by The NSF IGERT on Assistive Technology at the University of Pittsburgh and Carnegie Mellon University and the NSF Graduate Research Fellowship Program.

Keywords: Machine Perception, Rehabilitation Engineering, Stroke Rehabilitation, Applied Computer Vision

Abstract

In this thesis, we propose a novel measurement methodology for long term monitoring and assessment of stroke survivors' functional motion. Our long-term goal is to measure and provide feedback about changes in functional movements that happen in everyday environments across days, months or years.

The research contained in the thesis consists of three parts. First, therapists were interviewed to determine common observations they make about stroke survivors functional activity for the purpose of assessment. Protocol analysis was used to gain insight onto parts of the body as well as motor symptoms that were particular points of focus.

Next, a prototype system was built to measure functional movement and its ability to capture statistics that discriminate between levels of functional impairment was tested. The technologies explored were primarily chosen because they are inexpensive and robust. They include simple computer vision and force sensing devices. Several different score prediction paradigms were tested, including some that are task specific and others that score movements, like lifts or grasps, across many tasks. We illustrate the accuracy of each paradigm, and discuss several of the underlying statistics found to correlate strongly and consistently with functional score. These statistics included the measured variance about the elbow and motion of the torso.

Finally, we present initial results from an effort to bring low cost and automated assessment technologies to the home. Cameras and force sensing devices were installed in the home of a single stroke survivor, and desktop activity was measured in this environment over the course of two weeks.

We believe that the technologies we highlight in this thesis have the capacity to inexpensively enhance the quality and character of therapy that stroke survivors receive after hospital discharge. This is primarily because they can focus therapy on real world tasks that take place in the home. We expect the tools we prototype here ultimately to enable: 1. Quantification of long term changes in functional mobility; 2. Evaluation of interventions that relate to functioning; 3. Motivating feedback about real world movement quality.

Acknowledgments

Emily Eckel, Bambi Brewer, Chris Atkeson, Deva Ramanan and Elizabeth Skidmore have all provided ideas and comments that significantly shaped my research ideas and direction. Laura Dabbish and Frank Broz have been there to read drafts of my papers, give me comments, attend my (often unfinished, stuttering) practice talks. Garth Zeglin taught me to use resistors and a bread board. Carolyn Rose and Bilge Mutlu helped me understand protocol analysis. Carlos Guestrin introduced me to ridge regression and Larry Wasserman helped me understand power analysis. Sidd Srinivasan watched my practice talks and gave me pep talks when required. Nancy Baker, Nancy Pollard, and Jie Yang all gave me ideas, encouragement, and comments on this manuscript. Thank you so much!

Contents

Introduction

Not long ago, most doctors thought recovery from stroke spanned weeks or months after an injury. Recent research, however, has shown that significant motor gains can take place several months, and even years after the acute phase of care and long after individuals have been discharged from a hospital [78, 134]. This means that the period where rehabilitation may be useful is much longer that was originally thought possible. As a result, research has placed a stronger emphasis on rehabilitation paradigms for chronic stroke survivors and those who are no longer in hospitals.

Research in neuroscience, occupational and physical therapy stresses several therapeutic principles that correlate strongly with positive long term rehabilitation outcomes after stroke. Some of these principles include: frequent and repetitive practice of movement with body parts most affected by stroke [64]; practice of motions that are functionally meaningful [79, 85]; and practice of motions that are personally meaningful to the individual [101].

In the United States the therapy administered to facilitate recovery, however, rarely addresses these principles overall. This is largely due to restrictions imposed by insurance providers. Medicare, for example, has routinely enforced caps on the amount of outpatient therapy it will reimburse and a moratorium on caps is set to expire [4]. This means most people in this country do not receive the duration or intensity of monitored practice that research suggests to be worthwhile. In addition, monitored therapy rarely takes place in the home or involves the same objects that clients regularly use. Medicare currently will only fund home visits by therapists for those who are physically unable to leave the house [3].

In this thesis, we propose a method to augment long term monitoring and assessment of stroke survivors' functional activity in the home. To do this, we cheaply and robustly quantify functional motion of the upper extremities of stroke survivors in ways that can both discriminate between levels of disability and detect changes in mobility over time. Moreover, we do this in a fashion that is consistent with existing functional assessments. Our ultimate goal is to measure and provide feedback about changes in real functional movements that happen in everyday environments. We want to do this over long periods of time, such as months or years.

The technologies we explore in the these were primarily chosen because they are cheap and robust. These include simple computer vision and force sensing devices. As a part of the work, we built a prototype system to measure functional movement and tested its ability to capture statistics that discriminate between levels of functional impairment. We also deployed our system to an individual's home and explored issues surrounding functional measurement in this environment.

This thesis makes contributions to different but potentially overlapping fields of research. On a technical level, the work represents both a novel application area and set of evaluation metrics for vision based kinematic trackers. It also expands upon the study of functional force perception, which is changing as the capacity to measure force improves. For the clinical community, the work illustrates an approach to assessment that is appropriate for homes or workplaces, and which can theoretically operate automatically, over long stretches of time.

We believe that the technologies we highlight hold the capacity to inexpensively enhance the intensity of therapy that chronic stroke survivors currently receive. They can additionally focus therapy on real world tasks in the home. We expect, then, the tools prototyped here ultimately to enable: 1. Quantification of long term changes in real functional mobility; 2. Evaluation of interventions' that relate to real world functioning; 3. Motivating feedback about real world movement quality.

Organization of the Thesis

The organization of this document is as follows:

In Chapter 1, we present clinical and technical background. We describe a range of movement features that are typical after stroke and present assessments that are popularly used to quantify impairment. We also describe technologies currently employed in stroke rehabilitation as well as technologies most appropriate for inexpensive, home-based therapy.

In Chapter 2, we describe the system we developed for observation and assessment of functional movement. We describe a range of features the tool is capable of extracting both from visual and force data.

In Chapter 3, we present results from a protocol analysis study with therapists. This study was designed to identify visually perceptible features that therapists regularly use to make functional assessments of their clients.

In Chapter 4, we attempt to automate functional assessments with our engineered devices. Our goal is to automatically discriminate between levels of disability as accurately as a human expert. We also examine features our automated system prioritizes in order to make assessments, and compare these features to those prioritized by humans in Chapter 3.

In Chapter 5, we deploy our engineered devices to a single stroke survivors' home. We explore the issues surrounding this deployment and outline future work that will make use of the data that was acquired in this environment.

In Conclusion, we discuss remaining obstacles that must be tackled before a feedback system can be developed, and outline related projects for the future.

Chapter 1

Background Work

Figure 1.1 illustrates the kind of measurement and assessment system we are looking to enable through our research. At the left is a stroke survivor at home unloading his dishwasher. Invisibly situated in the environment are low cost cameras and force sensing devices that measure his activity. In this particular example, the character and quality of grasps are being evaluated; each time the stroke survivor grasps an object in the dishwasher, movement data corresponding to his grasp is being extracted and a numeric assessment of this data is being automatically made. The research described in this thesis seeks to enable the development of systems, like the one depicted in Figure 1.1, that evaluate the quality of real world lifts and grasps of objects in a fashion that is consistent with existing clinical assessments. Like existing clinical assessments, we seek to make evaluations that are sensitive to the variation in monitored performance over days or weeks, and that can discriminate any one person's level of disability from that of others who have been impaired by stroke.

Automated assessments at home hold the capacity to enable retrospective review of performance data with therapists. They can also allow therapists to intervene should a stroke survivor's ability to perform functional tasks at home deteriorate, or to alter the character of practice that takes place at home so as to better encourage recovery. In short, they can augment the quality of rehabilitation be measuring performance outcomes as they relate directly to real world perfor-

Figure 1.1: Idealized assessment system in the home of a client. Our goal is to create an automated system that invisibly measures the quality of real world functioning. Here, a system evaluates functional grasps of a stroke survivor as he unloads the dishwasher. Every grasp is assigned a “motor quality” score that is validated against other clinical assessments, like the FMA or AMAT. The recordings enable retrospective review of functional performance changes that take place in the real world.

mance.

In this chapter we describe the literature that has informed our progress toward the automated assessment of functional movement. The work comes from several different fields of research: occupational therapy; rehabilitation engineering and science; computer vision and human computer interaction. In our review of the literature, we focus on rehabilitation of the upper body, in particular on the hands and arms. We are particularly concerned, moreover, with assessment as it relates to functional object manipulations, of the variety that commonly take place in the home.

We first review clinical literature that describes common characteristics of upper extremity movement after stroke as well as guidelines for effective therapy. Therapy guidelines, of course, are constantly being revised as research matures and new evidence is generated; what we present here, then, serves more as a snapshot of the state of the art than a definitive guide. We also look at various assessment paradigms for the upper extremity that are commonly employed during rehabilitation after stroke.

We then draw upon literature in computer vision, rehabilitation engineering and human computer interaction to describe ways technology can fill gaps in care. We place a particular emphasis on force measurement and computer vision devices and algorithms. We look at advances in kinematic tracking, for example, which promise soon to bring robust and clinically meaningful measurements of the body to a wide variety of environments. We also explore rehabilitation specific applications in force sensing and tracking, and the other force measurement paradigms that have inspired our work.

All of the technical and clinical work ultimately inform the development of the prototype system that is illustrated in Figure 1.2. Our prototype tracks the upper bodies of stroke survivors in three dimensions as they grasp and lift various functional objects. Force applied to objects is also measured during several functional tasks. Based on measurements, functional assessments of both “grasps” and “lifts” are automatically produced. These machine-generated scores are consistent with assessments made by human experts but, unlike the human scores, the machine-generate scores can be made continuously, for long periods of time, like weeks or months.

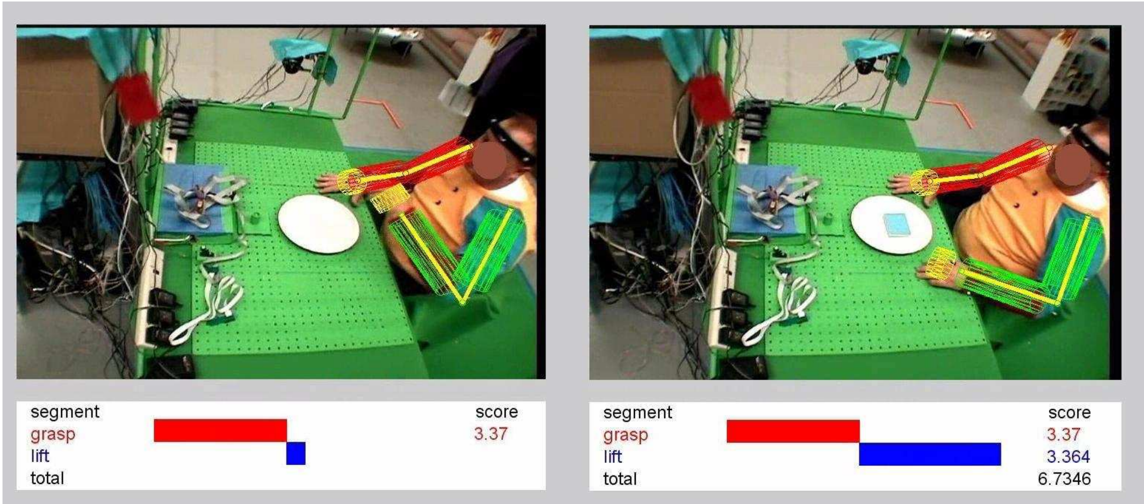


Figure 1.2: Current implementation of an automated assessment system. Our prototype tracks the upper bodies of stroke survivors in three dimensions as they grasp and lift various functional objects. Force measurements are made during several functional tasks, as well. Based on measurements, the quality of perceived 'grasps" and "lifts" is automatically evaluated.

1.1 Clinical Background

1.1.1 Common Symptoms after Stroke

The variety of sequelae that persist after stroke is large and depends upon features such as the nature of the physical injury and the age of the injured individual [56]. Despite this variation, stroke survivors are evaluated using fairly consistent criteria when determining motor progress for medical and insurance purposes. In this section, we make an effort to explain relatively common sequelae that are taken into consideration during assessments. Many have informed the development of instruments like the Fugl-Meyer Assessment [42], the Wolf Motor Functioning Test [147], or the Arm Motor Ability Test [62].

A. Hemiparesis. Six months after a stroke, more than 50% of stroke survivors will exhibit one-sided partial paralysis on the side of the body that is contra-lateral to a brain lesion; this is called *hemiparesis* [2]. Immediately after a stroke, many individuals exhibit no voluntary movement at all on the affected side. A complete lack of movement is called *flaccidity*. Weakness may be present even as recovery takes place; studies have found persistent upper limb weakness to affect the majority of stroke survivors [69].

Persistent weakness influences the maximum forces that stroke survivors are able to produce. In prior research, individuals with hand impairments were found to be significantly less likely to be able to produce the torque required to achieve functional goals, like turning a key in a lock [123]. In [75], stroke survivors' maximum finger forces were found to be reduced relative to controls by about 36%.

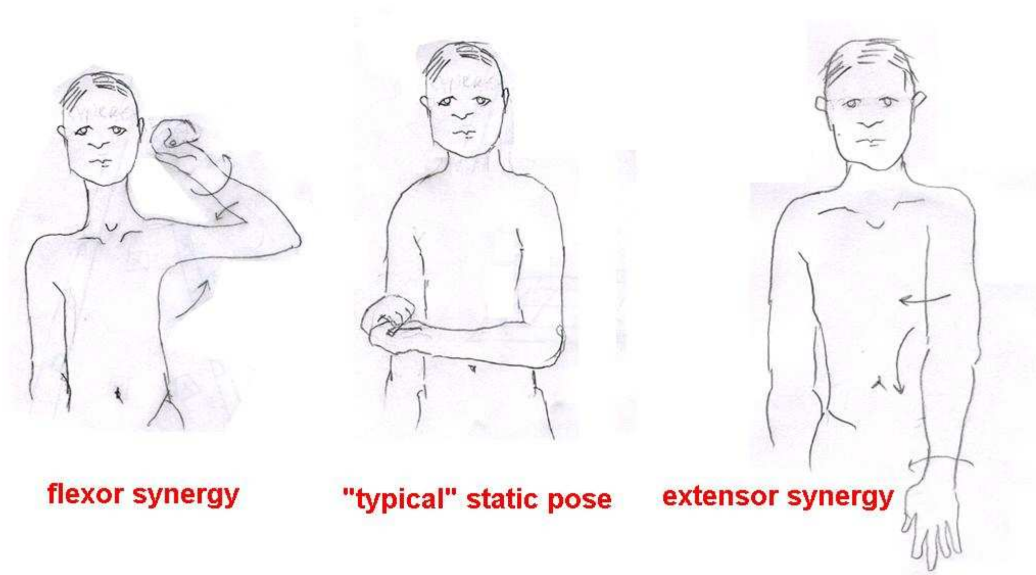


Figure 1.3: Flexor and extensor synergies.

B. Limb Synergies. Stroke survivors often exhibit stereotyped movement defined by patterned muscle tightness and restricted range of motion during volitional movement. These patterns of movement are termed *muscle synergies* and may be mediated in part through a reflex system [91]. Two typical muscle synergies that influence upper body motion in stroke survivors are the flexor and extensor synergy. The flexor synergy couples external rotation and abduction at the shoulder with elbow flexion and wrist supination. The extensor synergy couples internal rotation and adduction of the shoulder with elbow extension and wrist pronation. Some combination of flexor and extensor synergy is combined to form the canonical hemiparetic posture described by [20]. See Figure 1.3 for an illustration of these synergy patterns.

In 1970, Signe Brunnstrom created a mechanism to organize stroke survivors into categories that are largely defined by limb synergies [20]. These categories are known as the *Brunnstrom Stages* and were developed alongside protocols for interventions that relate to each stage. There is significant debate as to whether individuals pass through Brunnstrom Stages in a predictable manner as they recover from stroke, and whether the treatment options promoted by Brunnstrom are indeed optimal or advisable [144]. Nevertheless, her categorizations are useful to describe a wide range of post-stroke motor pathology and they have shaped research in occupational and physical therapy.

In the low Brunnstrom Stages, individuals remain largely flaccid on their most affected side and are unable to produce movement on their own. In the middle stages, voluntarily movement returns, yet it is dominated by synergy. If, for example, an individual attempts to flex at the elbow, he or she may also supinate the wrist without intention. Finally, at the highest Brunnstrom Stage, individuals can freely move about each joint independently. Motion about the elbow is not coupled with motion about the wrist. Motion of each finger can be solicited independently, without resulting in massed finger flexion or massed extension.

Patterns of constraint force consistent with the flexor and extensor synergies have been measured in the point to point motion of hemiparetic individuals, particularly as they are executed in a horizontal plane [108, 109]. Some combination of flexor and extensor synergies in stroke survivors has also been shown by EMG studies, albeit influenced by weakness [83].

An assessment designed to measure the impact of synergies on voluntary movement is called the Fugl-Meyer Assessment (FMA). The FMA scores individuals' ability to move in agreement with and outside of coordinated patterns that are consistent with synergies. Scoring takes the relative independence of joints into account [42].

C. Spasticity and Clonus. Coupled with the presence of muscle synergies, roughly 30-40% of stroke survivors [69] have velocity-dependent increases in their stretch reflexes, which is called *spasticity* [68]. This may manifest as hypertonicity (increased muscle tone), coupled with clonus (involuntary sets of rapid muscle contractions), exaggerated tendon reflexes, muscle spasms, or fixed joints.

Spasticity and synergy both contribute to the jerky upper body motion that is commonly measured in stroke survivors [63], as well as in disturbed coordination around joints [73]. In addition, spasticity and synergy contribute to lack of force control. Prior studies have shown a dominance of forces produced by the flexors among stroke survivors, which increases as affected muscles are used [23]. An inability to control the extent of flexion also contributes to excess grip force commonly measured among stroke survivors during functional motions [48]. Such grasping impairments are particularly significant in that grasping impairment is a major determinant of stroke survivors' capacity to achieve functional goals [122].

Spasticity and clonus also depend on the environment and task constraints. In a study of spinal cord injured patients, for example, spasticity in the lower limbs became noticeably worse at night [81]. Heat can also reduce the presence of spasticity and spasms, at least temporarily [22].

D. Sensory Disturbances. Even with some volitional motion intact, roughly 30% of stroke survivors still suffer some form of sensory impairment [69]. This sensory impairment may affect the modulation of grip forces during manipulation tasks. Healthy individuals have been seen to modulate their grip forces as the acceleration of the objects they are holding increases or decreases. Stroke survivors may be comparatively slow in modulating force during tasks where objects are moved and they may fail to correctly estimate forces required to keep an object statically elevated [96]. Disturbed timing of force corrections has also been found to differ between periods where exerted forces are decreasing rather than increasing [65]. The commonly perceived delay in the onset of muscle contractions during manipulations after a stroke [137] may be the result of impairment to descending motor pathways in addition to ascending ones.

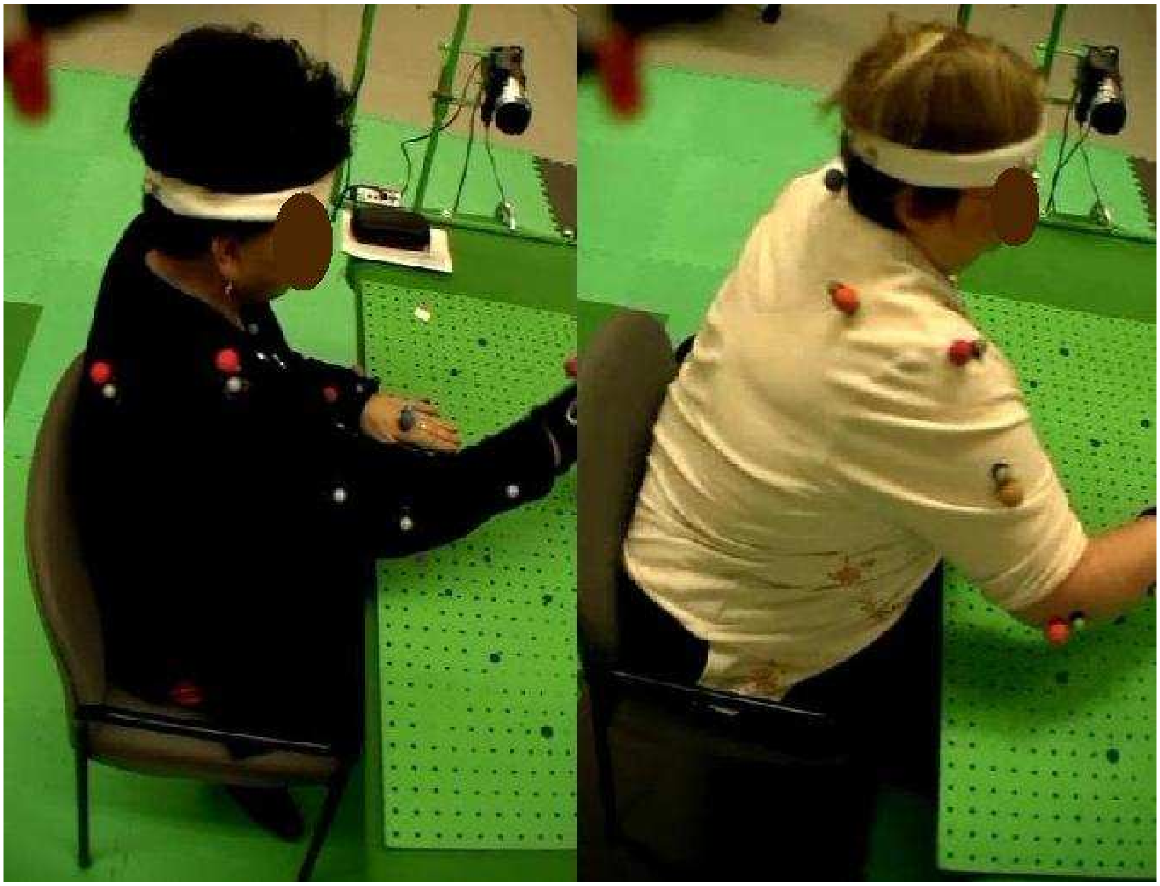


Figure 1.4: Torso compensation during reaches. At the right is an individual who never had a stroke reaching for an object. At the left is a stroke survivor reaching for an object at the same location. The stroke survivor uses much more motion of the torso to complete the reach.

E. Compensatory Strategies. Regardless of the Brunnstrom stage that represents an individual stroke survivor's movement ability, stroke survivors compensate for their deficits. As Bernstein first noted in 1967 [15], the human body has numerous degrees of freedom; any particular functional task can be achieved by several different combinations of motion about joints. Stroke survivors often use these excess degrees of freedom to compensate for their disabilities. They may, for example, recruit more proximal degrees of freedom when distal extremities are weak or otherwise impaired. In the upper body, hemiparetic individuals have been found to lean excessively into a reach so as compensate for an impairment at the elbow or shoulder [111], as is illustrated in Figure 1.4. Additionally, stroke survivors have been found to modify the orientation of their hands and wrists to compensate for weak or impaired extensors [13, 111, 112].

Opportunities to compensate are offered to stroke survivors not only by extra degrees of freedom in the body but also by the environments in which they operate. For example, stroke survivors may rely on the table to support the weight of their arms during reaches or they may use passive object properties, like friction, to maintain a grasp [112]. These kinds of grasps and manipulations may be fully functional. In [111], for example, those disabled by stroke were able to transport impaired hands to the same physical locations as their unimpaired counterparts. The impact of the environment on functioning is one reason that we stress measurement in the environments in which people actually live and work.

An assessment designed to measure the presence of compensatory trunk motion is called the Reaching Performance Scale (RPS) [74]. Preliminary trials of therapy in which the torsos of stroke survivors were restrained during therapy have also been shown to positively influence stroke survivors' coordination and range of motion [90].

There are, of course, several other ways in which stroke survivors can be impaired. An estimated 23% suffer *hemi-neglect* [69], for example, which is a disruption or blockage of the visio-spatial field on the side of the body that is most impaired by stroke. The symptoms above, however, provide a rough picture of several common physical impairments that influence gross upper body motion; it is admittedly incomplete.

1.1.2 Guidelines for Effective Therapy

Not very long ago, most doctors thought recovery from stroke could span only several weeks or months after an injury. Recent research, however, has shown that significant motor gains after a stroke can take place several months, and even years after the acute phase of care, and long after individuals have been discharged from a hospital [36, 38, 92, 133].

An increased awareness of the duration of recovery has led to the development of a variety of new therapies and interventions that target chronically impaired individuals. One example is found in the Constraint Induced Movement Therapy (CIMT) movement, which has been widely adopted in clinics. CIMT requires the restraint of clients' more functional, or mobile, side of the body, to ensure that the more impaired side of the body sees regular use. CIMT has been found to yield functional motor improvements as well as associated neural re-mappings as many as 12 years after a stroke [135].

Along with these new interventions has come a set of evidence-based guidelines for effective post stroke therapy. Evidence, of course, keeps mounting and guidelines keep being revised. Some guidelines as they currently stand follow.

A. To the degree possible, therapy should emphasize use of body parts impaired by stroke. Many stroke survivors have a tendency to rely on the parts of their body that are least damaged by a stroke. This tendency has been termed "learned nonuse" [130, 131] and is commonly seen as a barrier to effective therapy, despite the fact that it may produce functional motion. It has been estimated that 25-30% of stroke survivors exhibit learned nonuse [130]. Lack of limb use, in fact, does result in atrophy, both in terms of the plasticity of muscles and connective tissues as well as in terms of neural representation. Shrinking neural representations associated with limb movements have been evidenced by primate studies wherein limbs impaired by stroke were forcibly restrained. Increased use of an impaired limb, by contrast, can help maintain or expand the number of neurons whose activity is associated with that limb. This expansion, moreover, has been seen to take place not only days or weeks after a stroke but months or years after injury [76, 78]. Other gains as a result of the CIMT strategy include improved quality of motion and better long-term functioning vital to independent living [134].

B. Intense, repetitive practice of movements yields motor improvements. Research in occupational and physical therapy stresses repeated, intense practice of motion during therapy. In a research synthesis by Kwakkel et al. [67], for example, statistically significant improvements in ADL performances and in neuromuscular and functional outcome variables were associated with longer rehabilitation duration and intensity. The robotic rehabilitation community has contributed a substantial amount of additional evidence to support the use of intense and repetitive practice during therapy. Studies of clients who have made use of the MIT Manus [49], for example, have established that intense, repetitive practice of reaching motions can yield motor improvements relating to movement smoothness and speed [64].

An important feature of results produced by the MIT Manus is that they include not only those with moderate impairments but also severely impaired chronic stroke patients as well. Moderately and severely impaired patients were both found to benefit from a massed-practice therapy paradigm with intensive, frequent, and repetitive treatment of isolated motions [86].

C. Practice of functional tasks is preferable to practice where tasks are simulated. It has long been understood that task specific constraints inform the character of movements that individuals use. Research in occupational therapy, for example, has shown motor impaired individuals to use decidedly different movement strategies in contexts where they are pantomiming functional tasks than in environments where they actual execute tasks [55, 85]. Wu has demonstrated that motor impaired individuals will reach more smoothly to a real coin than to a non-coin [149]. A meta-analysis of studies comparing exercise programs in which functional tasks were practiced versus programs containing rote or elemental exercises showed functional activities to yield better quality of movement [79]. Several other studies have stressed the importance of supporting functional activities in therapy environments in order to achieve optimal recovery [77, 97, 135, 146].

If individuals train to tasks that do not emulate real world functioning, they may compromise the transfer of motor learning to real world tasks. In some early studies of reach retraining, for example, subjects were found to exhibit motor gains on tests resembling their training, but not in the context of functional assessments [143]. Therapists are therefore increasingly emphasizing the task specificity of re-training after stroke. Research in Constraint-Induced Movement Therapy, for example, and has shown that intense and task specific training can positively influence functional performance and associated cortical reorganization [133]. The potential effect of functional task practice on brain plasticity has been similarly stressed by work in [10, 11, 77, 97, 135, 146]. The use of natural objects and environments during rehabilitation is also recommended by occupational therapy texts [136]. Research relating to cerebral palsy has shown that learning in real task environments may prove most meaningful to those whose sensory systems are impaired; those least able to use the intrinsic feedback from their bodies may most depend on the environment to gauge performance success [142].

D. Practice at home may be cheaper than practice in the clinic, and at least as effective. The environments that have the most functional importance to people after a stroke are those in which they actually work and live. Moreover, rehabilitation at home may be substantially less expensive that rehabilitation in clinics. In a study of 110 stroke survivors in Sweden, for example, a home rehabilitation paradigm was shown to yield equivalent patient outcomes as traditional rehabilitation in terms of total motor capacity, manual dexterity, walking, self-reported independence in the activities of daily living, frequency of activities, and health-related quality of life [88]. Similar results were reported by the Domino Study Group [45], which followed 327 stroke survivors as they participated in either traditional or home rehabilitation. In the Domino study, the treatment of the home-based subjects was found to be significantly cheaper than the treatment of hospital-based counterparts. Moreover, these home based rehabilitated individuals had better household and leisure abilities at six months than those treated in clinics [44]. Similar results were found in a study of 86 stroke survivors in South Australia [8], although the potential negative effect of home rehabilitation on caregivers in the home was noted.

The home is also known to be an environment that is comfortable, stimulating and familiar to stroke survivors. In a study of stroke professionals' attitudes, 11 of 32 professionals felt a stimulating environment to be critical to patient motivation, and all said that the clinic did not provide this stimulating environment [87]. The home, then, may provide an attractive, motivating and cost effective alternative to traditional clinical environments for many therapy clients.

E. Practice in the context of motivating activity is vital. Keeping clients motivated through the repetitive practice of almost anything can be challenging. Several studies, however, have indicated that when clients are allowed to set their own goals for rehabilitation, participation in and compliance with therapeutic interventions increase [101]. In a recent survey of 32 therapists, more than half indicated they felt setting relevant goals to be a key to keeping clients of therapy motivated [87]. The Canadian Occupational Performance Measure (COPM) is an assessment that emphasizes individual client goals. It assesses clients' perception of motor progress relative to tasks that are most important to their quality of life. Its efficacy and validity, as well as its effect on motivation, has now been tested for a wide range of client groups, including stroke survivors [12].

Based on these guidelines, we focus in this thesis on assessment tools that are flexible, can be installed in real functional environments, are inexpensive, and can facilitate repetitive practice of a variety of tasks over the long term.

1.1.3 Guidelines for Assessment

Whatever the character of therapy an individual receives after a stroke, determining the effect of this therapy is crucially important. A positive outcome from therapy, however, can mean many things. It can mean people are able to independently perform tasks of daily living, or move smoothly and without hesitation. Assessments in physical and occupational therapy reflect these different interpretations of positive outcomes. Insurance companies tend to pay heed to assessments that are functional in nature, such as the Functional Independence Measure (FIM) or The Barthel Index of Activities of Daily Living (BI) [26, 80]. Assessments like these inform decisions to sponsor supportive care. The BI and the FIM are by far the most commonly used measures of functional disability [115], yet neither is sensitive to the particular motor strategy an individual uses to complete a task; rather, they measure the degree to which functioning requires assistance.

Research suggests that emphasizing function over underlying movement quality, however, can lead to the development of poor movement strategies that, while successful, may impede further recovery [9] and promote ”learned non-use”.

To make detailed measurements that relate to the underlying quality of motion, many researchers make use of the Fugl-Meyer Assessment (FMA) [115]. The FMA was designed to measure the effect of synergies on voluntary movement; it scores ability to move within and without coordinated patterns that are consistent with synergistic couplings. Scoring takes the relative independence of joints into account [42]. Emphasizing outcomes that do not relate to real world functioning, however, has been shown to be as problematic as training to increase functional outcomes alone. This is because training outside of functional activity is not guaranteed to result in functional gains [143].

Increasingly, assessments have come to emphasize quality of motion evaluations in the context of motions that have functional benefit. Examples of such hybrid assessments include the Wolf Motor Function Test (WMFT) [147], the Canadian Occupational Performance Measure (COPM) [12], and the Arm Motor Ability Test (AMAT) [62]. The AMAT is an assessment that inspires and informs much of the dissertation research reported here, so we explain it in some detail.

The Arm Motor Ability Test (AMAT) weighs functional capacity of the upper limbs against motor quality and speed. The test has been shown to have high inter-rater reliability, sensitivity to change, and concurrent validity with other leading assessments, including the WMFT [62]. As a result of the assessment, observed functional performance of stroke survivors is mapped onto two scales. One scale evaluates functional capacity and the other evaluates quality of underlying motion. Each task on the assessment is sub-divided into a set of components such as lifting a sandwich to the lips or dialing a number on the telephone. Individual ratings relate to these subcomponents and are summed to create an aggregate score. For the Quality of Movement scale,

The Arm Motor Ability Test (AMAT) Task/Subtask	OK to Delete?
Cut “Meat”	No
1. Pick up knife and fork 2. Cut “meat” (Play-Doh) 3. Fork to mouth	
Foam “Sandwich”	No
4. Pick up foam “sandwich” 5. “Sandwich” to mouth	
Eat With Spoon	No
6. Pick up spoon 7. Pick up dried kidney bean with spoon 8. Spoon to mouth	
Drink From Mug	Yes
9. Grasp mug handle 10. Mug to mouth	
Comb Hair	No
11. Pick up comb 12. Comb hair	
Open Jar	
13. Grasp jar top 14. Screw jar top open	
Tie Shoelace	No
15. Tie shoelace	
Use Telephone	
16. Phone received to ear 17. Press phone number	
Wipe Up Spilled Water (7 ml)	Yes
18. Wipe up water (six movements) 19. Discard towel in wastebasket	
Put on Cardigan (Jacket-Style) Sweater	No
20. Affected arm in sleeve, sweater over shoulder 21. Button two lower buttons	
Put On T-Shirt	No
22. Arms in T-shirt sleeves 23. Head through neck hole 24. Pull down and straighten shirt	
Prop On Extended Arm	Yes
25.Prop on extended affected arm, reach across body with unaffected arm, pick up small object	
Light Switch/Door	No
26. Pincer grasp of light switch and flip down 27. Grasp door handle, rotate handle, open door 28. Close door	

Table 1.1: Tasks on the AMAT. Those that [62] finds to be redundant are indicated.

the ratings are: 0 = no use, 1 = very poor, 2 = poor, 3 = fair, 4 = almost normal, and 5 = normal. The complete list of tasks and subtasks on the AMAT is given in Table 1.1.

To generate scores, the AMAT manual requests a clinician to pay attention to features like "dexterity of paretic fingers, the extent to which the head moves downward when the sandwich is brought to the mouth ... and the fluidity and precision with which movements are performed." [132]. "Fluidity of motion" may be influenced by conditions such as excess spasticity, weakness and pathological synergy. "Dexterity of the fingers" may reflect observations of grasping impairment, while head movement "downward" may reflect observation of a compensatory motion.

Prior studies have indicated a certain degree of redundancy in the Arm Motor Ability Test. Both the functional scale and the quality of motion scale, for example, were found to be highly correlated in practice [62]. In the same study, several tasks and subcomponents were found to relate poorly to aggregate scores. The authors therefore suggest removing one of the two scales, and deleting specific tasks. In our work, we have chosen to follow this advice, so we make use of a single scale and reduced task list. We also drop tasks we feel are not amenable to measurement with the devices we have engineered. Tasks that are dropped as per the suggestion of [62] are indicated in Table 1.1. We also dropped other tasks that were not confined to a desktop.

Whatever assessment that is used, the goal of the assessment remains primarily the same. An assessment must be sensitive to variation in disability and should be able to measure change. It should be repeatable, consistent, and relatively quick to administer. Enforcing these properties, however, may not be easy. This is partly because stroke survivors exhibit day-to-day biological variability in their performance; spasticity, for example, depends on temperature and time of day [22, 81]. Variability makes measuring performance outcomes in a repeatable and reliable fashion very difficult. Additionally, most standard assessments are reported on ordinal scales. It is impossible to numerically differentiate these scales, meaning that small changes in function may be lost.

1.2 Technical Background

There are many barriers to providing the kind of care that literature has shown to be most effective. For one, insurance providers, particularly in the United States, place strict limits on the duration, location and timing of therapy that they will fund. Medicare, for example, has routinely enforced caps on the amount of outpatient therapy it will reimburse and a moratorium on caps is set to expire [3]. Moreover, home visits are hard to come by; Medicare will currently only reimburse home visits for individuals who are physically unable to leave the house [4]. This is in spite of evidence that people regain motor skills most readily in those environments where they feel comfortable and where tasks have real functional benefit [85, 149].

In addition to the barriers from insurance, psychological barriers can negatively effect the quality and character of therapy. "Learned nonuse" is one of these psychologic barriers. In addition and as posited by [18], stroke survivors may develop a low "self-efficacy" [33]. They may feel they are incapable of activity and resist engaging in therapy due to a fear of failure post-injury.

Physical and psychological limitations on therapy have inspired technical researchers to explore the potential for engineered devices, like robots or virtual environments, to fill gaps in care. There are obvious benefits of engineered devices for rehabilitation. Engineered tools, for example, can deliver hours upon hours of care without ever fatiguing. These same devices can deliver tightly regulated regimens of therapy, and they can precisely measure even the subtlest of change in mobility as it takes place. Finally, engineered devices can attempt to make what may be boring and repetitive practice more interesting, by contextualizing practice within games or inside augmented or altered feedback paradigms [18, 107, 116].

In this section, we first describe several engineered systems that have sought to facilitate semi-supervised or independent rehabilitation in an increasingly diverse range of environments. We then move on to explain technologies that remain largely unexplored by the rehabilitation engineering community, and the ways these technologies may potentially be applied. We focus in particular on video-based kinematic tracking, and sparse, task specific tactile sensing. Such technologies are of interest to us primarily because they are inexpensive. In addition, they are attractive because they can be integrated into "natural" environments like homes and workplaces without being very obvious. The desire to maintain the integrity of the environment in which people live and function guides our research.

1.2.1 Engineered Rehabilitation Devices for the Upper Body

Machine assisted biofeedback is one of the oldest forms of machine enhanced post-stroke therapy; this perhaps began with the invention of electromyography (EMG) but the first published reports of its use in clinics appear in the 1950s and 1960s [93]. Its clinical importance is still unclear, but it nevertheless has many of the qualities of the kinds of systems we desire. EMG devices are relatively inexpensive, portable, and they can be used in the home. Moreover, the intervention appears equivalent to traditional physiotherapy interventions that similarly target strength or muscle coordination [46].

More recently, there has been a great deal of work from the robotics and graphics communities to engineer systems that retrain relatively complex reaching motions [38, 106]. The MIT-MANUS is perhaps the most well known and widely deployed of these systems; it consists of a manipulandum which operates in a plane. Individuals hold the manipulandum as they make reaching motions to targets on a table top; the device assists as assistance is required. Tests of massed, repetitive MANUS-guided reach practice have now been made with over 250 stroke survivors, and have been shown to produce significant benefits both in terms of increased motor control and strength [64]. Other robots [21, 59, 106] similarly use force feedback either to assist individuals as they reach [59, 106], to surreptitiously increase the difficulty of reaches as patients improve [18], or to otherwise alter the character of motor learning that takes place [117].

Unfortunately, the ability to translate motor gains acquired during periods of reach training to every day functional motion is not guaranteed. In early trials with the MIT-MANUS, for example, motor gains achieved through robot training were seen only for movements of the form that were practiced. Motor gains, moreover, did not influence tests of functional performance of everyday activities [143]. Additional research has therefore focused on partnering robotic rehabilitation devices with movements and situations that more directly relate to real tasks. Noteworthy examples of the integration of functional simulations in virtual rehabilitation environments include that found in the Automated Constraint-Induced Therapy Extension (AutoCITE) system [82, 132], those recently developed for the MIT-MANUS [37], and that found in the Activities of Daily Living Exercise Robot (ADLER) therapy system [57]. The T-WREX [116] also is being used to simulate functional tasks, and Brewer has extended finger retraining with the PhanToM device to target functionally inspired pincer grasps [18].

Task specific training for stroke survivors in VR environments may in fact be similar to task training in the real world with a human coach. Several studies indicate VR training transfers well to motor performance real world environments [53] and others indicate that VR training may outperform real world training in some ways. Specifically, in [114], individuals who were trained in a VR environment to execute a "steadiness test" were able to translate what they had learned to real world performance even while task interference was taking place. Those trained in the real world, by contrast, had greater difficulty performing the learned task in the presence of task interference. There is still relatively little data about the benefits and drawbacks of VR training for motor impaired populations, however; further research is necessary. See [50] for a thorough review of results in Virtual Environment training after stroke.

In this thesis we seek to enable task training in environments in which functional movements are ordinarily situated. Toward this end, we have chosen measurement technologies that are not ordinarily deployed in VR systems. Rather than a tracking system requiring markers [100, 128], for example, we have selected technologies that, with some additional development, we believe can be cheaply and easily integrated into pre-existing functional environments like homes and workplaces.

We additionally seek to emphasize the development of technologies that are flexible and low cost. Related efforts to engineer low cost and home appropriate therapy systems have focused on interactions with desktop computers that are mediated by a mouse [107], a force feedback joystick [99] or force feedback driving wheel [52]. There is now considerable enthusiasm at the prospect of employing the Wii for home based rehabilitation games, as well. Desktop computer mediated interfaces, however, do not readily allow for motor practice that is task specific or which relates to those tasks that may be most motivating to an individual. Computer vision and task specific force sensing provides a relatively flexible set of input devices.

1.2.2 Kinematic Tracking for the Upper Body

A kinematic tracker is a device that can follow human beings in images and which is conscious of the location of limbs. For a kinematic tracker, humans are not simply blobs; they are fully articulated objects with limb segments that hinge at joints. It is possible to get some information about joint angles from these devices, while with with a blob based tracker it is not.

Tracking people's kinematics is a problem that has received substantial attention from the computer vision community [24, 103, 119, 121]. As tracking algorithms and tools improve, the computer vision community moves closer to producing reasonable alternatives to laboratory based motion capture that do not require markers. In some situations, markerless pose estimation and tracking has been shown to be accurate relative to commercial systems that track the reflections of infra red light off of markers [7]. The potential application of such tools to the field of rehabilitation science and technology, however, still largely remains unexplored.

Vision based kinematic tracking tools hold many qualities that make them desirable for rehabilitation applications. The most obvious is that vision based tools are inexpensive relative to commercial motion tracking devices; a VICON camera currently costs over $10,000 and an 6 camera system retails for more than $75,000. In addition, computer vision based toolscan be more easily moved around than motion capture technologies. They also image both the motion of individuals as well as the environments in which they operate, which makes interpretable review of motion possible in a way that is somewhat more difficult with systems that do not store images. Finally, computer vision tools increasingly prove to be robust, even in situations where there are no markers on the body or bodies being tracked [24, 103, 119, 121]. The main obstacle to the use of kinematic tracking in rehabilitation is that much of the tracking work is still at the research stage, and the use of existing software for visual tracking requires considerable effort and expertise.

Based on the potential strengths of these tools, we engineered a kinematic tracker for our research. We decidedly simplified the human tracking problem, however, in an effort to make our system reliable, robust to a variety of environments, and ready to be deployed. In this thesis, for example, we make use of colored markers to delineate limb segments in the upper body and we also make use of more cameras than may be required. This is to insure that we reliably detect limb segments, that we are robust to occlusions, and that we are guaranteed a certain degree of spatial accuracy in our tracking. In the future, however, we plan to use automatic tools to acquire and track the appearance of limb segments, as in [103]. We also plan to drop unnecessary cameras from the system, and build kinematic reconstructions with increasingly small camera sets. Reconstructions achieved with a single camera are found in [95, 124, 139]. The system we currently have is amenable to the iterative inclusion of more complicated algorithms to facilitate cheaper and more complex tracking.

Kinematic trackers frequently place priors on the body's kinematics and/or dynamics. A prior is basically a belief about the likely configurations of the body, and it is used to guide searches for body parts between frames and to discard detected configurations that seem improbable. Priors on human motion and/or kinematic configurations can be trained from sets of 2D images of the body [58] or 3D data from motion capture [17, 119, 120]; they may also be based on some engineered knowledge about the shape of the body [61] or content of the motion taking place. The difficulty, however, is that many kinematic trackers fail easily. They may detect body parts in places that they do not exist and, once detections have taken place, they may fail to recover the position of the body. This is due in part to the fact that priors are usually very complicated;

training based on a small set of data may fail to capture the true range of foreseeable body configurations. Humans can move very quickly and in unexpected ways, so forming a sound belief about where a particular body will be from one time point to the next can be very difficult.

To augment the predictive power of priors on the human body, many modern applications make additional use of bottom-up part detectors at each frame. A bottom up part detector is a device that responds to patches of an image that look like limb segments, irrespective of the patches' locations in the image, or proximity to other limb segments. A bottom up detector, then, does not depend on a generative dynamic or kinematic prior to suggest possible poses; rather, it introduces assemblages of individually detected body as pose possibilities. The use of body part detectors at each frame makes tracking look a lot like object recognition, where the object of interest is a human body. This way of looking at the problem has become popular and has made kinematic trackers more robust; examples can be found in [103, 121].

In our work, we stress a similar bottom-up approach to part detection and tracking in the interest of robust tracking. In our system, the appearance of body parts is made relatively easy due to the colored swatches of cloth we ask individuals to wear on limb segments as they are being tracked. At each frame, we consider not only locations for limbs that are dynamically or kinematically reasonable, but also locations in the frame that bear the right colors. Parts of the frame that are correctly colored represent our "bottom up" limb hypotheses; parts that are kinematically or dynamically reasonable, by contrast, are top down.

Future work will explore the development of tools to remove the colored swatches. More specifically, we must integrate automatic appearance acquisition for limb segments into our system. There are several examples of automatic limb detection and appearance acquisition. In [39], for example, filters are engineered to respond to pairs of oriented edges and patches of appearance that conform to the size and appearance of a limb segment. Potential "parts" are then assembled to form probable body configurations according to some kinematic prior. In [102], parts contained in probable assemblages are used to further refine appearance models. Another example of automatic appearance acquisition is found in [104], where appearance models are built when the body configurations conform to stereotyped “poses”. These poses are common to many human activities; one, for example, is a side view of an open legged stance that is generated mid gait cycle. Among stroke survivors, of course, stereotyped poses may be slightly more difficult to detect and parameterize.

Perhaps the main contribution of this thesis to the field of computer vision is its use of clinical metrics to evaluate the quality of kinematic tracking. Recent years have seen many instances in which kinematic reconstructions, be they based on single or multi-view camera input, are numerically judged relative to motion capture data [7, 72]. Evaluation criteria include squared error in 3D point reconstruction, joint angle error, or false-positive/missed-detection rates based on the overlap between estimated body configurations and configurations provided by motion capture. The right evaluation metric is not clear [41]. We circumvent these issues altogether. In our case, a good kinematic reconstruction is one that can relate well to and predict a clinician’s functional score.

1.2.3 Force Sensing in Hand Therapy

Our vision system focuses on the motion of the torso, shoulders and elbows. Finer motion of the hands and fingers, however, is not captured with vision based tools. To augment the data stream at the hands and fingers, we additionally explore the use of very sparse arrays of force sensors placed on objects. The force sensors that we use are thin, low cost, flexible printed circuits. When pressure is applied to the sensors, the resistance of these circuits change. The resistance of each circuit can then be read electronically, and converted to units that represent Newtons. Each sensor that we use costs roughly $7, and we limit ourselves to the use of 8 sensors on the surface of any object.

There are several motor symptoms common to stroke survivors that we can hope to detect with sparse arrays of force sensors. For example, we know that stroke survivors frequently exhibit a lack of force control that is manifest both in the amount and the timing of force they produce [48**?**]. We also know grip forces produced by the flexors tend to dominate in stroke survivors, and that these forces increase as affected muscles are used and may result in excess force during functional manipulation [23]. Disability in hands is known to significantly inhibit functional capacity [122], which has made devices to help retrain hands of great interest in the field of rehabilitation science and assistive technology.

The range of force sensing applications that target hand rehabilitation after stroke is currently small but is quickly growing and becoming diverse. Many existing systems have been inspired by results generated in robotic rehabilitation research for the arm. There is substantial variation, however, in the precise configuration of devices and the amount of force information used to guide training. In [84], for example, a VR environment is tested which tracks the kinematics of the hand during periods of grasps, and its use in therapy is evaluated against two devices that assist users in the control of force they produce. One is a pneumatic device, that utilizes EMG controlled bladders to facilitate finger extension; the other is a user-controlled orthotic. This orthotic has a cable which assists the users as they need it, and the force on the cable is monitored. Both assistive devices generated positive preliminary results when integrated into repetitive reach and grasp therapies, while the practice without assistance yielded no change in functional hand status. A similar rehabilitation application for the hand which requires force sensing is found in [18]; here force that is produced during pincer grasps between thumb and index finger is monitored with two force feedback devices. As with [84], use of this device in the context of a repetitive practice paradigm has yielded promising hand therapy outcomes across a variety of virtual feedback conditions.

There are many devices used in hand rehabilitation applicationse. The CyberGrasp glove[1], for example, facilitates very high degree of freedom kinematic tracking of the hand, yet it is both very expensive and heavy (more than 500g); this makes it unsuitable for most clinical applications. The pneumatically powered glove in [84] is light by comparison. PhanTom devices can deliver force to individual fingertips, but are complicated to use in multi-fingered grasping applications.

Instead of using a wearable device, we sense the force that is produced by clients of therapy on the surface of objects. There are several reasons for this, but perhaps the most significant is that we seek to respect channels of intrinsic feedback that come normally with functional manipulations. Moreover, we want to facilitate easy transfer of learning to real world domains. We do not, however, currently offer any kind of assistance to individuals as they produce force on objects, nor do we offer any haptic corrections of applied force as in [31]. At this point, we simply seek to demonstrate that measurements at the object level can discriminate level of disability and capture change in functional status over time.

Force sensing on object surfaces, of course, has its own challenges. For one, force sensing resistors are small, not terribly flexible, they conform poorly to curved or deformable surfaces, and large arrays can become expensive. Despite all of these difficulties, force measurement at the surface of objects has been explored in a variety of user input and robot control tasks. In [98], for example, relatively high resolution contact forces were used to constrain the configuration of synthesized three fingered grasps with reasonable accuracy. High resolution patterns of pressure recorded at the interface between a body and chair have also been used to posit configurations of the torso and legs [129]. The authors of [129] use their predicted configurations to allow sitters to control an animated car. Finally, reconstructions of full body posture have been generated from foot pressure data in [150], also for the purpose of animation and avatar control.

Chapter 2

The Measurement Device

Our goal in this thesis is to inexpensively and robustly quantify functional motion of the upper extremities of stroke survivors in ways that can both discriminate between levels of disability and detect changes in mobility over time. Moreover, we seek to do this in a range of environments and in a fashion that is consistent with assessments like the Arm Motor Ability Test (AMAT).

Typically when therapists use the AMAT to assess function, they watch as clients perform upper body tasks, like raising a comb to the head or dialing a telephone. Each subtask is scored on a 6 point scale (0 to 5) according to the quality of the underlying motion (i.e. its smoothness, fluidity) as well as functional efficacy. Those who receive a 0 may be completely flaccid on the hemiparetic side and unable to complete any given task on the assessment. Those who receive a 5, by contrast, may have motion that is indistinguishable from individuals who have never had a stroke. Various post-stroke symptoms may be manifest among individuals with intermediate scores, such as muscle synergy (i.e. increased flexion in characteristic patterns [20]), spasticity, or jerky, uncoordinated motions.

The research in this thesis suggests we can automatically perform assessments that are consistent with expert humans. Unlike human observers, however, automated systems can operate at any time of day, in a variety of locations and for a very long period of time.

The assessment tools we have chosen to use were selected first and foremost based on their

Figure 2.1: The experimental setup. At left is a desktop instrumented with cameras and at the right is a colorful jersey used to facilitate tracking.

Figure 2.2: The views provided by cameras.

cost and their ability to be integrated into living and work environments. In the future instantiation of our system, we do not want individuals to notice the presence of measurement devices. We do not want them wearing markers or using objects with wires, for example, as these may negatively influence the character of their motion as well as their comfort. At this time, however, we are using markers on the body and wires placed on objects simply to prove that automated assessments are possible. To the degree possible, we have made engineering decisions to facilitate easy integration of markerless and wireless technologies.

2.1 Hardware

The hardware elements we use to make measurements of desktop functioning are:

1. Cameras. A desktop has been instrumented with eight commercial camcorders, as illustrated in Figure 2.1. One pair of wide angle views exists to image the motion of the torso and arms in context, a second pair images mid range views, and the last two pairs focus on the hands and torso respectively. For stereo reconstructions, we currently make use of six of the eight cameras; we do not use data from the two close up views of the hands.

Each camera (a JVC GRDV 2000 or GRDV 3000) has 1 CCD and records progressive scan video at a rate of 30 frames per second. Cameras are synchronized with flashes from a red LED; one LED is located in the corner of each camera's view. A single flash denotes the start of a motion and a second flash denotes its stop.

Video is compressed in real time to MPEG2 format with an Axis 250 MPEG-2 Video Server. The computer which interfaces with the camera network is a 3 GHz PC with a 800 MHz front side bus. All communications between the computer and cameras are mediated by an ethernet network.

2. Markers. During recording sessions, we currently request each subject to wear a colored jersey, as in Figure 2.1. This allows us to easily localize and track individual limb segments of the upper body. In addition, the table top we use has been painted green to make background segmentation straightforward. The green covers only a limited portion of the background, however, meaning that there is still substantial background variation to be dealt with during processing.

3. Instrumented Objects. The system contains a selection of objects that have been instrumented with force sensing resistors. The objects for which we report results in this thesis include telephone keys, a selection of wooden blocks, and a spoon. The telephone and the spoon relate to tasks on the Arm Motor Ability Test (described in Chapter 2; see Table 2.1). The wooden blocks correspond to tasks on the Action Research Arm Test (ARAT) [140]. Images and schematics of the objects are shown in Figure 2.3 through 2.7.

In each object, 6mm square force sensing resistors are sandwiched between plates that link solid interior and exterior shells. The force sensors that we use are thin, low cost, flexible printed

Figure 2.3: A selection of force sensing objects. These include a force sensing phone, a cube and eating utensils.

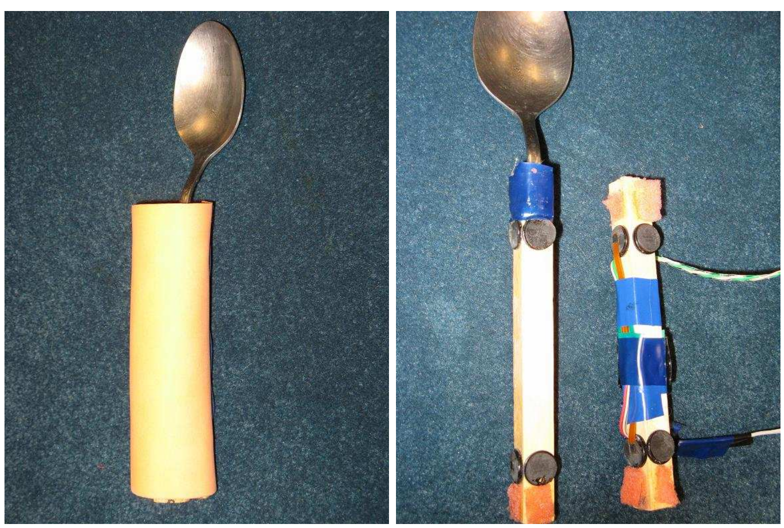


Figure 2.4: The interior and exterior of the force sensing spoon. Force sensors are placed at junctures between plates. Plates are attached to one another with compliant foam.

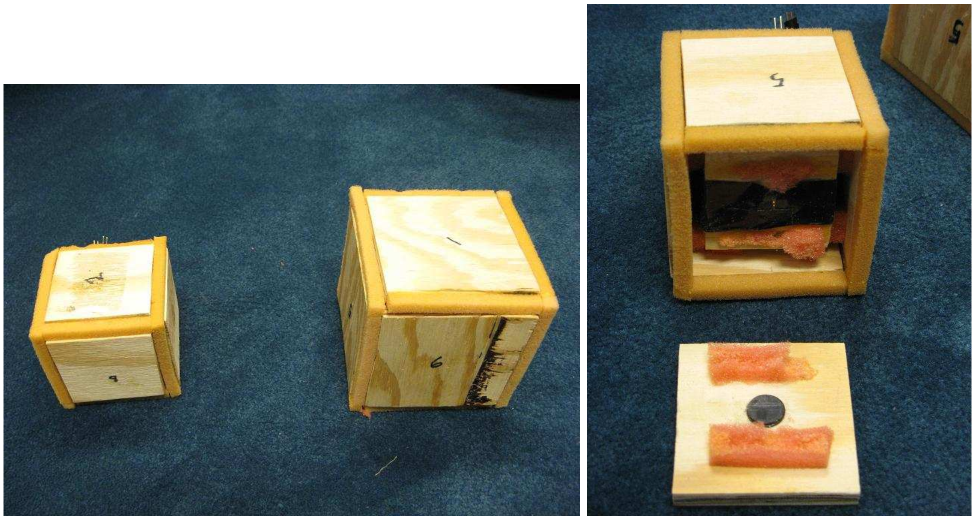


Figure 2.5: The interior and exterior of a force sensing cube. The construction is similar to the spoon; sensors are placed between plates at junctures.

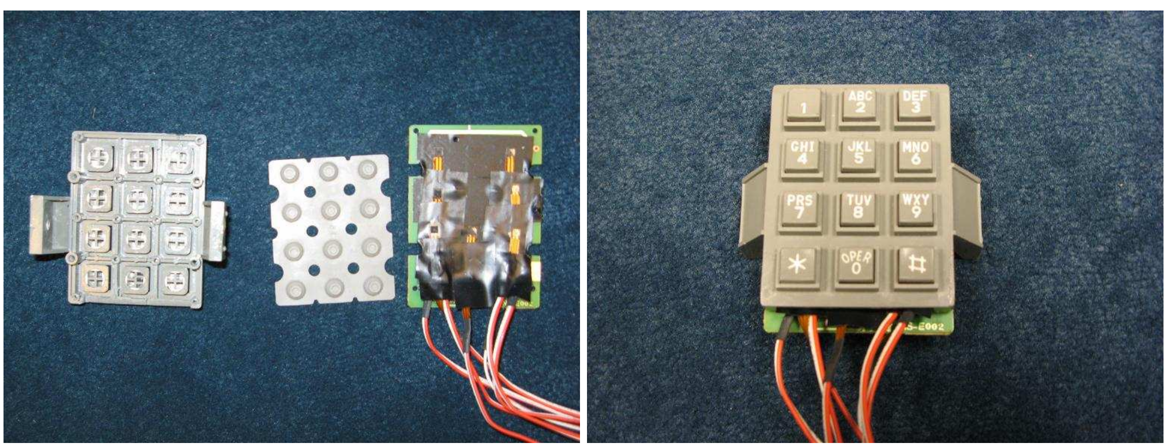


Figure 2.6: The interior and exterior of the force sensing telephone. Here, force sensors are placed at the juncture between the button and a rigid interior plate.

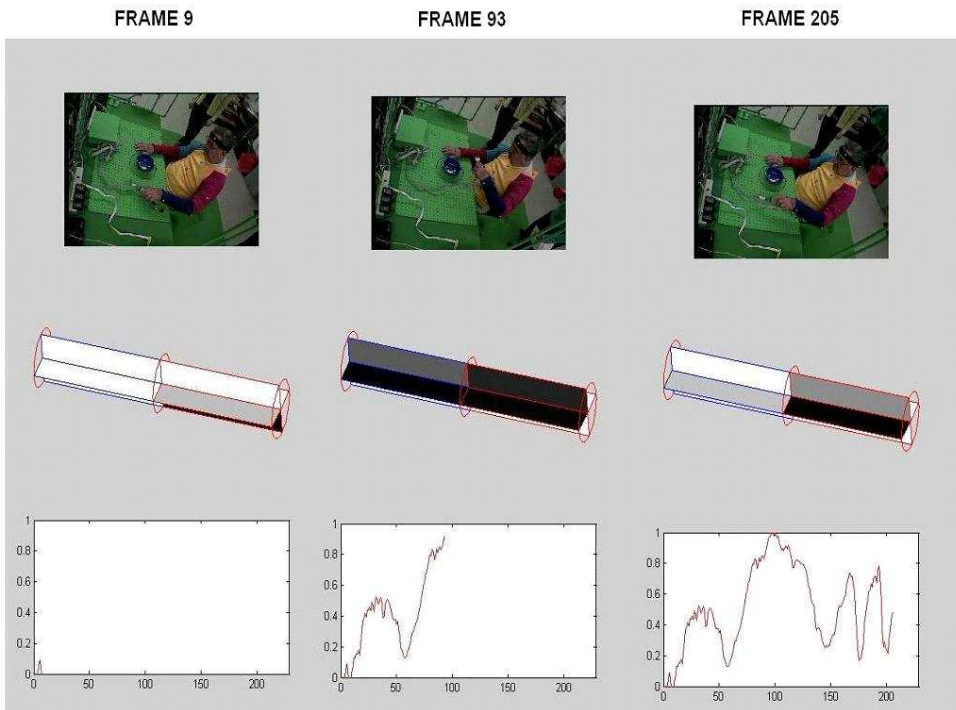


Figure 2.7: A force sensing object while being manipulated. In the top row is the individual using the spoon at three instants in time. In the middle row, the locations of the force perceived by the handle of the spoon are illustrated. The darker the color of a location, the more force recorded at that location. At the bottom is a force profile from the spoon. Force from all sensors is summed and illustrated over time.

circuits. When pressure is applied to the sensors, the resistance of these circuits change. The resistance of each circuit can then be read electronically, and converted to units that represent Newtons. The interior and exteriors of objects are connected to one another with a compliant foam material. The goal in this design is to maximize the transmission of force that is generated during grips and manipulations through the force sensing resistors. All objects contain between 4 and 8 force sensing resistors. The interiors of objects are illustrated in Figures 2.4 through 2.6.

When our system is running, force is sampled at 300 Hz and subsequently down-sampled to match the rate of the video. We use Data Translation's DT 9800 board with 16 channel inputs and USB link to perform analog to digital conversion. Acquired force data is synchronized with video by means of a single 5V electronic pulse that is fed to channel zero of the analog to digital converter; this is subsequently matched with the blinking of LEDs.

2.2 Software

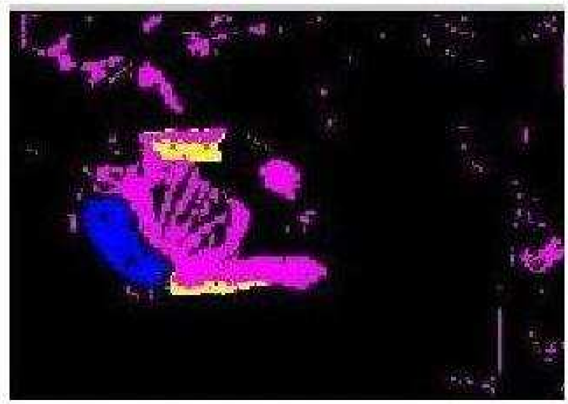
There have been two iterations of software in our system and more are planned for the future. In both iterations, force and vision sensing were involved. In this section we describe the software developments that relate to both components. Measurements reported in subsequent chapters were generated with the most recent instantiation of the system.

2.2.1 Vision

Our vision software creates three dimensional estimates of upper body limb segment locations, as follows:

A. First, six of the eight cameras are calibrated. We omit the two cameras that feature close up views of the hands. To calibrate cameras, we use the Matlab Camera Calibration toolbox, which is based on the work of [151]. This requires a planar pattern to be viewed by each camera

(a) Original image.



(b) Detected blobs. Many areas look like skin, as is indicated by pink.

(c) Filtered parts, assembled into an arm.

(d) Complete 3D reconstruction, back-projected onto image.

Figure 2.8: Vision processing stages. In the top left is the original image. To the right, clusters of pixels corresponding to possible limb segments are detected. These limb segments are then aggregated into ellipses, as illustrated in the bottom left. Finally, ellipses are triangulated across views to make a three dimensional reconstruction of the upper body. A reconstruction, back projected onto the original image, is illustrated at the bottom right.

at several different orientations. A single image of a planar pattern shared across all cameras determines the world origin. We rotate and translate the camera centered calibration matrices supplied by the Calibration Toolbox so they align with the corner of this shared planar pattern.

B. Second, beginning and end segments of a selected motion are located based on LED blinks. A simple piece of software scans selections of frames in each camera that correspond to the location of LEDs. When these areas become saturated with light, the onset of a recoding is said to begin. Unfortunately, the software is somewhat sensitive and is triggered by quick changes in lighting, changes in the position of the LED relative to the camera, and light background colors encroaching on the LED location. Data, then, is currently corrected by hand.

C. Camera frames are extracted from the video and down sampled. Elements of video corresponding to motion of interest are decompressed and stored as sets of 740 x 480 JPEG frames. These sets are then further down-sampled using bicubic interpolation by a factor of four, to speed processing. All JPEG images that are processed by the kinematic software are 185x120 pixels in dimension. The low resolution of these images in encouraging and means that we may be able to use substantially cheaper cameras with less resolution (around $10 USD) in future system iterations, assuming color quality is good.

D. In each frame from a given camera, limb segments are detected based on their color. For each color of the jersey, we train a quadratic logistic regression, as in [103]. One regression is trained for each camera and each subject. To train, regions corresponding to colors of interest are selected in images. Pixels inside selections are labelled 'positives' and pixels in a 50 x 50 pixel area around that selection are labelled 'negatives'. To be more specific, training involves using the Newton-Raphson method to locate coefficients β that maximize this value, summed across all 'positives' and 'negatives':

$$l = \frac{y * log(p)}{(1-y) * log(1-p)}. \tag{2.1}$$

Here, y is a vector of ones and zeros corresponding to 'positives' and 'negatives' in training data. The variable p represents the probability that a given y will be one assuming a given input, x:

$$p = \frac{1}{1 + exp(-x * \beta)}. \tag{2.2}$$

In our application, x is made of RGB values for selected pixels along with their squares and an offset variable.

Trained coefficients, or β, are then used to map pixels in all other frames onto either zero or one using a thresholded version of p. A one indicates membership in a part of interest. An example of an image after pixels have been mapped onto likely parts is illustrated in Figure 2.8.

After detection, pixels corresponding to parts are grouped into blobs, or contiguous regions. These regions are parameterized as ellipses.

D. 'Potential limb segments' are filtered in each frame with simple kinematic trees. To remove blobs that do not relate to limb segments, we use two very simple kinematic trees in a fashion inspired by [120]. Kinematic trees are tools that facilitate reasoning about connections between limb segments. Each potential limb segment is a blob, and the presence of a blob increases the probability that blobs for related limbs will be found close to it. Mathematically, the probability that a given set of blobs relates to a limb segment can be expressed like this:

$$P(p_1, p_2, p_3|I) = \prod_{(i,j)\epsilon E} P(p_i|p_j) \prod_{i=1}^{3} P(p_i|I). \tag{2.3}$$

In this equation, each p_i represents the strength of our belief that a given limb segment, i, can lie at a a particular location in the image. Our model has three limb segments, so this is why the equation contains p_1, p_2 and p_3. These three segments are the upper arm, the lower arm and the hand.

$P(p_i|p_j)$ represents the strength of our belief that the ith limb segment can lie at a given location, assuming we already know where the jth lies. The term E denotes the complete set of linkages between body parts that we choose to reason about. In our model, we have a link between the upper and lower arm, and one between the lower arm and either hand. $P(p_i|I)$, by contrast, represents the strength of our belief that the ith limb segment can lie at a given location, given the image patch that is at that location. If we are looking for a blue image patch at a location that is green, for instance, $P(p_i|I)$ will be very low.

Strengths of beliefs are determined by training. We use the same images selected for color training to additionally train beliefs. Based on parts selected for training color classifiers, we compute the relative angles between limb segments and the relative differences between adjacent segment centers. $P(p_i|p_j)$ is then defined to be a normal distribution centered at these idealized angles and locations. $P(p_i|I)$ is similarly determined based on the color, size and shape of a given blob relative to blobs we selected during training.

An example of a 'filtered' arm is illustrated in Figure 2.8.

E. Filtered 'parts' are combined across multiple views to make a three dimensional reconstruction of arms' locations. After filtering, we are left with a set of ellipses in each image that are adjacent to one another and that represent our belief about the location of arms. To reconstruct a three dimensional picture of the location of arms from sets of corresponding ellipses, we must find corresponding points across images.

In our application, we use points located at the extremes of ellipses to create correspondences across frames. Two points are located on the major axes of each ellipse, and two on the minor axes. Sets of points are then combined across multiple views by locating three dimensional lines that, when projected, pass through camera centers as well as individual points. Reconstructed points in 3 dimensions are determined to be those which are closest to a complete set of corresponding 3D lines. This 3D point is defined as [27]:

$$\sum_{i}^{N}(w_i(Id - u_i u_i'))P = \sum_{i}^{N}(w_i(Id - u_i u_i'))c_i. \tag{2.4}$$

Here, i is an index referring to an individual camera view, P is a reconstructed three dimensional point, c_i is the ith camera's center, and u_i is the direction of the ray extending from the camera's center to the point. Id is the identity matrix. Finally, w_i is an independent weighting factor for each view, indicating the level of "trust" we have in that view. We can use the value for $P(p_1, p_2, p_3|I)$, determined during filtering.

Any four three-dimensional points reconstructed from corresponding ellipses define a cylinder in space. These cylinders we use to estimate the location and orientation of each target limb segment. In practice, we do not allow the width of cylinders to exceed a pre-defined threshold. A reconstruction is illustrated in Figure 2.8.

F. Finally, results are refined by locating the three dimensional intersections of cylinder axes. The closest three dimensional points between the cylinder corresponding to the upper arm and the cylinder corresponding to the lower arm are used to define the locations of the elbow. These locations are in turn used to adjust the estimated positions of cylinders that correspond to both the lower and the upper arms.

A picture of elbow angles that are estimated using the kinematic tracker is shown in Figure 2.9.

2.2.2 Force

Our force software augments the three dimensional reconstructions of arms with information about the hands and fingers. Force data is processed as follows:

A. Sensors on objects are calibrated. To calibrate objects, known forces were applied to each

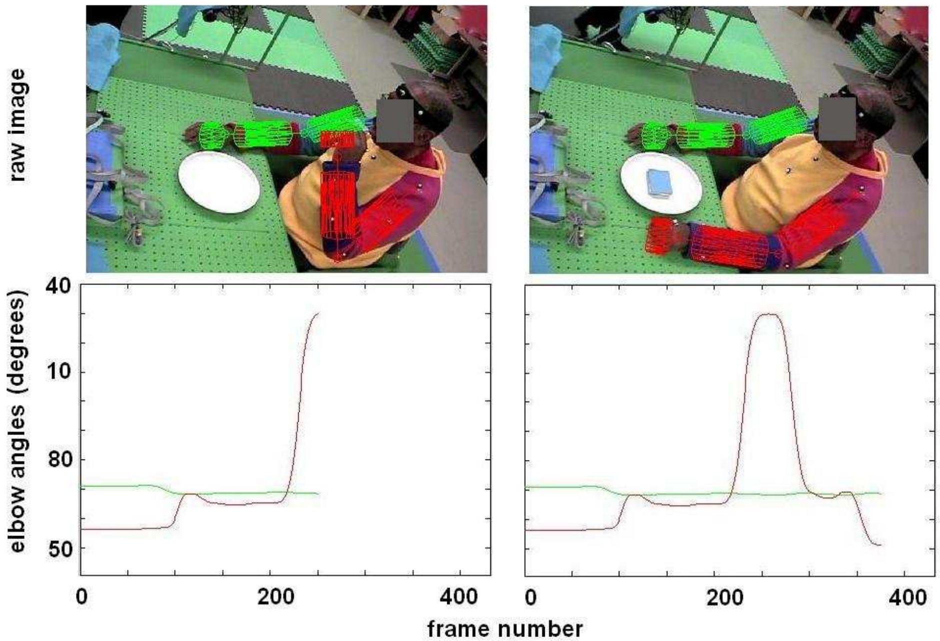


Figure 2.9: Elbow angles as reconstructed by the vision system. Elbow angles are shown at two different points in time during a grasp and lift motion. The angles relating to the left limb are illustrated in red, while those relating to the right limb are illustrated in green.

sensor on the objects individually. More specifically, we measure the resistance from each sensor when no force is applied, as well as when forces of 1, 2 and 4 Newtons are applied. A bicubic interpolation was then used to connect these measurements, and to estimate the number of Newtons that correspond to different voltage outputs.

Examples of calibration curves are illustrated in Figure 2.10. These curves are for the sensors inside the spoon. A significant issue here is that fact that, during calibration, sensors are exposed while, during manipulation, sensors are sandwiched between plates. This means that the exact number of Newtons recorded by each object depends on the ability of force to be transmitted through external plates. Calibrations, then, only provide a very rough idea of the number of Newtons recorded on object surfaces; they are by no means precise.

B. Beginning and end segments of a selected force data are located based on the presence of a 5V voltage pulse. As with the vision data, a simple piece of software scans force data to detect

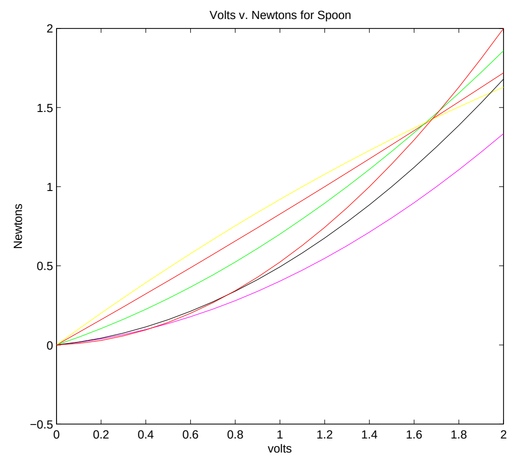


Figure 2.10: Calibration curves for one object.

the presence of a 5V indicator pulse. When the pulse is detected, the onset of a force recording is said to begin. The same technique is used to detect force recording terminations.

2.3 Remaining System Challenges

There are several modifications of the system that we would like to see in the future and which we believe will substantially enhance the measurement accuracy of the tools we have developed. Some of recommended system modifications follow.

2.3.1 Short Term Modifications

A. Multi-camera calibration. The calibration routines we currently use minimize re-projection errors in individual cameras. What would be preferable, however, is to perform calibration across all cameras. A multi camera calibration would minimize the re-projection errors of three dimensional points to which that all cameras simultaneously contribute location information. A camera calibration toolbox of this kind can be found in the work of [127].

B. Utilization of dense correspondences to make three dimensional reconstructions. Our reconstruction method turns dense regions of color into into a small number of points. Three dimensional reconstructions are based on the triangulation of these points. Condensing such rich information about parts into representative points, however, may discard useful position information. In future iterations, denser methods of part reconstruction should be explored. In the literature at the densest reconstruction extreme is the work of [24], which employs voxel carving to construct parts. For real time tracking a sparser set of correspondences may be required.

C. Utilization of three dimensional, as opposed to or in addition to two dimensional, kinematic tree filtering. We currently employ kinematic trees only in two dimensional images, as in [39]. This ensures that the arm segments in each two-dimensional image appear reasonable from frame to frame. There is, however, no such constraint on the appearance of three dimensional reconstructions from frame to frame. A three-dimensional kinematic filter, as in [119] may reduce errors the manifest at the reconstruction stage. A three dimensional tree, for example, can inform estimates as to the position of the elbow in three-dimensional space. Currently, this position is constructed on a frame by frame bases, and is located at the intersection of upper and lower arm cylinders. When these cylinders become almost parallel, the intersection of cylinders can change very radically from frame to frame. With a three dimensional filter, spurious elbow reconstructions could be eliminated.

2.3.2 Long Term Modifications

Desirable long term software modifications include:

A.Development of a real-time system. In order to enable feedback, we must operate in real time. Currently, it takes roughly 2 seconds per frame to generate a three dimensional estimate of the body's location from six corresponding frames. This is because out system software exists in Matlab, and it operates on JPEG images, not on video directly. Several real time tracking systems do currently exist, however, such as the one developed by [148].

B.Removal of visual markers. In addition, we want to integrate markerless tracking of the kind found in [103, 119, 139]. This requires re-engineering the kinematic tree we have developed to make it reason about more complicated limb appearances that contain texture information in addition to color.

C. Automatic appearance acquisition for limbs. Automatically locating bright colors in images is fairly straightforward; we do not need to tell our system the initial location of limb segments. In systems where limbs share more features with the background, however, the initial location of limb segments is not obvious. Because this is the case, we want to integrate automatic acquisition of limb appearances in some future iteration. Automatic acquisition of the appearance of limbs is discussed in [103, 104].

C. Automatic segmentation of interesting motion. Motion segmentation requires some limited form of activity recognition. We discuss this in our conclusions and outline of future work.

D. Wireless force tracking. Finally, the objects we have developed are currently connected to a Data Translation board with long wires. Wires influence people's style and comfort when it comes to manipulating objects. In the future, we must explore wireless options for communicating force information to the computer.

2.4 Prior Software Iterations

2.4.1 Vision

A prior iteration of the vision system tracked the position of balls placed on joints of the body instead of swatches of color. For the sake of completeness, we detail some tracking variations in our prior system here.

A. In our prior system, we tracked spherical balls placed on joints. The location of each ball was identified in the first frame from each camera by hand. Each target was followed from frame n to frame $n+1$ using the mean-shift algorithm of [29].

The mean shift algorithm is a mode seeking algorithm, which attempts to shift the window of focus in an image such that the distribution of colors found in that window matches the distribution of colors in a target. The closeness of a match between two color distributions is defined by the Bhattacharya coefficient, which is as follows:

$$\Sigma_{n=1}^{N}\sqrt{\frac{P(candidatecolor == color(n))}{P(targetcolor == color(n))}}. \tag{2.5}$$

Here, $\Sigma_{n=1}^{N}$ represents a summation over all pixels in a window, $P(candidatecolor)$ represents a distribution over all colors in a candidate window, and $P(targetcolor)$ represents a distribution of color in a target window. In our implementation, we used a search window that was 25 pixels wide. We additionally weighted elements of the histogram by their distance from the center of each window, as in [29].

In order to avoid losing targets at any point, we also made use of simple circle detectors at every frame. This involved first subtracting the mean color of each target from an image, and dividing by the target's color variance. Resulting colors were then thresholded and those closest to zero used to form candidate 'blobs' corresponding to markers.

Only circular blobs were considered candidate markers. "Circularity" was detected by evaluating the eignenvectors and eigenvalues of each blob's spatial covariance. Symmetric blobs have eigenvalues that are roughly equal. The likelihood that a given blob corresponded to a marker, then, was said to be inversely proportional to:

$$d(eig(1); eig(2)), \tag{2.6}$$

where $eig(1)$ is the eigenvalue corresponding to the first eigenvector of the blob's spatial covariance, $eig(2)$ the second, and $d(x; y)$ a simple distance function.

Finally, goodness of fit achieved by our simple circle detector was weighed against the fit achieved by the mean shift tracker. Goodness of fit was defined by the Bhattacharya coefficient; the solution with the highest Bhattacharya coefficient was saved.

B. Detected makers were combined across multiple views to make a three dimensional reconstruction of joint locations. The same technique to combine multiple views was used for this iteration. The sole difference is that correspondences were based on points at detected marker locations, not on regions corresponding to parts of a colored jersey.

This prior iteration of the vision software was abandoned simply because the reliable detection of markers was difficult given their relative size. The region based tracker facilitates more robust localization and tracking of body parts.

Chapter 3

Assessment by Human Experts

We begin with an exploration of the ways in which human therapists assess stroke survivors. There are many reasons we want to do this. First, we want to establish targets for our automated system. We want to make sure that our automated system assigns scores that are comparable to scores from human experts, and which agree with human experts as much as humans agree with one another.

More importantly, perhaps, we seek to gain insight into the kinds of features that experts think to be important when determining the capabilities of stroke survivors. There are two reasons to want this insight. First, the features that humans emphasize will help us select foci for automated perceptions. If, for example, the human experts think that range of motion at the elbow is a diagnostic feature, we will make sure that this feature is measured, and measured as accurately as possible, by the machine. Second, we want to determine how best to provide feedback about measured movement to therapists. With our automated system, we desire to emphasize features that humans care about; we do not want to distract from clinicians' lines of reasoning, instead of supporting them.

Because we are most interested in assessments that relate to real world functioning, we focus primarily on assessment in the context of the Arm Motor Ability Test (AMAT). This is because the elements on the AMAT directly emulate functional activity in a way that other assessments, like the Fugl-Meyer Assessment, do not. We also look at select elements of the Action Research Arm Test (ARAT), as the ARAT is more commonly used than the AMAT. Based on the work we report here, we establish the variation, repeatability and consistency of humans as these features relate to either assessment. We also determine those features the humans focus on during AMAT assessment, and get some sense as to the variation in their assessment criteria.

The same data that we show to human experts in this chapter will be used to train an automated assessment device in the next chapter. In this chapter, then, we record video and force data from several stroke survivors in a laboratory environment; we then show the video to human experts and ask for their opinions of the stroke survivors' functional capability. In the next chapter, we feed the same video and force information to a machine, and train the machine to determine functional health in a way that conforms with the opinions of the human experts.

3.1 Background on the Assessments

The Arm Motor Ability Test (AMAT) is an upper body assessment of stroke survivors that was developed in 1987. Items on the assessment are contextualized within functional tasks, like eating a sandwich or using a telephone. Therapists are required to subdivide each of these tasks into constituent components, like grasping or lifting; each component is then scored independently on a 0-5 scale. A zero indicates inability to use the affected side of the body during functioning, while a five indicates performance that is indistinguishable from that of a person who has never had a stroke. A complete list of the tasks on the AMAT can be found in Table 1.1.

The AMAT manual provides criteria that therapists are expected to use in order to make assessments [132]. For example in the sandwich task, the manual asks therapists to attend to features like the "dexterity of paretic fingers, the extent to which the head moves downward when the sandwich is brought to the mouth ... and the fluidity and precision with which movements are performed." A more complete list of criteria with which to judge movement quality as suggested by the manual can be found in Table 3.1.

Score	Criteria to judge movement quality
0	No movement is initiated.
1	Partial range of motion accomplished but movement is dominated by synergy; there is gross lack of coordination between limb segments, or limb is non-functional for weight bearing activities.
2	Movement is accomplished but is influenced by synergy or is accomplished by excessive compensatory movement of the trunk, head, or contra-lateral upper extremity. Lacks either proximal control or fine motor ability, or movement is performed very slowly, or limb is minimally useful in weight bearing activities.
3	Some isolated movement exists, but it is influenced to some degree by synergy. Alternately, movement is performed slowly, or moderate lack of coordination and lack of accuracy is manifest. Primitive grasping patterns are present and weight bearing activities are performed with difficulty.
4	Movement is close to normal but slightly slower or lacks precision, fluidity or precise coordination. Able to perform weight bearing activities but with mild hesitancy or mild difficulty.
5	Normal movement, fluid and coordinated activity. Speed of movement appears to be within normal limits.

Table 3.1: Criteria to assess movement quality from the AMAT manual.

There are, however, still a wide range of possible observations that an individual expert may find to be relevant during a given evaluation. A particular therapist may consider “dexterity”, for example, to be a function of the speed or smoothness of a particular movement. Dexterity may also be judged on the affected side relative to the unaffected, or it may be judged independently. This means these measures are fairly subjective.

Despite this possible variation, the AMAT has high inter-rater and intra-rater reliability. In [62], for example, two therapists used the AMAT to assess 33 stroke survivors at two points of time; Spearman correlations comparing the two therapists were found to be over 0.97 [126]. The internal consistency of the AMAT, as judged by Cronback’s alpha [30], was over 0.93, as was the test-retest reliability.

The Action Research Arm Test (ARAT) is a similar assessment of functioning, in that it is highly repeatable and consistent. Where the AMAT focuses on gross upper mobility in the context of explicitly functional tasks, the ARAT focuses primarily on dexterity of the hands and fingers. Sections are devoted to grasping, gripping, pinching as well as gross movement. There are 19 elements on the assessment in total. Items are arranged in order of difficulty; if a person cannot complete a task, tasks that are more difficult are automatically scored at zero. This saves time during the administration of the test; in practice it takes no more than 10 min to administer [32].

On the AMAT, each element is scored on a 4 point scale, where 0 reflects no motion and 3 reflects movement that appears “normal”. Differentiating between a score of 2 and 3, however, requires slightly more subtle judgements. Wagenaar et. al. set time limits for each item to help discriminate scores in the middle of the range [144]; scores are also automatically docked if individuals lean excessively during task performance.

As with the AMAT, the ARAT has been shown to be reliable and valid despite the potential variability in assessment criteria. In [141], for example, 2 therapists were asked to assess 20 stroke patients both in the laboratory and based on video tape. Therapists were also asked to review video tapes of stroke survivors at two different points of time. Inter-rater and intra-rater correlation coefficients were both found to be over 0.98. A similar study wherein 3 therapists assessed 105 stroke patients resulted in an inter-rater correlation coefficient (ICC) of 0.98 [51].

3.2 Protocol Analysis

Scores that agree on the AMAT or ARAT may be determined by therapists using different criteria. Some therapists, moreover, may not be fully aware of the criteria they are using to make their decisions [145]. Decisions that take place in clinics occur under conditions of uncertainty [28], which admits additional possibilities for variation in the features to which clinicians attend.

Protocol analysis is a methodology for verbally eliciting thought sequences in an effort to shed light onto underlying reasoning as it takes place [35]. The technique assumes it is possible for subjects to verbalize their task-related thoughts without altering their thoughts or their performance of tasks. In a study of pediatric nurses who were asked to verbalize their decision-making, no impact of verbalization was found on clinical performance [70].

Protocol analysis has often been used to explore variations in clinical reasoning. One study, for example, used the technique to explore methods of expert physical therapists in the treatment of cerebral palsy [34], and to differentiate these methods from those of novices. Others have used the technique to study clinical decisions made by nurses [43, 138]. More recently, the method has been deployed to understand user-interface requirements as they relate to a variety of computer supported medical tasks [54]. In [54], the working behavior of pediatric oncologists was examined so as to design interfaces that seamlessly integrate into pre-existing patterns of clinical reasoning.

In this work, we use protocol analysis to identify variations and commonalities in assessment criteria that are used by therapists to reason about perceived motor performance. We do this partly to understand those observable and measurable performance features that are weighed heavily by humans. More importantly, perhaps, we seek to learn features that are important to therapists so as to optimize the way we provide feedback. The ultimate goal is to provide

Therapist	Years in Neuro-rehabilitation	AMAT familiarity	ARAT familiarity
1	11	Basic	Intermediate
2	15	Basic	Basic
3	11	Intermediate	Basic

Table 3.2: Demographics on expert therapists.

feedback with an automated system that conforms with existing patterns of clinical reasoning.

3.3 Methods

The subjects for the protocol analysis were three expert Occupational Therapists. All were certified and licensed with at least 10 years experience in neuro-rehabilitation after stroke. Two had AMAT specific training, and all had at least basic familiarity with the AMAT and ARAT. Some demographic information on the experts is found in Table 3.2.

Each therapist was asked to assess a sample of eight stroke survivors. Stroke survivors were all at least one year post stroke, had a stable medical condition, and Mini Mental Scores, which are tests for cognitive functioning, over 24 (where the maximum is 30). The protocol was approved by the Carnegie Mellon Institutional Review Board and all stroke survivors gave consent. Data from this same population was used to automate assessments; results from this automation effort are reported in Chapter 4.

All stroke survivors were tested for spasticity, pain and sensory deficits. To test for spasticity, the Ashworth Scale was employed at the elbow and wrist [71]. A higher number on this scale indicates that more velocity dependent reflex activity was recorded when parts of the body were passively moved by the therapist. The Visual Analog Scale (VAS) was used to determine the pain individuals experienced both at rest and during motion [118]. This is a test in which individuals self-report about pain using a ten point scale, where ten is the highest. The Semmes-Weinsten test was used to measure tactile deficits. This test requires a therapist to touch a blind-folded client with filaments of varying diameters. Location specific deficits are detected when individ-

ID	Age	Gender	Years Post-Stroke	M.M.E.	Lesion Site	Dominance	Aphasia
1	75	Male	2	25	Left	Right	None
2	60	Male	1.5	27	Right	Right	Expressive
3	47	Male	22	26	Left	Right	Expressive
4	82	Male	12	28	Right	Right	None
5	64	Female	4	28	Left	Right	Expressive
6	58	Male	13	29	Right	Right	None
7	78	Female	7	30	Left	Right	Expressive
8	63	Female	35	29	Left	Right	None
ID	Ashworth Elbow	Ashworth Wrist	Touch	Vision	Audition	Pain (rest)	Pain (motion)
1	0	0	Intact	Impaired	Intact	5	5
2	1	1+	Impaired	Intact	Intact	10	5
3	1+	1+	Impaired	Impaired	Intact	0	0
4	0	0	Intact	Impaired	Intact	0	0
5	1	1+	Impaired	Impaired	Impaired	0	0
6	0	0	Intact	Intact	Intact	4	0
7	2	3	Impaired	Intact	Impaired	0	0
8	0	0	Impaired	Impaired	Intact	0	0

Table 3.3: Basic demographics of participating stroke survivors. The M.M.E. is a short test of cognitive functioning. The Ashworth tests for spasticity about various joints.

ID	Past Conditions	Marital Status	Caregiver	Education	Former Occupation
1	Gout, Cardiac	Married	Spouse	High School	Gas Man
2	Cardiac	Married	Spouse	Grade School	Bus Driver
3	Cardiac	Single	None	Grade School	Mechanic
4	Cardiac	Married	Spouse	High School	Sales
5	None	Divorced	None	High School	Sales
6	Cardiac	Married	None	High School	Warehouse
7	Osteoporosis	Separated	Child	Graduate School	Librarian
8	Osteoporosis, Cardiac	Divorced	None	College	Caregiver

ID	FMA Wrist	FMA Hand	FMA Arm	FMA Speed	FMA Total
1	10	13	36	5	65
2	5	13	23	3	44
3	7	13	21	5	46
4	7	14	26	6	53
5	0	4	9	0	13
6	10	14	35	5	64
7	3	9	21	6	39
8	10	14	36	4	64

Table 3.4: Lifestyle demographics and Fugl-Meyer (FMA) scores of the subject pool.

AMAT Task/Subtask
Task 1: *Cut "Meat"*
1. Pick up knife and fork 2. Cut "meat" (Play-Doh) 3. Fork to mouth
Task 2: *Foam "Sandwich"*
4. Pick up foam "sandwich" 5. "Sandwich" to mouth
Task 3: *Eat With Spoon*
6. Pick up spoon 7. Pick up dried kidney bean with spoon 8. Spoon to mouth
Task 4: *Comb Hair*
9. Pick up comb 10. Comb hair
Task 5: *Open Jar*
11. Grasp jar top 12. Screw jar top open
Task 6: *Use Telephone*
13. Phone received to ear 14. Press phone number
ARAT Task
Task 1: *Lift 10 cm block*
Task 2: *Lift 2.5 cm block*
Task 3: *Lift 5 cm block*

Table 3.5: AMAT and ARAT tasks chosen for protocol analysis.

uals fail to feel filaments. Finally, all stroke survivors were assessed using the upper body and coordination portion of the Fugl-Meyer Assessment (FMA) [42], which tests for the presence of muscle synergy patterns. The combined high score on these portions of the test is a 74; a higher number indicates less synergistic coupling was detected.

Basic demographics of the stroke survivors and results of assessments are listed in Table 3.3 and 3.4. The mean age of the pool was 66 years old and subjects were, on average, 12 years post stroke. All subjects were right handed and five of the eight had lesions on the left side of the brain.

After screening, each stroke survivor performed six elements of the Arm Motor Ability Test and three elements of the Action Research Arm Test in a laboratory environment. The AMAT and ARAT elements selected for performance are all listed in Table 3.5.

Both the AMAT and ARAT were administered by therapist number three. Locations of

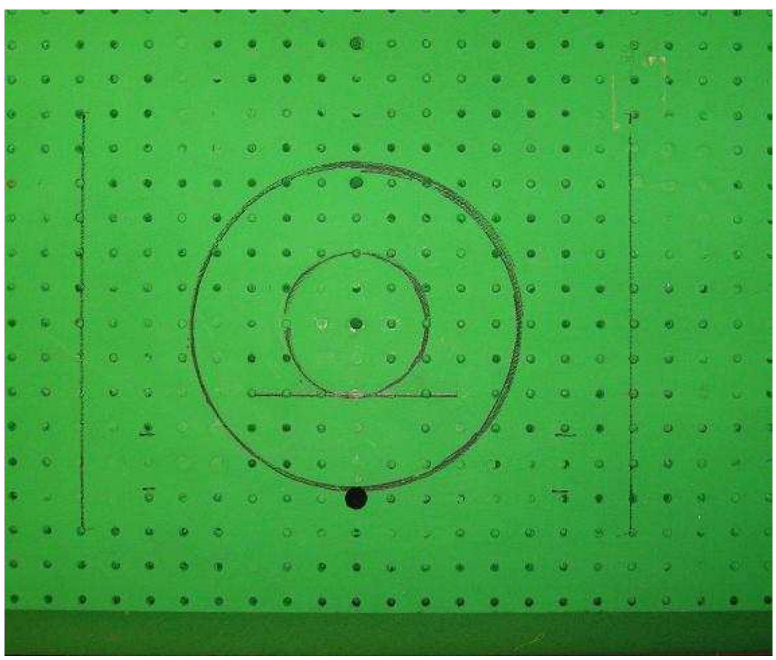


Figure 3.1: Object locations for the AMAT and ARAT.

objects were placed in standardized locations on the desktop; these locations are illustrated in Figure 3.1. The starting position for each subject was with torso touching the back of the chair, elbows at roughly 90 degree angles, and hands located in designated spots on the counter top. This is also illustrated in Figure 3.1. Stroke survivors were asked to perform tasks at the command of an investigator and at a comfortable movement speed, and to return their hands to the starting locations at each movement's completion. During performances, all stroke survivors were measured with a motion capture device as well our measurement instruments, described in Chapter 2.

Performances of assessments were video-taped from eight angles, as illustrated in Figure 2.2. Each camera was approximately 1 meter from the subject, and all cameras focused on the motion of the upper body and hands. Care was taken to provide views that showed the position of the back relative to the chair, so that determinations of leaning behavior could be made based on review of the video.

For the inter-rater reliability study, the therapists were asked to review the recorded video

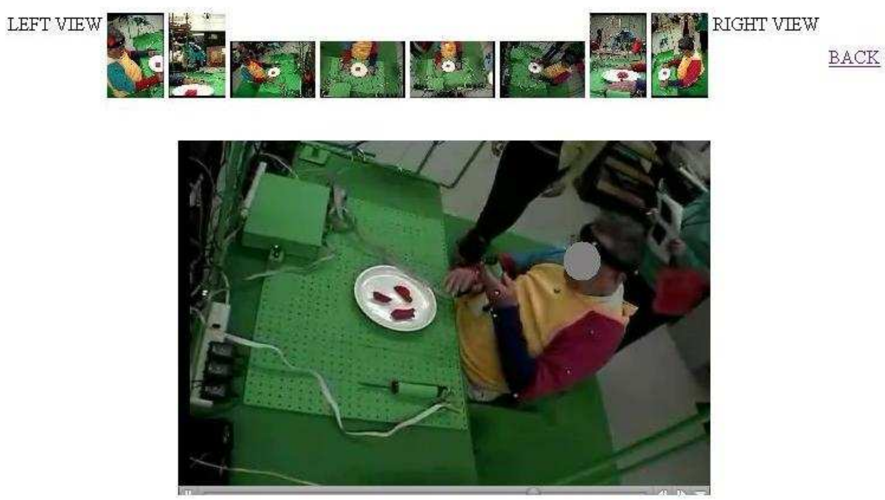


Figure 3.2: A screen shot of the interface used by therapists to make assessments. At the top are the views that were recorded of each stroke survivor. Clicking on a view plays the video corresponding to that view in the main window.

tape of each stroke survivor. The eight recorded views were provided to therapists and were organized according to the interface illustrated in Figure 3.2. Clicking on any one of the viewpoint icons at the top of this interface would make the video corresponding to that viewpoint play in the main window on the screen.

For the intra-rater reliability study, therapist number three was asked to make two assessments of the stroke survivors at different periods of time. The first assessment was made in the lab and was based on observations of task performance in real time. The second assessment was made after 12 months had passed, in an effort to erase recall of the earlier assessment period. These second assessments were based on recorded video, which was organized, as before, according to the interface illustrated in Figure 3.2.

Finally, during all video based assessments, therapists were asked to verbalize their decision making process. The instructions provided to each therapist requested that they generate a constant stream of verbalizations while reviewing video; if therapists were silent for more than a few seconds, they were prompted by the investigator. Audio recordings of all verbalizations were subsequently transcribed and coded for data analysis.

3.4 Reliability Results

Reliability of assessments was determined based on the intra-class correlation coefficient (ICC) [105]. The ICC assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. In order to determine these variations, we must make some assumptions about the sources of variation. For our inter-rater results, we assumed that each rating was generated by a different individual and so scores are drawn randomly. For the intra-rater situation, we assumed the raters to be the same and so scores are drawn from the same distribution.

We computed ICCs relating not only to the whole battery of AMAT and ARAT tasks, but relating to each task on the AMAT and for each subject individually. This was done in an effort to identify tasks or subjects that therapists found particularly difficult to assess. Finally, in order to identify potential rater bias, we employed a series of two tailed paired t-tests comparing each

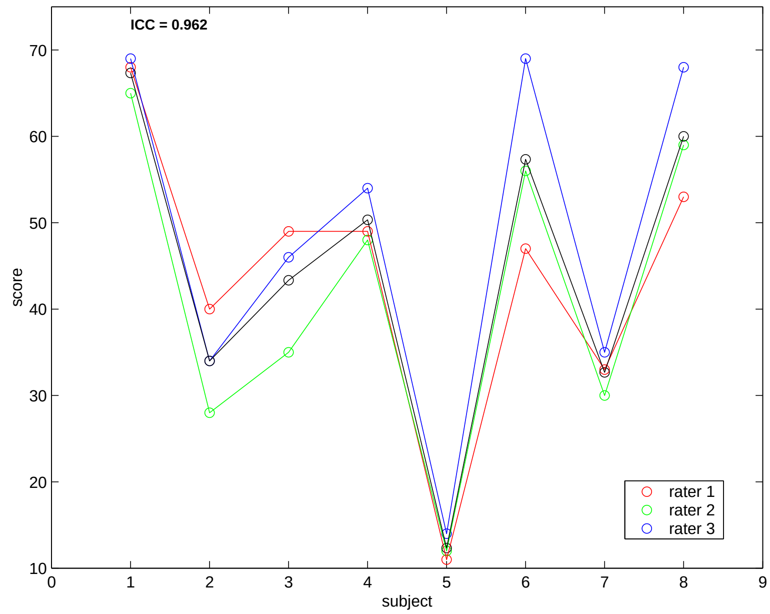


Figure 3.3: AMAT scores from all therapists. These are scores that have been summed over all six tasks in the battery.

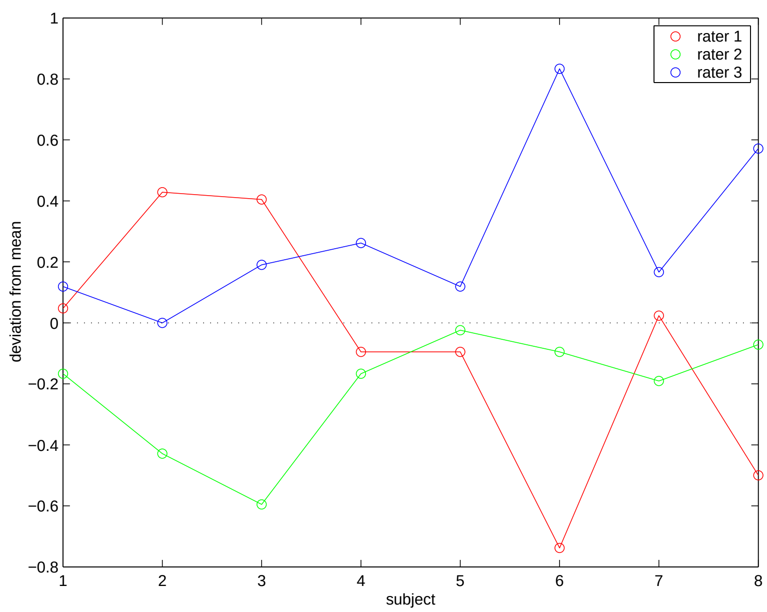


Figure 3.4: Deviation in therapists' aggregate AMAT scores from the mean.

Therapist	1	2	3
1			
2	0.4736 (-0.24, 4.49)		
3	0.1809 (2.51, 7.24)	0.0011 (4.64, 9.36)	

Table 3.6: Table of p-values for paired t-tests between raters' scores on the AMAT.

therapist to the others.

1. Inter-rater reliability on the AMAT. Aggregate AMAT scores generated by each therapist are shown in Figure 3.3. A different color indicates each therapist and the mean of all scores is drawn in black. The ICC computed across therapists was 0.96. Paired t-tests revealed a tendency of therapist number three to assign relatively high scores to subjects. P-values for two-tailed paired t-tests relating scores from this third therapist to the first and second were 0.18 and 0.001 respectively. The complete set of calculated p-values for t-tests is found in Table 3.6. The 95% confidence intervals surrounding mean differences between therapists' scores reflect therapist three's positive bias; the upper and lower bounds on these intervals are found inside parentheses in Table 3.6. A graph illustrating deviations between therapists' scores and the means of all scores is shown in Figure 3.4.

Dividing AMAT scores by task and subject reveals certain tasks and subjects that admit more

Therapist	1	2	3
1			
2	0.052 (-0.24, 4.90)		
3	0.296 (-2.07, 3.07)	0.048 (-0.74, 4.40)	

Table 3.7: Table of p-values for paired t-tests of AMAT scores for subject number three.

Therapist	1	2	3
1			
2	0.076 (-1.07, 4.07)		
3	0.000 (1.10, 6.24)	0.071 (-0.40, 4.74)	

Table 3.8: Table of p-values for paired t-tests of AMAT scores for subject number six.

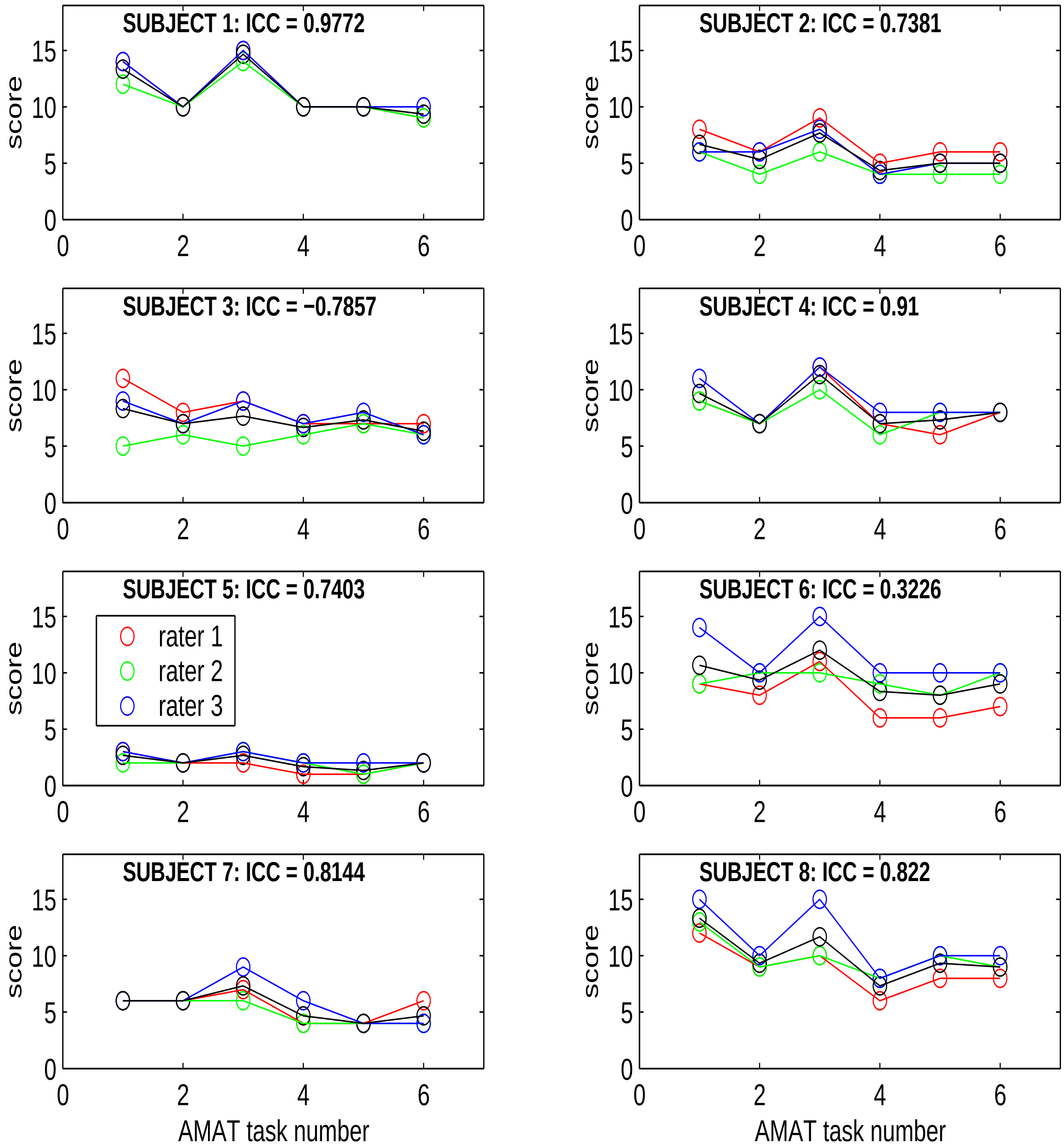


Figure 3.5: AMAT scores organized by subject.

variability in scoring than others. Subjects number three and six, for example, were particularly disputed by the therapists. The ICC values for these subjects' scores were -0.78 and 0.30 respectively. For scores related to subject number three, paired t-tests revealed an inclination on the part of therapists three and one to assign relatively high scores. The p-values for paired t-tests as well as 95% confidence intervals surrounding mean differences between paired scores are given in Table 3.7. For subject number six, the third therapist is once again found to be more inclined to assign high scores than her colleagues. P-values for paired t-tests related to the assessment of subject six are given in Table 3.8.

ICC values across tasks suggest the most highly contended tasks to be the "Cutting Meat"

Therapist	1	2	3
1			
2	0.15 (-1.1146, 3.6146)		
3	0.41 (-1.6146 3.1146)	0.02 (-0.36462 4.3646)	

Table 3.9: Table of p-values for paired t-tests of AMAT scores for task number one.

Therapist	1	2	3
1			
2	0.05 (-0.99, 3.74)		
3	0.11 (-0.99, 3.74)	0.00 (0.39, 5.11)	

Table 3.10: Table of p-values for paired t-tests of AMAT scores for task number three.

Therapist	1	2	3
1			
2	0.50 (-1.99, 2.74)		
3	0.08 (-1.24, 3.49)	0.05 (-1.61, 3.11)	

Table 3.11: Table of p-values for paired t-tests of AMAT scores for task number four.

task (task number one), the "Bean and Spoon" task (task number three) and the "Combing Hair" task (task number four). ICCs for these tasks were 0.926, 0.920, and 0.945 respectively. All other tasks featured ICCs over 0.95. For disputed tasks we also report differences between therapists. As in all prior tests, therapist number three had a tendency to produce scores that were slightly

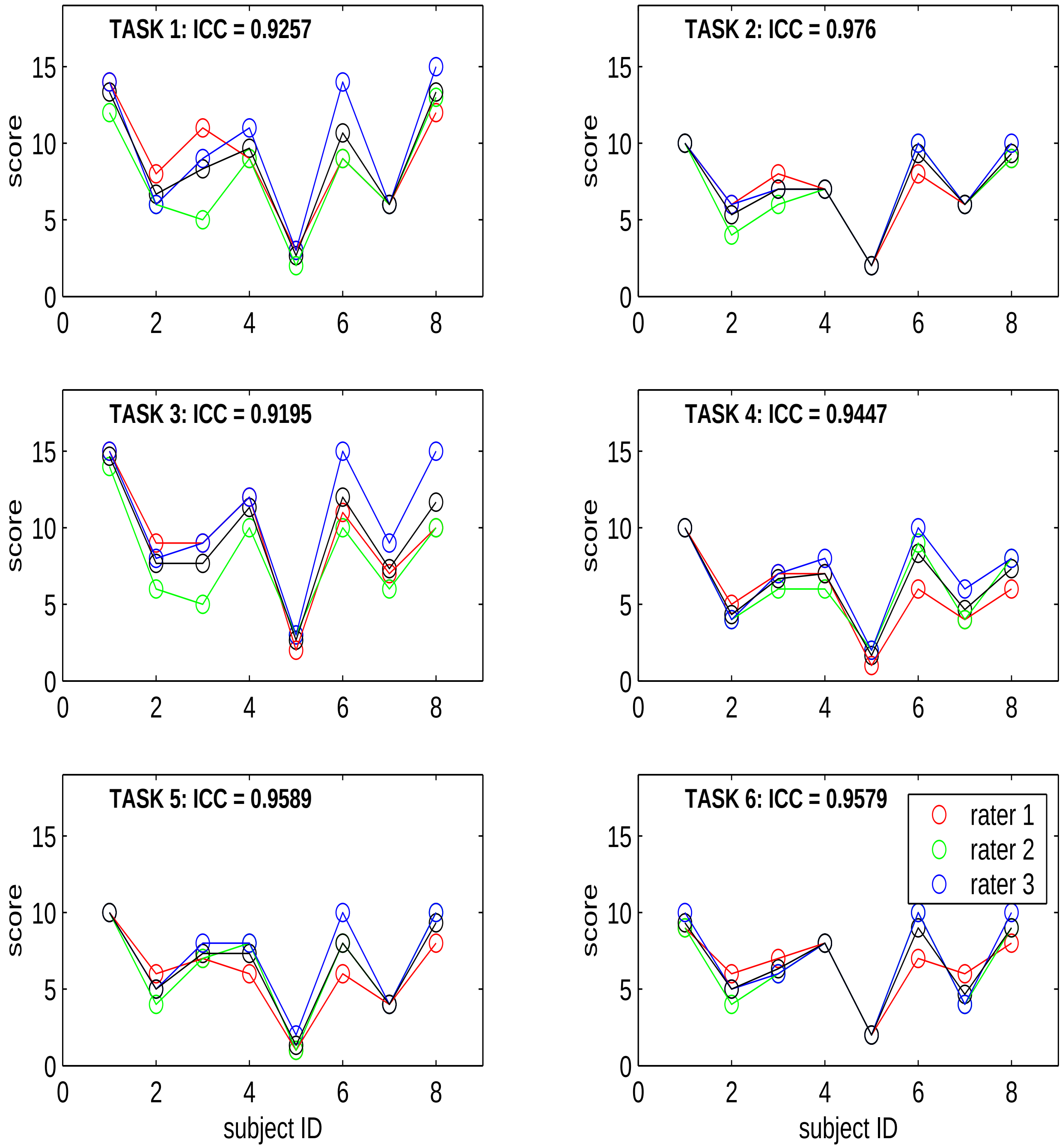


Figure 3.6: AMAT scores organized by task.

higher than those of the others. The most notable exception is found in task number one, where therapists one and three produce comparable scores. Therapist number two has a tendency to produce scores that are lower than her colleagues. The notable exception here is task number four, where therapist number one and two score consistently. P-values for all paired t-tests, and corresponding 95% confidence intervals, are found in Tables 3.9, 3.10 and 3.11.

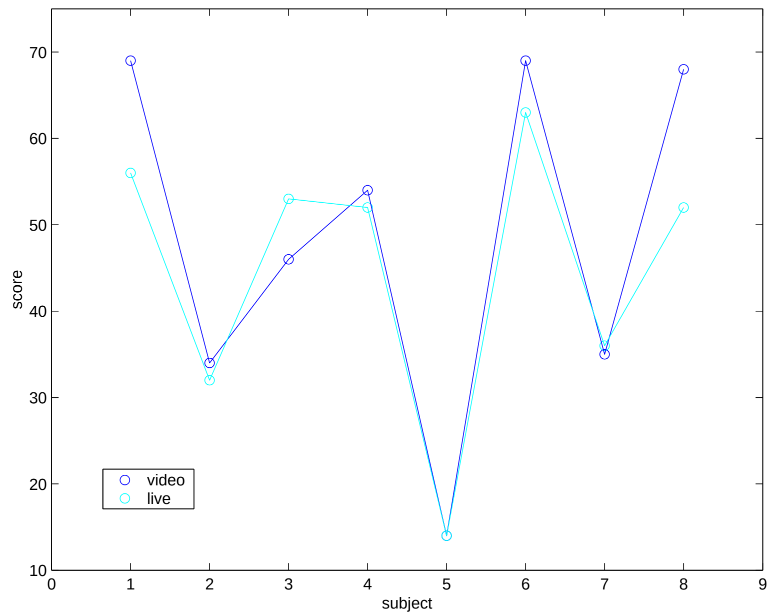


Figure 3.7: AMAT scores from one therapist at different times.

2. Intra-rater reliability on the AMAT. The AMAT scores of the single therapist that was asked to make assessments at two time points are shown in Figure 3.7. The computed ICC relating the first and second set of scores was 0.90. A paired t-test between the first and second set of scores suggested a slightly positive, but not statistically significant bias in ratings assigned to individuals at the later time point and based on video. The p-value for a paired t-test was 0.18, and the 95% confidence interval surrounding the mean difference between ratings was bound from above by 6.24 and below by 1.51. A graph illustrating deviations in scores at either time point is shown in Figure 3.8.

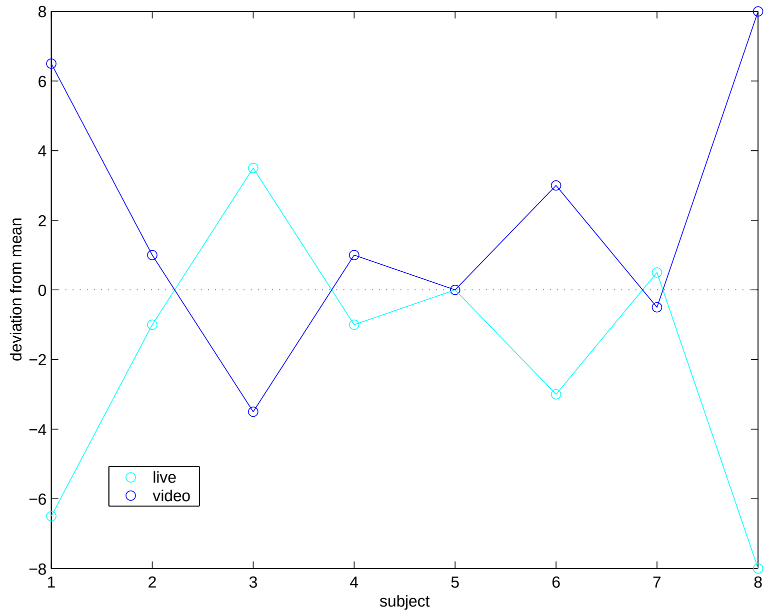


Figure 3.8: Deviation from the mean of one therapist's AMAT scores at two times.

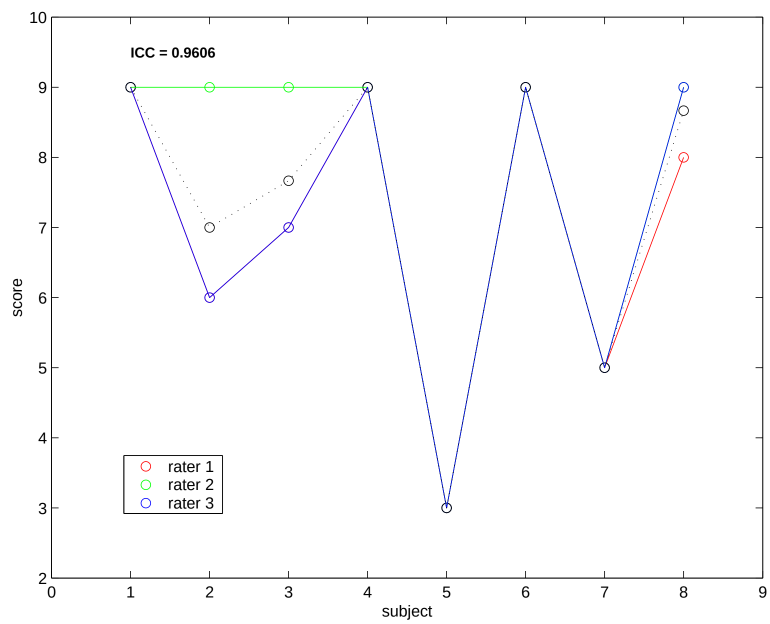


Figure 3.9: ARAT scores from all therapists.

Therapist	1	2	3
1			
2	0.1114 (-1.61, 3.11)		
3	0.3506 (-2.24, 2.49)	0.1803 (-1.74, 2.99)	*

Table 3.12: Table of p-values for paired t-tests on the ARAT.

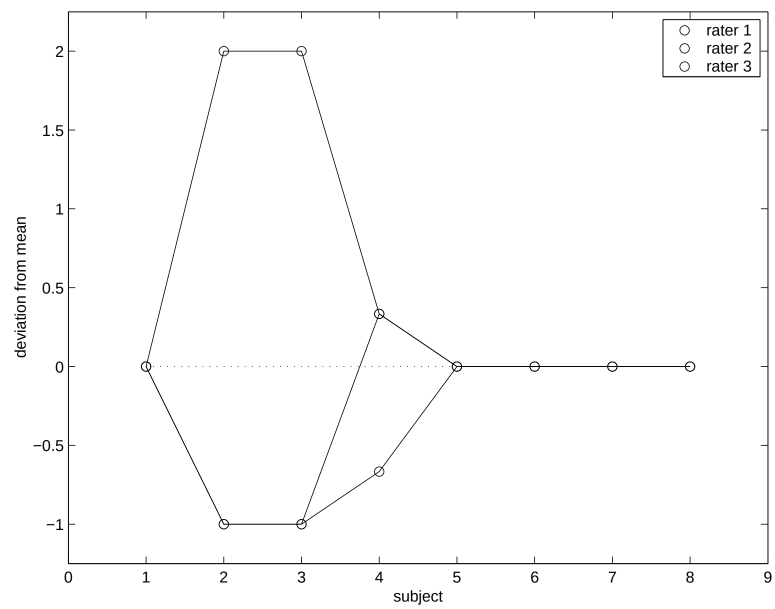


Figure 3.10: Deviation in therapists' ARAT scores from the mean.

3. Inter-rater reliability on the ARAT. Aggregate ARAT scores generated by each therapist are shown in Figure 3.9. On this assessment, the ICC value relating therapists to one another was 0.96. Paired t-tests showed no consistent bias between therapists; results from these tests are given in Table 3.12. As before, the upper and lower bounds on the 95% confidence intervals surrounding mean differences are placed inside parentheses. A graph illustrating deviations between each therapist's scores and the mean is shown in Figure 3.10. Interestingly, therapist number two assigned the maximum score to five of the eight subjects. The range of scores assigned to these same individuals by the same therapist on the AMAT, however, is 37 points, which represents more than 50% of the AMAT scale. A fairly strong ceiling effect is therefore present on the ARAT for this therapist, and, as a result, her scores do not discriminate between subjects as effectively as the AMAT.

4. Intra-rater reliability on the ARAT. ARAT scores from the single therapist asked to make assessments at two time points are shown in Figure 3.11. The computed ICC relating the first and second set of scores was 0.98. A paired t-test between the first and second set of scores

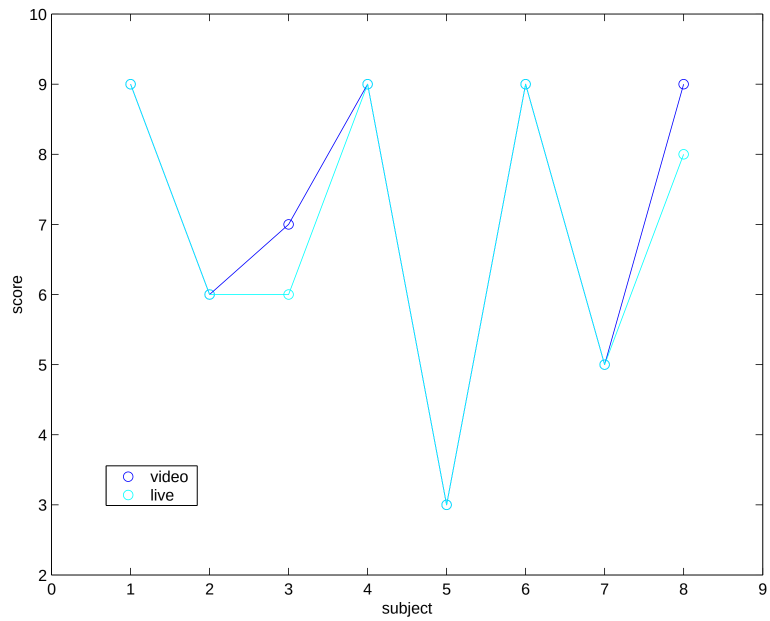


Figure 3.11: ARAT scores from one therapist at different times.

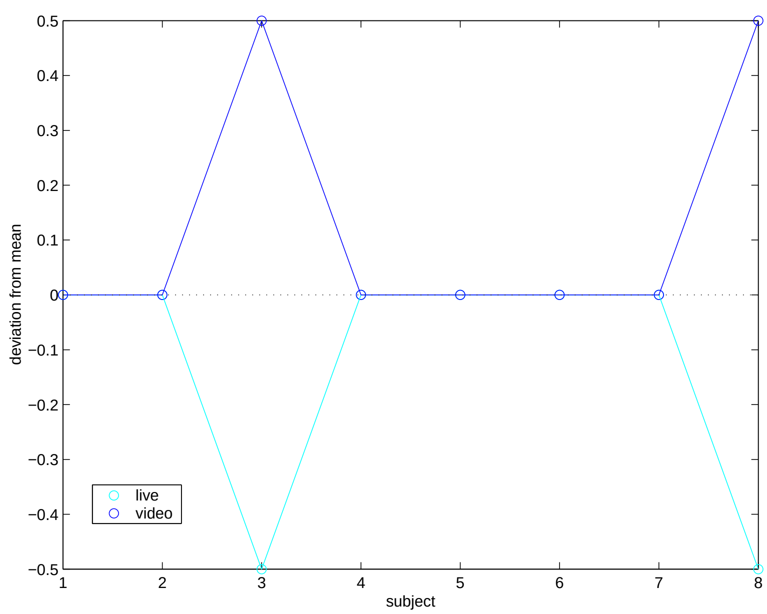


Figure 3.12: Deviation from the mean of one therapist's ARAT scores at two times.

Therapist	correlation
1	0.9004
2	0.7188
3	0.9543

Table 3.13: Correlations between AMAT and ARAT scores.

suggested no significant difference in ratings assigned at either time point. The p-value for a paired t-test was 0.18, and the 95% confidence interval surrounding the mean difference between ratings at either time point was bound from above by 2.6 and from below by -2.1. A graph illustrating deviations in scores at either time point is shown in Figure 3.12.

5. Correlation between AMAT and ARAT scores. Finally, we report the degree to which scores assigned by therapists on the AMAT agree with scores on the ARAT. Spearman correlations between each therapist's scores on either assessment are shown in Table 3.13. The correlation is particularly low for rater number two, as this therapist assigned the maximum score to five subjects on the ARAT but not on the AMAT . Correlations between AMAT and ARAT scores from the other therapists are comparatively high; all are over 0.90.

3.5 Discussion of Reliability

All of the reliability results, both intra and inter-rater, are comparable to, but slightly lower than, those found in previous literature. An ICC of 0.96 for the inter-rater reliability score of the AMAT is slightly lower than the correlation reported by [62]. Our number, however, reflects mean agreement between three therapists instead of two. For the ARAT, an ICC of 0.96 is slightly lower than the ICC of 0.98 reported by [51]. In [51], however, assessments of more than 105 individuals were compared; our sample size and task set is tiny by comparison.

The intra-rater agreements are also comparable to results reported in prior literature. An intra-rater reliability ICC of 0.90 compares that of 0.93 reported in [62]. An intra-rater ICC of

0.98 on the AMAT is equal to that reported by [141].

Our intra-rater results indicate that, for select items, the ARAT is a more repeatable assessment than the AMAT. This seems reasonable, given that the portion of the ARAT that is our focus is composed of a small set of tasks that all require very short movements. The portion of the AMAT that we have selected, by comparison, is composed of a wide variety of relatively complex tasks that must be divided into sub-tasks and scored independently.

While the ARAT is quick and repeatable relative to the AMAT, it fails to discriminate as accurately between levels of disability. This is most manifest in scores from therapist number two. Therapist number two assigns five individuals the same score when using the ARAT; these same individuals receive scores on the AMAT that cover more than 50% of the AMAT's complete range. The overall correlation between the AMAT and ARAT scores from this therapist is only 0.72, indicating fairly poor correspondence.

The tradeoff illustrated by these results is not surprising; assessments that are quick and easy to repeat may fail to discriminate between levels of disability in the same way that a more complicated, lengthy assessment can. Lengthy yet sensitive assessments, then, can perhaps benefit most from automated or semi-automated interventions.

In this thesis, we focus primarily on automating assessments that correlate with the AMAT. The AMAT was chosen because is relates to explicitly functional task performance, and it has been shown to feature high inter-rater and intra-rater reliability. More importantly, however, the AMAT is a sensitive instrument that is capable of discriminating between a wide variety of functional impairments. When we automate assessments, we will look first and foremost to match or correlate perfectly with AMAT scores that are assigned by expert therapists. Moreover, we will seek to match AMAT scores at least to the same degree that trained therapists match one another. We will additionally expect to guarantee the repeatability of automated assessments in a way that is provably difficult for humans.

3.6 Protocol Analysis Results

Identical assessment scores may not necessarily reflect the same assessment heuristics. Where one therapist may focus on the hands of a subject to assign his or her scores, another may focus on movement of the elbows or shoulders. To better understand the variation in underlying heuristics used by therapists, we look more closely at the verbalizations that were made as assessments took place.

Verbalizations were analyzed both qualitatively and quantitatively. To begin a qualitative analysis, areas where raw scores assigned by the therapists disagreed were the first point of focus. The hypothesis guiding this focus was that areas in which raw scores disagreed would correspond to areas in which assessment heuristics differed.

Based on qualitative results, emergent themes in verbalizations were selected for coding. Codes were assigned to all transcripts of the therapists' assessments. Distributions of codes were then analyzed to see if areas of disagreement could be measured numerically. Distributions of codes were also used to determine features of particular interest to the therapists.

3.6.1 Qualitative Overview of Verbalizations

Some themes that characterize areas of expert disagreement follow:

1. During assessments, therapists sometimes attended to different parts of subjects' bodies. This is most apparent in transcriptions related to subject six. Words that therapists use to describe this subject as he performs the "telephone" task follow:

Therapist one: "*Given the leaning, he is definitely at a three ... to pick up the phone. To dial the number he is pretty efficient and precise. There is a little bit of synergy b/t the thumb and fingers during (finger) extension (to push the buttons).*"
Therapist two: "*He could turn (the receiver) around in his hand, that was pretty good. I think he*

does well pushing the buttons; he has more trouble hanging (the receiver) up with his uninvolved side."

Therapist three: "*Picking up the phone he kind of twirls it around in his hand but I am not sure if that is spasticity. I think it is confusion about the set up and the wires. He does a fine job of pressing those numbers.*"

All therapists similarly focus on the fingers and hands of the stroke survivor as the task is performed. The first therapist looks closely at the "thumb and fingers" of the subject, and both the second and third look at the way the subject twirls the telephone receiver "in his hand". Only the first therapist, however, comments upon the tendency of the subject to lean as he reaches and she removes points from his score for this feature in particular. All therapists, therefore, focus on hands, but only one looks at the torso in order to make a functional judgement. This difference in focus, moreover, reflects a difference in resulting score. This first therapist assigns a combined score of seven to the subject in question, while the others give the same performance a combined score of ten.

2. Therapists sometimes differed in their analysis of cognitive issues. This is most apparent in the group's discussion of subject eight's performance of the "comb task":

Therapist one: "*She overshoots to get to the comb. It is a jerky motion. There is kind of an imprecision ... some of it is being silly but then some of it is lack of control. Especially on the left side, there is a jerk.*"

Therapist two: "*She just lacks fine coordination and she is really exaggerated with her motions. She lacks fluidity.*"

Therapist three: "*For the comb task she gets a four for all the fumbling to pick up the comb. I think that is a motor planning issue. Her orientation in space seems like a problem. And combing, her gesture ... has an odd velocity to it. It is nothing like actually combing your hair. So I* *will give her a four. But I don't think this is because of spasticity, I think it is other things. Lack of coordination and motor planning and overacting to compensate for that.*"

The first therapist calls the subject's task performance "silly" but does not attribute a major cognitive deficit to this subject. Therapist number two similarly calls the observed motion "exaggerated". Therapist number three, by contrast, sees "overacting" as a form of compensation for a possible "motor planning" deficit. She further postulates the subject to be having difficulty with "orientation in space".

These different evaluations of cognitive deficits *do not* explain a disagreement in raw scores, however. Therapists two and three assign the same score to this task performance, yet therapist three paints a very obviously different picture of the subject's cognitive faculties than her peers. It is therapist number one that assigns the lowest score to this task performance.

So what can explain the difference in assigned score? The low score from therapist number one may relate to a stronger emphasis on, or at least a different interpretation of, perceived movement jerk. More specifically, therapist one seems to see the jerk in the subject's movement as a symptom of a low level motor control problem while the others see the low level control in this subject as intact. Therapists two and three see jerk as a product of "exaggeration" or "overacting". Both see "odd velocity" or lack of"fluidity", yet they both believe the movements to be controlled. The first therapist, by contrast, is the only therapist who sees jerk as related to "lack of control", which may at least partly explain the difference in the relative weight placed on this feature.

3. Therapists sometimes placed different weights on creative and functional problem solving. The following transcriptions relate to assessments of the "spoon and bean" task from subject three:

Therapist one: "*I am giving him a three to pick up spoon based on his in-hand adjustments. He*

used an unusual grasp; it is a really interesting grasp. He is ... dominated by synergy but accurate."
Therapist two: "*He is very slowly picking up the spoon and getting the bean on the spoon. His hand is very pronated!*"
Therapist three: "*I am giving him a three for picking up spoon. I can see problem solving going on. He has learned to function with spasticity; once he gets the spoon (into his hand) he can make it work for him.*"

Therapist number three sees the subject as "problem solving" with the spoon and credits the individual for being able to "make it work for him" despite his spasticity. Therapist number one is similarly impressed by the grasping strategy of this subject, and calls his grasp "really interesting". Therapist number two, however, focuses on the speed and orientation of the hand, without commenting on the subject's "problem solving" ability or capacity to complete the task. For this task, therapist number two gives the subject a score of five, while the other two therapists, both of whom attend to functional problem solving skills, give the same subject a score of nine.

An emphasis on creative problem solving across the therapists, of course, is not the only difference between the therapists' comments. It is, however, the difference that seems to predict the disagreement in scores. Therapists also reveal different attention to features like "synergy", "spasticity", speed of motion or orientation of the wrist.

Despite differences, expert therapists were found to share several common assessment heuristics. Some examples of similar attention to observed movement features follow.

1. Therapists similarly emphasized grasping heuristics. The following transcriptions illustrate this; all relate to the same performance of the jar task:

Therapist one: "*He grasps the jar top pretty well ... He kind of hooks his hand over the top (of* *the jar) in order to get his fingers around it.*"

Therapist two: "*He grasped the jar top fine. He is holding the neck (of the jar) in order to twist the jar top open. He lacks fine precision in the grasping of the jar lid.*"

Therapist three: "*(This subject) surprises me. He has some wrist movement while he is turning which is very nice to see.*"

All three of the therapists are extremely attentive to very specific details of the grasp that is used during the jar opening task. Therapist one notes the manner in which the subject "hooks his hand" over the jar top and therapist two similarly notes the lack of "fine precision" during the grasp. Therapist three, however, sees "wrist movement" during the grasp that is not attended to by the other therapists. While all three are very attentive to the hands and fingers of this subject, therapist number three assigns the subject a comparatively high score. She gives him an eight for the task, while the others assign a seven.

2. Therapists placed comparable emphases on compensatory motion. The kind of compensations most commented upon by the expert therapists included use of the table to stabilize movements, leaning during reaches or manipulations, as well as use of the uninvolved arm. One example, relating to the a subject's performance of the "comb" task:

Therapist one: "*To pick up the comb she really leans forward with her trunk and uses that in exchange for her lack of shoulder. Combing is not very well controlled. You see shoulder elevation.*"

Therapist two: "*She brings her head to her hand.*"

Therapist three: "*I have to give her a three. She is just putting her whole body into it. She has a very weak wrist. I think part of the problem is weakness as much as spasticity ... and for the combing she does not do the back (of her head) and she is ducking a lot*".

All three therapists similarly comment on the subject's use of the head and torso to complete the task. Therapist number one, however, is the only therapist that comments on the use of the shoulder during the task performance, while therapist number three comments on the orientation of the wrist. While all three therapists focus on the torso, then, not all focus equally on other body parts, like the shoulder or wrist.

For this task, therapists one and two agree, while therapist number three assigns a comparatively high score. Results from the paired t-test results, however, indicate that therapist number three has a general tendency to assign higher scores than her peers. This may explain the difference in scores here, although differences may also be explained by differing emphases on features like "shoulder elevation", or an unequal weighting on torso compensation.

3. Therapists were often similarly attentive to the presence of muscle synergies. The following verbalizations relate to a single performance of the "sandwich" task:

Therapist one: "*There is some synergy in the hand*."
Therapist two: "*There is a good deal of synergy in that*."
Therapist three: "*She gets a three for picking up the sandwich because it is definitely influenced by synergy ... but she does accomplish (the task) with effort*."

On this task, all of the therapists comment on the synergy manifest in the subject's movement. Moreover, they all equally weight the impact of this feature, and assign identical scores to the task performance. Each of the three therapists give this particular task performance a six.

3.6.2 Quantitative Exploration of Verbalizations

The qualitative analysis of the therapists' verbalizations revealed that, at times, therapists similarly attend to features like smoothness of motion or the use of the torso during reaches. At other times, therapists disagree as to their emphasis on features like cognitive capacity, or hand orientation. Moreover, discrepancies in the ways that the therapists reason about the task performance often seem to relate, at least in part, to discrepancies in the numeric scores they assign.

In an effort to quantitatively characterize the verbalizations of the various experts, codes were developed for features that seemed to best capture areas of disagreement among the experts as well as areas of shared emphasis. Histograms of codes were then compared to one another in an effort to see if areas of disagreement and agreement could be captured numerically.

Two different varieties of codes were developed for the transcripts. The first variety described the parts of the body to which a therapist was attending while she spoke. These parts of the body were relatively coarsely considered and included the “head and torso”, “arms” and “hands or wrist”. Body location codes were used to capture moments when one therapist perceived leaning with the torso where the others did not, for example.

The second variety of codes described movement features commonly used to characterize the impact of stroke. These included features like “smoothness”, “speed”, “spasticity” and “muscle synergy”. Movement feature specific codes were designed to capture moments when one therapist saw “synergy” where another saw “spasticity”, or where one focused on the “speed” of a motion while the others focused on movement “jerk”.

A complete set of codes developed to label transcriptions can be found in Table 3.14. This same table provides some examples of transcriptions that were said to correspond with the various codes. All codes was considered to be binary variables associated with the transcription for a given therapist, subject and task performance. Two instances of the word “hand” in a specific transcription, then, resulted in a single assignment of the “HAND” code to that transcription.

A potential issue here is that several of the movement quality codes are coupled. Smoothness, for example, is affected by the presence of synergy as well as spasticity. Speed may be affected by strength or weakness. The codes, however, reflect the use of words that appear frequently in the transcripts themselves.

In order to verify the consistency of codes, two judges were asked to independently code

CODE	**Example of corresponding text from transcription**
BODY PART CODES:	
HAND	"I am going to give him a three because he doesn't extend all his *digits* into a full lumbrical *grip*."
ARM	"While cutting the meat he is abducting his *shoulder*."
HEAD/TORSO	"While combing his hair he is really moving his *head*."
MOVEMENT FEATURE CODES:	
STRENGTH/WEAKNESS	"I think part of the problem is *weakness*."
SYNERGY	"There are definitely *synergy patterns* going on in this guy."
COMPENSATION	Compensation with the unaffected arm: "He has to *use the right (unaffected) hand* to position the fork." Compensation with the torso: "She kind of fixes her elbow in flexion and *uses her trunk* to push the buttons." Compensation with the table: "It is almost like *she is counterbalancing her body by putting her uninvolved hand under the table* so that she can leverage her involved arm up to her mouth."
SPASTICITY	"He has interesting functional use with *spasticity*."
COGNITION	"He makes a lot of *decisions*; I am not convinced that the *lack of decisions* is stroke related, it might just have to do with aging."
SPEED	"The task is performed *slowly*."
SMOOTHNESS	"I put combing at three given the *tremor* you see there." "There is kind of an imprecision and based on the *jerks*, some of it is being silly but then some of it is lack of control especially on the left side there is a *jerk*."

Table 3.14: Codes applied to transcriptions.

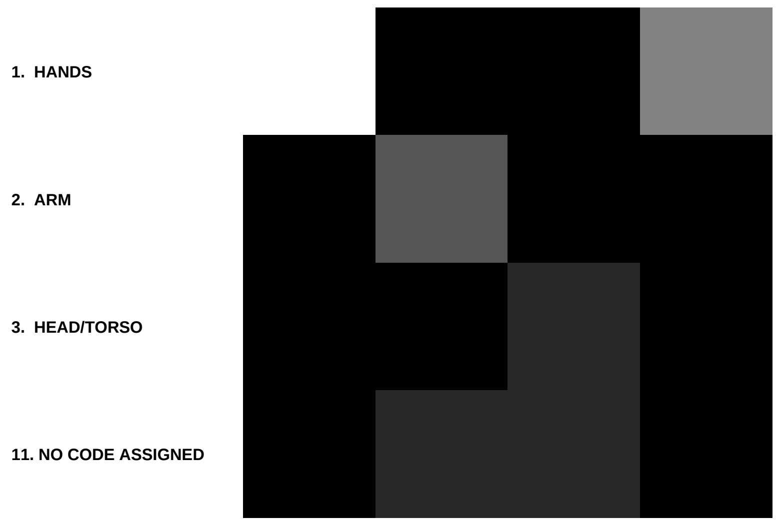
selections of recorded data. Selections were randomized across the subjects and tasks on the AMAT. Eighteen recorded selections were used for coding verification, representing slightly more than 10% of recorded data.

Matrices of coding agreement are found in Figure 3.13. Brightness along the diagonal of these matrices is scaled to reflect the number of instances where the two independent judges agreed in the the application of that code to a particular transcription. When judges did not agree, an entry was made in the “No Code Assigned” column.

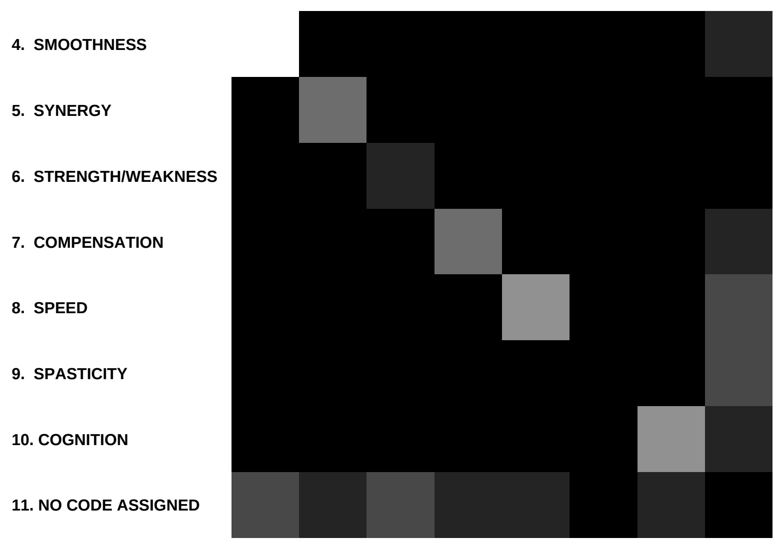
Cohen’s Kappa was used to measure agreement between judges [25]. For body location specific codes, the Kappa value was 0.45, while for movement feature specific codes the Kappa was 0.52. The Kappa related to the use of all codes in total was 0.55. Disagreements resulted when, for example, comments relating to “grasps” were interpreted by one judge as involving the arm, and another the hand. Despite these kinds of disagreements, the computed Kappas indicate moderate agreement between the judges in the use of codes.

One judge completed coding the entire set of transcriptions from the therapists. A total of 297 codes were assigned to the transcriptions in total. Therapists number one and three were assigned approximately 11 terms per transcription on average. Therapist number two was assigned only 7 codes per transcription, on average. The overall frequencies of codes that were applied to transcriptions are illustrated in Figure 3.14. The ICC value relating the three therapists’ distributions of body specific codes was found to be 0.85, while for movement feature specific codes it was 0.57.

Computed ICCs indicates that therapists were, for the most part, relatively consistent in their emphasis on specific body parts during assessments, yet less consistent in their emphasis on particular movement features. The hands were by far the most prioritized body part by the therapists, and they represented roughly 50% of all assigned codes. The torso followed, representing 27% of assigned codes, and the final 22% of codes related to arms. Smoothness and compensation were the most prioritized movement features among the therapists, each representing roughly 22% of assigned movement codes.

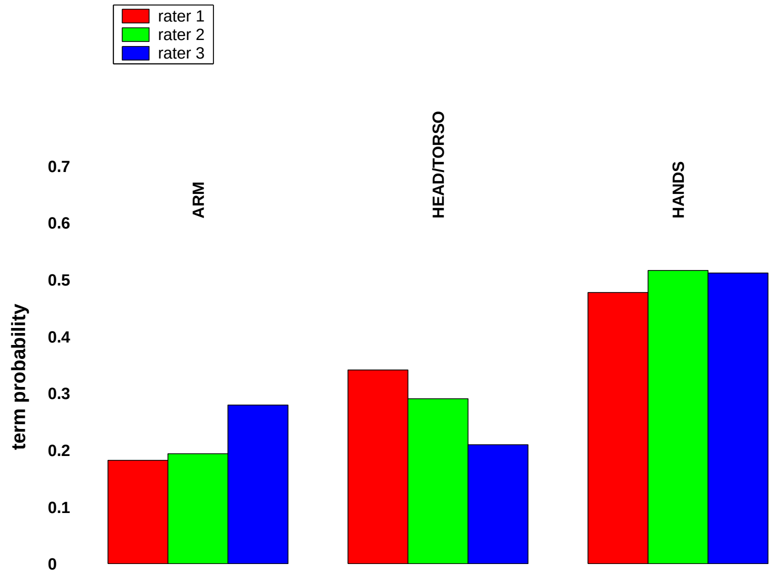


(a) Body Locations

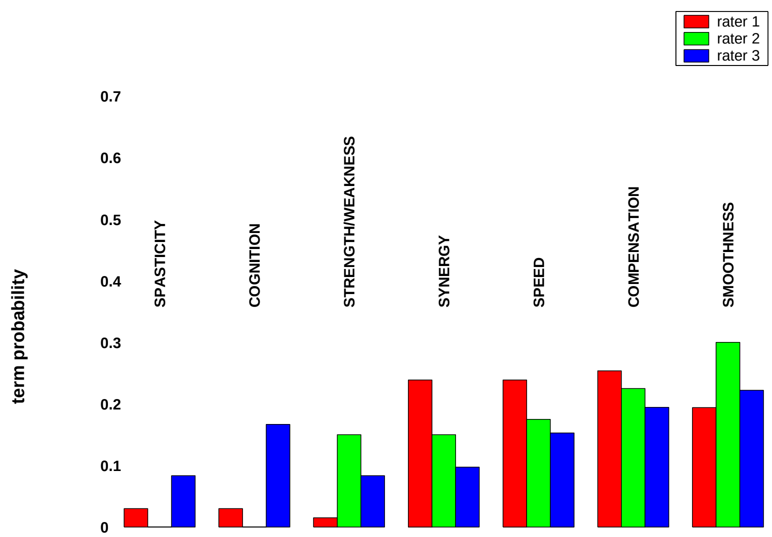


(b) Motor Features

Figure 3.13: Matrices of agreement in code assignment between two independent coders. Brightness along the diagonal is scaled to reflect the number of instances where two independent judges agreed in the the application of a code. Entries in the "No Code Assigned" column reflect disagreements.



(a) Body Locations



(b) Motor Features

Figure 3.14: Frequencies of assigned codes. Each color represents data from a different therapist.

3.7 Discussion of Protocol Analysis Results

Protocol analysis proved to be a useful tool to explore assessment heuristics regularly employed by therapists when conducting the AMAT. Results indicate therapists to be relatively consistent in their focus on different parts of the body during assessments, and slightly less consistent in the focus on different movement features.

Different assessment heuristics do seem to relate to different assignments of scores, but this is hard to capture quantitatively. An effort to code transcriptions revealed a loose, and quantifiable, relationship between attention to different parts of the body during assessments and resulting differences in score assignments, but this relationship was not found to be statistically significant. It also proved to be much easier to code for attention to different body parts that for attention to different motor symptoms. This may be in part because some of the codes used for motor symptoms are coupled. Speed, for example, can be the influenced by synergy, and smoothness can be influenced by spasticity.

The protocol analysis results have two different implications on the development of an automated assessment system. These are as follows:

First, the results indicate different parts of the body and movement features that might be of use when developing an automated system, and they indicate the relative weighting that experts place on each of these features and body parts. The therapists prioritized observations relating to the hands and fingers, as should, conceivably, an automated assessment device. Moreover, the experts were found to prioritize the smoothness, compensatory strategies, and speed; an automated system should perhaps do the same.

Second, the results point to a dilemma that a system which provides feedback about functional assessments must address. The same numeric score on an assessment was found to relate to several different assessment heuristics on the part of individual therapists. Even if a machine can produce numeric scores that are identical to an expert, then, there is still some ambiguity about how to report results. Which details will be most valuable to a given therapist? The variety of heuristics discovered through protocol analysis indicate that a variety of feedback options are required.

Chapter 4

Predicting Assessments in the Lab

4.1 Introduction

Recovery from stroke is a long term process. Stroke survivors have been shown to make significant gains several months and even years after the acute phase of their care, and long after they have been discharged from a hospital [36, 38, 92, 133]. A growing understanding of the duration of stroke recovery has led to an increased interest in therapies for chronic stroke survivors, and therapies that can be used cost-effectively outside of clinics in individual homes.

Some evidence-based guidelines for therapy after stroke include: emphasis of body parts most impaired by stroke; intense, repetitive practice of movements; practice of functional tasks; and practice in motivating and comfortable environments, like the home. Unfortunately, insurance in the United States rarely covers the kind of therapy that is most recommended by research. Medicare, for example, has routinely enforced caps on the amount of outpatient therapy it will reimburse and a moratorium on caps is set to expire [4]. Home visits can be even harder to come by; Medicare will currently only reimburse individuals who are physically unable to leave the house [3]. A significant gap, then, exists between the kinds of therapy that is available to stroke survivors, and the kinds of therapy the literature suggests for optimal outcomes.

In this thesis, we explore the potential for simple vision and force sensing technologies to fill gaps in post-stroke health care. We specifically envision a situation in which cheap, ubiquitous technologies are integrated into the homes and workplaces of stroke survivors for long term monitoring after hospital discharge. There, we expect technology to measure changes in movement quality as they relate to real functional behavior. The benefits of this kind of measurement are many: it can help therapists to better judge outcomes as they relate to real world tasks, enable coverage of long term rehabilitation that takes place in the home, and facilitate novel forms of motivating feedback.

The need for home-based rehabilitation options is well understood in the rehabilitation science community. Many example systems that have been developed, however, focus primarily on interactions with desktop computers and are mediated by external input devices. In [107], for example, mobility measurements are made with a mouse, while in [99] a force feedback joystick serves as an input device. There is now considerable enthusiasm at the prospect of using the Wii for home based rehabilitation games as well. Although desktop systems have demonstrated promising potential to capture changes in mobility that take place at home, they do not readily allow for practice that is task specific or flexible. Moreover, these example systems require an individual to hold or wear a device as they practice.

Computer vision and task specific force sensing hold the potential to provide a desirable set of input devices for rehabilitation applications. They can potentially track movement changes that take place in relatively natural environments. While commercial motion tracking technologies currently require markers to be placed on the body, future trackers may ultimately require none [24, 103, 119, 121]. Force sensing embedded in functional objects similarly promises to create measurement situations that are relatively flexible and "natural" in that they do not require gloves or other potentially uncomfortable mediating devices. Video camera systems and force sensing systems are also comparatively cheap. A commercial motion capture device can cost as much as $10,000 for a single camera, and a whole system may cost over $75,000. In the research we present here, each camera is worth less than $800 and we can potentially suffice with cameras that cost under $20 each. The force sensing resistors that we use are also affordable;

each currently costs roughly $7, and there are no more than 8 in a single instrumented object. Finally, a potential benefit to be derived from video based tracking systems is that they can provide interpretable feedback. Commercial motion tracking systems, by comparison, image solely the motion of markers, which may make easily interpretable review of recorded data challenging.

In this chapter, we show that simple vision and force sensing tools can be used to discriminate levels of physical and functional health among stroke survivors across a variety of tasks. We do this by taking measurements from stroke survivors as they perform desktop tasks and relating the measurements to assessments of health on the Arm Motor Ability Test (AMAT). The hypothesis that guides the research is as follows:

Inexpensive vision and force sensing tools can measure statistics of ordinary functional motion that both correlate with and predict the impairment of stroke survivors.

Our results demonstrate the possibility of measurement devices that can extract clinically meaningful measurements of stroke survivors cheaply and in functionally important environments, like homes and workplaces.

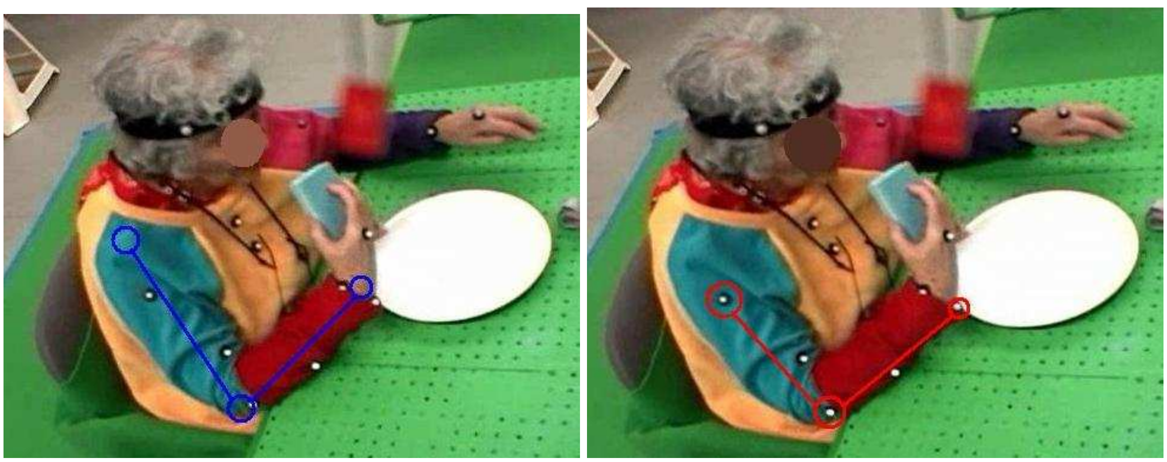
4.2 Methods

This chapter relates to the same data that was acquired and analyzed by human experts in the previous chapter. Where the last chapter revolved around the analysis of the data as made by humans, however, this chapter revolves around analysis of the same data as made by machines. The methods for data collection were previously described; we clarify and repeat some of the salient details in these methods here.

Basic demographics of the stroke survivors and results of their intake assessments are listed in Table 3.3 and 3.4. One stroke survivor (ID #5) was excluded from the machine analysis due to flaccidity on the impaired side, which was reflected in an FMA score below 25. The remaining subjects in the pool were all right handed and five of the seven had lesions on the left side of the brain.

Recording of AMAT performances took place in a motion capture lab at Carnegie Mellon University. Prior to any recording, all subjects were outfitted with a set of motion capture markers in addition to a colorful jersey of our design. The markers and jersey were used to facilitate easy tracking with both a commercial motion capture device and the tracker of our design (described in Chapter 2). Tracks from the commercial motion capture served as a tracking "gold standard" for the purpose of evaluation. They also facilitated the search for descriptive kinematic statistics.

The locations of motion capture markers are illustrated in Figure 4.1. Markers were placed



(a) Points for Reconstruction from Video (b) Points for Reconstruction from Motion Capture

Figure 4.1: Points from video and motion capture that were used to reconstruct 3D kinematics.

on the head, at the sternum and clavicle as well as at the base of the neck. In addition, markers were placed on the acromion process of either arm, the lateral epicondyle of either humerus and the styloid processes of the ulna. Only a subset of these markers, however, were used to estimate the motion of the upper body. This subset was chosen to most directly match the tracking capacity of the video based system, and it included markers on the acromion processes, elbows and wrists.

The kinematic statistics generated by the video-based system, by comparison, related to three

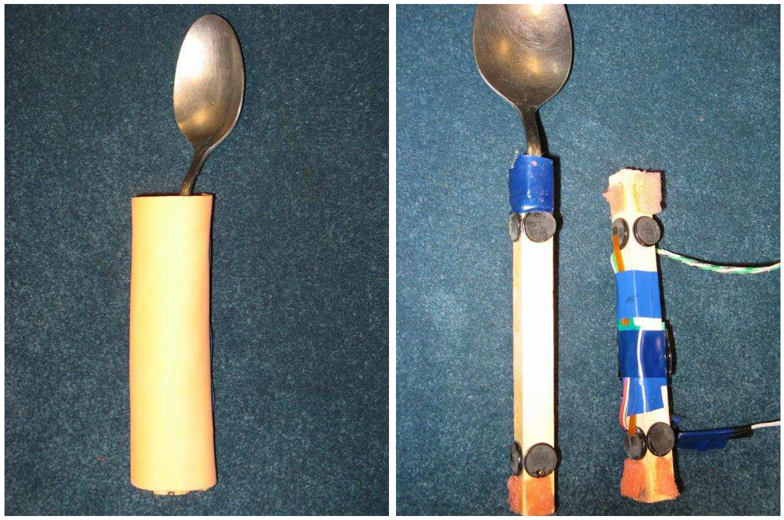


Figure 4.2: Force sensing spoon

point estimates from each arm. One of these points was located at the posterior end of the cylinder fit for the upper arm. The second was located at the anterior end of the lower arm's cylinder, and the final one was located at the point closest to the axes of upper and lower arm cylinders. Points used to estimate upper extremity kinematics from either system are illustrated in Figure 4.1.

Force measurements of the subjects were also made with an instrumented spoon during the performance of one task. The spoon was instrumented with eight 6mm force sensing resistors located at junctures between internal plates. The spoon, and its internal construction, is illustrated in Figure 4.2. It weighed approximately 8 oz, its handle was cylindrical, wooden, covered with a very thin layer of compliant foam, and was 9 inches in length.

All subjects were asked to perform six elements of the Arm Motor Ability Test in the motion capture lab, four of which were selected for the purpose of machine analysis. These four elements are listed in Table 4.1. The other two elements were omitted either because they were bi-manual tasks, or tasks which were frequently occluded across video and motion capture cameras by an assisting human therapist. All elements of the AMAT were guided by a trained Occupational Therapist, and the starting locations for objects were standardized according to AMAT instruc-

AMAT Task/Subtask
“Sandwich” Task
1. Pick up foam “sandwich” 2. “Sandwich” to mouth
“Spoon” Task
3. Pick up spoon 4. Pick up dried kidney bean with spoon 5. Spoon to mouth
“Comb” Task
6. Pick up comb 7. Comb hair
“Jar” Task
8. Grasp jar top 9. Screw jar top open

Table 4.1: AMAT Tasks

tions [132]. Starting positions for subjects required that the torso touch the back of the chair, elbows be at roughly 90 degree angles, and hands be located on designated spots on the counter top.

Subjects performed tasks at the command of an investigator and at a comfortable movement speed. Once they were finished, they were asked to return their hands to their starting locations. While motions were made, data pertaining to the motions were captured with the commercial motion capture system, the video cameras and the force sensing object, if it was used. Measurement devices were synchronized with a button press by the investigator at the start and close of every task.

The video camera angles that were used are illustrated in Figure 2.2. Each camera was approximately 1 meter from the subject, and all cameras focused on the motion of the upper body and hands. Six views were used for the purpose of kinematic reconstruction. Care was taken to provide views that showed the position of the back relative to the chair.

The “verified” AMAT scores that served as ground truth for our automation efforts were the means of those reported in the previous chapter. These scores were generated by three therapists, each of whom was asked to retrospectively review the video of stroke survivors’ AMAT performances in the lab. "True" scores for each subject were said to be the average across this expert group. All of the expert raters were certified and licensed occupational therapists with at least 10 years experience in neuro-rehabilitation for stroke survivors. Two had AMAT specific training, and all had at least basic familiarity with the AMAT. Some demographic information on the experts used to produce "verified" AMAT scores are found in Table 3.4.

4.3 Data Analysis

4.3.1 Extracting Movement Features from Recorded Data

From the recorded data, several movement features were extracted with machines. Some were purely based on kinematics while others were based solely on the force data. A third category of statistics related recovered kinematics and force data. These statistics, as well as the clinical justification for their measurement, follow.

Kinematic Statistics:

1. Movement smoothness. Prior work has shown stroke survivors' point to point arm movements to be jerky relative to the motion of people who never had a stroke [63]. There are many ways to parameterize this jerkiness, however [113]. A common form involves summing the third derivative of the hand's trajectory through space as it moves, as follows [40]:

$$\sum_{t=1}^{T}\left(\frac{d^3x(t)}{dt^3}\right)^2 \tag{4.1}$$

Here, t is a time index, and $x(t)$ is the hand trajectory which has been indexed by parameter t.

An alternative smoothness metric is to compute the number of peaks in a speed profile; this has been done in prior studies of the motion of healthy subjects [19] and stroke survivors [60]. Fewer peaks represent a motion that has fewer periods of acceleration and deceleration, and is therefore relatively smooth.

In our work, we compute smoothness as both the sum of squared jerk AND the number of peaks in the velocity profile of the "end effector". The "end effector" was defined as the point located on subjects' wrists. Before computing any derivatives, however, position data relating to these were smoothed with a median filter of width 5. They were additionally smoothed with a fourth order butterworth filter.

2. Torso displacement. In 1967, Bernstein noted that the human body has numerous and redundant degrees of freedom [15]; any particular functional task can therefore be achieved by several different combinations of motion about joints. Stroke survivors may capitalize on these excess degrees of freedom to compensate for their disabilities. In the upper body in particular, hemiparetic individuals have been found to lean excessively into a reach so as compensate for an impairments at the elbow or shoulder [111].

To measure torso compensation during reaches, we computed the motion of a point on the body that corresponds loosely to the position of the sternum. This point was located equidistant between point estimates for either shoulder. The mean, variance and total displacement of this point was computed as a measure of torso motion during task performances.

3. Use of the unimpaired arm. Extra degrees of freedom are found not only in the torsos, but in the less impaired side of stroke survivors' bodies. Learned compensations that rely heavily on parts of the body less impaired by stroke have been characterized as "learned nonuse" [130, 131] and is commonly seen as a barrier to effective therapy. It has been estimated that 25-30% of stroke survivors exhibit learned nonuse [130].

As a coarse measurement of "learned nonuse" we compute the motion of the less impaired end effector during tasks. More specifically, we compute the mean speed and total distance traveled by the less impaired wrist. Once again, before any derivatives of motion were calculated, data was smoothed with median and butterworth filters.

4. Shoulder abduction and flexion, elbow flexion. Stroke survivors may exhibit stereotyped movement defined by patterned muscle tightness and restricted range of motion around the joints during volitional movement. These patterns of movement are termed muscle synergies and they may be mediated in part through the reflex system [91]. Two typical muscle synergies that influence upper body motion in stroke survivors are the flexor and extensor synergy. The flexor synergy couples external rotation and abduction at the shoulder with elbow flexion. The extensor synergy couples internal rotation and adduction of the shoulder with elbow extension. Some combination of flexor and extensor synergy is combined to form the canonical hemiparetic posture described in [20].

To capture the influence of synergies, we estimate abduction and flexion of the shoulder as well as flexion of the elbow. Flexion of the elbow is computed as the angle of the two vectors that connect the elbow to the the shoulder and wrist. Shoulder angles are slightly more difficult to compute. Flexion is computed by projecting the vector connecting the elbow and shoulder onto the plane defined by the body's midline (the sagittal plane). The angle between this projection and a vector that is normal to the table is said to reflect flexion. Abduction is computed by projecting the arm vector onto the plane defined by the table. The angle between this projection and the vector which connects the two shoulders is said to reflect abduction. Measurements are complicated by the fact that the table top in not orthogonal to the coronal or sagittal plane; this means that measurements of flexion and abduction carry redundant information.

We additionally compute the degree to which motion about the elbow is coupled with abduction at the shoulder. The statistic that we use to measure this "coupling" is mutual information [94]. This is a statistic that has been previously used in graphics research to help characterize "naturalness" in human movement [110]. Mutual information between joints is defined as follows:

$$I(x,y) = \sum_{y \epsilon Y} \sum_{x \epsilon X} p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{4.2}$$

where x represents an angle recorded at the elbow, y is an angle at the shoulder, $p(x, y)$ is the joint probability density function joining elbow and shoulder angles, while $p(x)$ and $p(y)$ are the marginal distributions of angles for either body part. In our research, we estimate probability densities with very coarse histograms spanning 5 bins.

5. Mean and maximum speed. Prior studies of the kinematics of stroke survivors have shown that, during recovery, movements appear to be composed of many short sub-movements. This results in speed profiles that are not only jerky but slow relative to healthy controls. As subjects recover, the speed profiles of stroke survivors tend to feature shallower valleys between velocity peaks, and the mean speed tends to become significantly higher [113]. To be consistent with this prior research, we compute mean and maximum speed of the impaired wrist during movements in an effort to capture the relationship between speed and impairment.

"Movement Arrest Period Ratio" (MAPR) [14] has also been used to reflect recovery after a stroke. This statistic is the proportion of time that the recorded speed of a motion exceeds a given percentage of the peak speed. We compute MAPR as it relates to the movement of the point located at subjects' more impaired wrist.

Force Features

1. Mean, variance and maximum recorded force. Prior studies have shown a dominance of forces produced by the flexors among stroke survivors, which increase as affected muscles are used [23]. An inability to control the extent of flexion contributes to excess grip force commonly measured among stroke survivors during functional motions [48].

To capture force disturbances, we measure the peak, average and variances in forces produced by subjects during object manipulation. All forces are measured in volts, converted to Newtons, and then summed across all the sensors placed on an object. These summed forces are then median and gaussian filtered with filters that are a sixth of a second in width. The gaussian filter's standard deviation is a thirtieth of a second.

2. Smoothness of force. Jerky motion [113] influences not only smoothness in measured kinematics but smoothness in the application of force to objects. To capture this, we measure smoothness of force production. As with the kinematic data, smoothness in force profiles is measured as the sum of squared jerk as well as and the number of force profile peaks.

2. Force MAPR. Finally, as with the kinematic data, we measure the proportion of time that the recorded force exceeds a given percentage of the peak force. This we define a force related "Movement Arrest Period Ratio" (MAPR).

Kinematic and Force Statistics:

1. Distance between the points of peak acceleration and peak force. Healthy individuals modulate their grip forces as the velocity of the objects they are holding increases or decreases. Peak force production in these individuals tends to coincide with the peak acceleration of objects. Stroke survivors may be comparatively slow in modulating force during tasks where objects are moved [96]. Disturbed timing of force corrections has also been found to differ between periods where exerted forces are decreasing rather than increasing [65].

To measure the relationship between acceleration and force modulation, we compute the squared time between peak force production and peak acceleration, as measured with either tracking device.

4.3.2 Predicting AMAT scores

Prediction of AMAT scores requires first **choosing features** to select those that are highly discriminatory, and then **regressing** chosen features onto AMAT scores in an effort to create a function that can predict scores in the absence of a clinician. We then perform an **evaluation** of the constructed regressions using a cross-validation technique.

1. Choosing features. To choose features, we use a correlation based metric. More specifically, we compute the correlation between the mean of AMAT scores assigned by humans (and denoted Y) and sets of kinematic or force features that correspond with these scores. In the equation for correlation which follows, we assume there are i machine generated features for each subject, and each set of a given feature is denoted X_i. $R(i)$ represents the computed correlation between the ith feature set and the AMAT scores produced by humans:

$$R(i) = \frac{cov(X_i, Y)}{\sqrt{var(X_i * var(Y))}} \tag{4.3}$$

To build the best linear regression, we seek features for which correlations strongly depart from zero. These features are the ones that, when evaluated individually, discriminate between scores assigned to the various subjects in the context of a linear model. To quantify the deviation of each computed correlation from zero, we compute the following test statistic, t, for each correlation:

$$t = \frac{R(i) - 0}{SE(R(i))} \tag{4.4}$$

Here, $SE(R(i))$ is the estimate of the standard error in the ith feature's correlation with AMAT scores:

$$SE(R(i)) = \sqrt{\frac{1 - R(i)^2}{n - 2}} \tag{4.5}$$

Computed t statistics are distributed according to t distributions. Those t statistics that lie at the tails of these distributions indicate significant departures from zero, and can be identified by computing associated p-values. Smaller p-values indicate t statistics that are located farther from the center of the distribution.

As explained in great detail in [47], there are several limitations to correlation based criteria for the purpose of choosing features. For one, $R(i)$ can only detect linear dependencies between AMAT scores and input variables. Moreover, as [47] illustrates very clearly, there are instances in which input variables that are completely useless independently can provide performance improvements for the purpose of prediction when considered in conjunction with other features. Alternative feature selection methods involve selecting subsets of features that collaboratively yield strong prediction performance.

In this work, however, we are at a distinct disadvantage in that we have an extremely small set of data with which to explore complex or multi-dimensional relationships between variables. For feature selection, then, we limit ourselves to strictly linear and uni-dimensional methods, and hope for richer data sets with which to explore subsets of feature interactions in the future. In the meantime, we capture feature interactions that the clinical literature suggests to be probable by computing input features that make multi-variate associations, like mutual information between joints or the distance between peaks force production and acceleration.

2. Producing regressions. In the same way that our variable selection methods are linear, so is our prediction methodology. More specifically, we use ridge regressions that have been trained on subsets of recorded data to automatically predict the AMAT scores of stroke survivors. Before employing ridge regression, however, we first normalize computed statistics so as to make each feature comparable with others. This involves centering each feature with respect to the mean for that feature across stroke survivors, and dividing by the feature's standard deviation.

Ridge regression is a form of penalized linear regression. Ordinary linear regressions may suffer when input variables are correlated with one another or collinear; collinear inputs create solutions that are not unique and therefore variable. Ridge regression handles this problem by constraining those coefficients that a regression can produce. This results in better generalization of the regressions to unseen data. More specifically, ridge regression minimizes the following expression:

$$SSE_\lambda = \Sigma_{j=1}^{n}(Y_j - \Sigma_{i=1}^{m} X_{i,j}\beta_j)^2 + \lambda\Sigma_{j=1}^{n}\beta^2 \quad (4.6)$$

In our case, j corresponds to the identification number of a subject and i corresponds to a particular feature; m is the total number of features and n the number of subjects. Xi, j is the value of ith feature for subject j and Y_j is his or her AMAT score. The values for β represent coefficients relating input variables, X, to output variables, Y. The term $\lambda\Sigma_{j=1}^{n}\beta^2$ is a soft constraint on coefficients. It is used to produce solutions that are more numerically stable, less prone to over fitting, and, as a result, typically more capable of generalizing to data outside the training set than ordinary least squares.

3. Evaluation of predictions. In order to evaluate our predictions of scores, we employ a cross validation technique. Cross validation requires that we repeatedly create regressions using a subset of recorded data and test prediction accuracy on data outside of this subset. In our case, we repeatedly use data from six subjects to train a ridge regression function relating motor statistics to functional scores. We then use constructed functions to predict the score of the seventh stroke survivor, and evaluate the accuracy of our prediction against the mean of scores assigned by human experts. As a measure of the accuracy of this prediction, we report the sum of squared error in scores across all subjects.

One of the potential subtleties here is the fact, during a single iteration of cross validation, we construct seven different regression functions to predict scores. Each regression maps six individuals' features onto six AMAT scores and the resulting regression is then used to predict the score of the seventh. Several different sets of regression coefficients are therefore generated to make predictions, which weight and re-weight features differently each time. In this research we are primarily interested in prediction outcomes and not the underlying weighting of features used to make predictions. We therefore shoot for stable predictors, more so than stable weights on sets of feature inputs.

The form of regression that we have chosen to use has two different parameters that impact performance. The first is the p-value used to select promising features. The second is the λ value that is used to constrain ridge regressions. In our results we report first the effects of the p values used to select features. Then, we select p values which minimize the sum of squared prediction error. Finally, we explore the impact of the choice of λ on these optimized predictors.

In the results which follow, we first report on key movement statistics that were found to correlate strongly with functional score. We also look at our ability to measure these statistics accurately with our devices relative to motion capture, and the degree to which all measured statistics depend on their task constraints.

Next, we look at the capacity of measured statistics to predict functional scores. We construct several different regressions, some of which are trained on data from task segments and others which are trained on data from complete tasks. Each regression is evaluated relative to the others, and relative to regressions trained solely with motion capture.

The sections which follow, then, are:

1. An analysis of statistics that correlate with AMAT scores, as computed over complete tasks.

2. An analysis of statistics that correlate with AMAT scores, as computed over task segments.

3. Prediction of AMAT scores, for a range of different regressions.

4.4 Analysis of Complete Tasks

In this section, all statistics were computed across entire task performances, from the initial onset of movements to the end of movements.

4.4.1 Kinematic Correlations with AMAT scores

	Sandwich Mocap	Video		Spoon Mocap	Video	
Torso Statistics						
mean displacement	**0.066** (-0.74)	**0.047** (-0.76)	***	**0.085** (-0.71)	**0.31** (-0.46)	**
sum displacement	**0.092** (-0.71)	**0.094** (-0.69)	***	**0.016** (-0.78)	**0.017** (-0.6)	***
variance displacement	**0.17** (-0.58)	**0.01** (-0.88)	*	**0.22** (-0.54)	**0.42** (-0.37)	
Shoulder Statistics						
mean abduction	**0.49** (0.32)	**0.59** (0.25)		**0.56** (0.27)	**0.51** (0.31)	
variance abduction	**0.54** (0.29)	**0.6** (-0.23)		**0.0076** (0.74)	**0.059** (0.76)	***
mean flexion	**0.17** (0.61)	**0.7** (-0.16)		**0.83** (-0.088)	**0.65** (-0.2)	
variance flexion	**0.33** (0.44)	**0.64** (-0.22)		**0.0047** (0.92)	**0.24** (0.53)	**
Elbow Statistics						
mean flexion	**0.24** (-0.47)	**0.15** (-0.62)		**0.48** (-0.32)	**0.2** (-0.54)	
variance flexion	**0.053** (0.78)	**0.029** (0.82)	***	**0.049** (0.77)	**0.055** (0.77)	***
Inter-joint Statistics						
abduction/elbow info	**0.12** (0.67)	**0.66** (-0.22)		**0.11** (0.65)	**0.26** (0.52)	
flexion/elbow info	**0.52** (0.32)	**0.26** (-0.51)		**0.23** (0.53)	**0.042** (0.79)	*
Speed Statistics						
mean speed	**0.56** (0.28)	**0.67** (0.21)		**0.0052** (0.69)	**0.051** (0.44)	***
max speed	**0.58** (0.23)	**0.78** (0.16)		**0.64** (0.23)	**0.26** (-0.5)	
MAPR	**0.0005** (-0.93)	**0.069** (-0.74)	***	**0.95** (0.014)	**0.74** (0.16)	
length of segment	**0.71** (-0.18)	**0.71** (-0.18)		**0.068** (-0.74)	**0.068** (-0.74)	***
Smoothness Statistics						
jerk	**0.53** (0.29)	**0.75** (-0.14)		**0.46** (-0.34)	**0.25** (-0.52)	
peaks	**0.61** (-0.25)	**0.64** (-0.23)		**0.021** (-0.84)	**0.018** (-0.86)	***
Unimpaired Statistics						
mean other side	**0.5** (-0.32)	**0.76** (-0.16)		**0.041** (-0.79)	**0.2** (-0.56)	**
sum other side	**0.6** (-0.25)	**0.7** (-0.19)		**0.033** (-0.81)	**0.069** (-0.74)	***

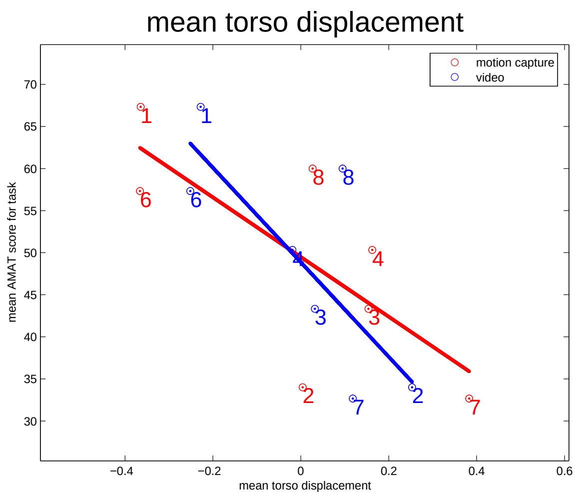
Table 4.2: Measured kinematics for the sandwich and spoon tasks. These statistics were computed over data from entire task performances.

P-values that relate computed kinematic statistics to the mean of expert assigned AMAT scores are reported in Table 4.2 and 4.3. In parentheses next to these p-values are correlations between statistics and AMAT scores. A single asterisk in a column indicates a significant p value was detected at the .1 level for the video based system only. A double asterisk indicates significance at the .1 level for the motion capture based statistic only. A triple asterisk indicates significance at the .1 level for both the motion capture and the video based system.

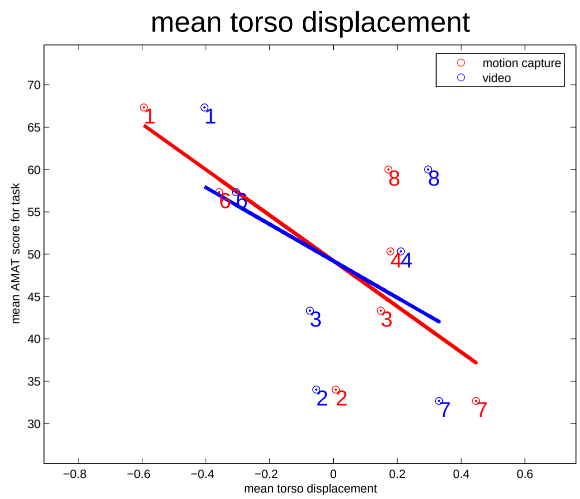
Several kinematic statistics were strongly connected to functional health across all analyzed

	Comb Mocap	Video		Jar Mocap	Video	
Torso Statistics						
mean displacement	**0.16** (-0.61)	**0.25** (-0.53)		**0.06** (-0.75)	**0.033** (-0.81)	***
sum displacement	**0.27** (-0.39)	**0.49** (-0.31)		**0.0084** (-0.9)	**0.0052** (-0.92)	***
variance displacement	**0.17** (-0.61)	**0.22** (-0.56)		**0.27** (-0.49)	**0.099** (-0.7)	*
Shoulder Statistics						
mean abduction	**0.57** (0.26)	**0.82** (0.1)		**0.63** (0.23)	**0.5** (0.32)	
variance abduction	**0.48** (0.33)	**0.64** (-0.22)		**0.31** (0.45)	**0.39** (0.37)	
mean flexion	**0.91** (0.029)	**0.23** (-0.55)		**0.44** (-0.39)	**0.082** (-0.45)	*
variance flexion	**0.52** (-0.31)	**0.78** (0.13)		**0.13** (0.66)	**0.28** (0.48)	
Elbow Statistics						
mean flexion	**0.37** (0.43)	**0.53** (-0.29)		**0.14** (-0.65)	**0.25** (-0.52)	
variance flexion	**0.14** (0.47)	**0.4** (0.39)		**0.28** (-0.48)	**0.28** (-0.49)	
Inter-jointStatistics						
abduction/elbow info	**0.0014** (-0.95)	**0.85** (-0.093)	**	**0.24** (-0.48)	**0.69** (0.18)	
flexion/elbow info	**0.25** (-0.51)	**0.3** (0.47)		**0.81** (-0.1)	**0.64** (0.21)	
Speed Statistics						
mean speed	**0.48** (0.32)	**0.13** (0.63)		**0.29** (-0.48)	**0.49** (-0.33)	
max speed	**0.41** (0.36)	**0.33** (0.45)		**0.23** (-0.54)	**0.21** (-0.51)	
MAPR	**0.21** (0.53)	**0.24** (0.52)		**0.42** (0.31)	**0.22** (0.53)	
length of segment	**0.79** (0.13)	**0.79** (0.13)		**0.085** (-0.71)	**0.085** (-0.71)	***
Smoothness Statistics						
jerk	**0.49** (0.32)	**0.66** (0.22)		**0.27** (-0.51)	**0.55** (-0.29)	
peaks	**0.83** (0.11)	**0.88** (0.078)		**0.28** (-0.5)	**0.057** (-0.76)	*
Unimpaired Statistics						
mean other side	**1** (-0.021)	**0.015** (-0.19)	*	**0.77** (-0.14)	**0.57** (-0.27)	
sum other side	**0.74** (0.13)	**0.56** (-0.18)		**0.026** (-0.83)	**0.078** (-0.68)	***

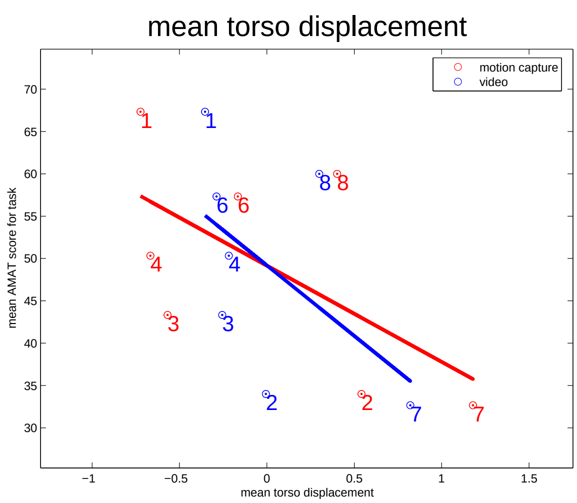
Table 4.3: Measured kinematics for the comb and jar tasks. These statistics were computed over data from entire task performances.



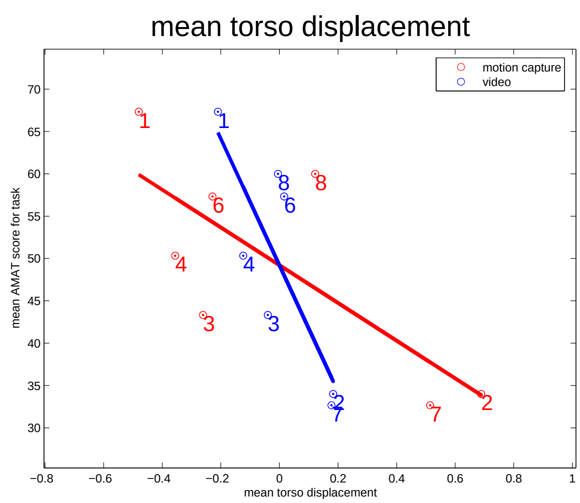
(a) Sandwich



(b) Spoon



(c) Comb



(d) Jar

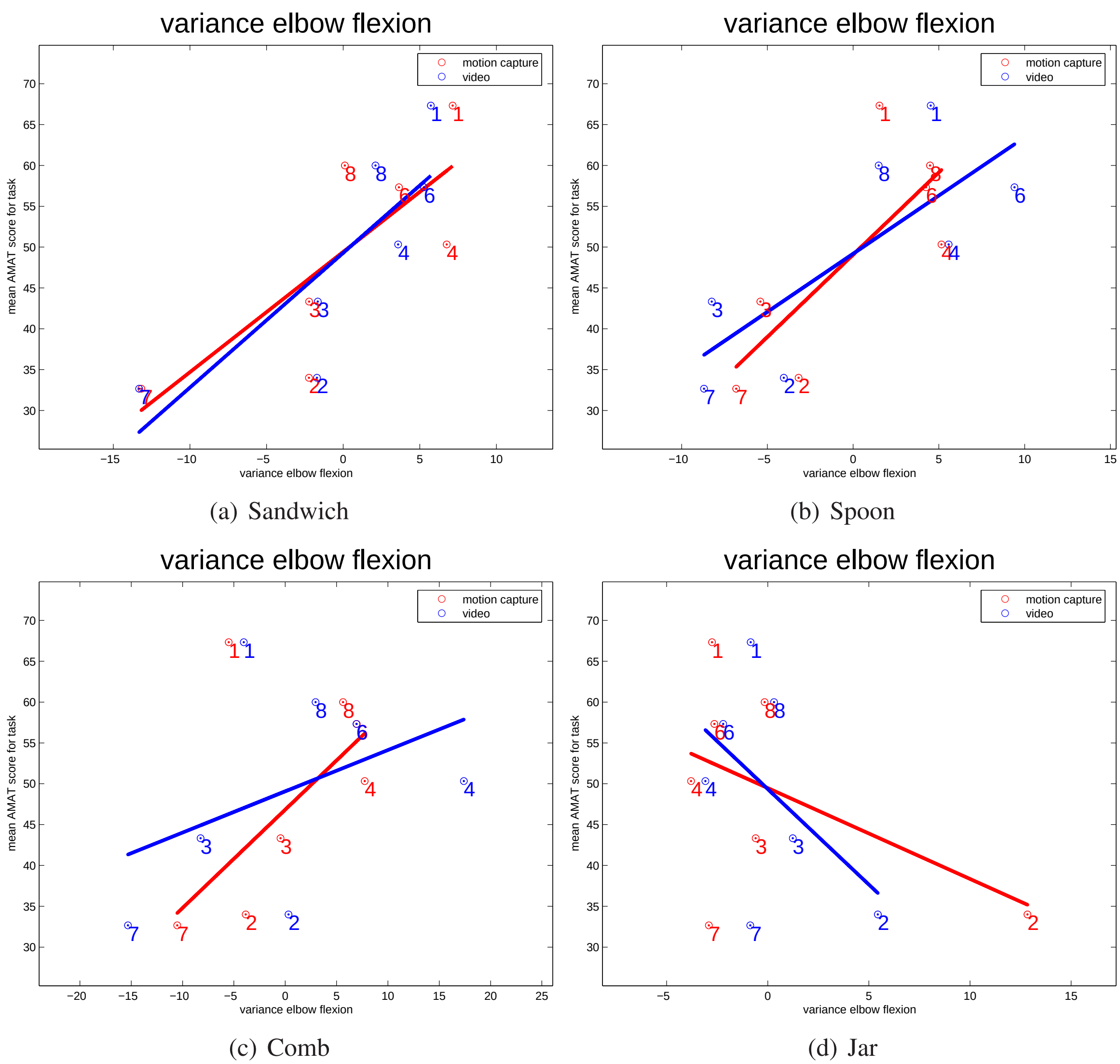
Figure 4.3: Mean torso displacement vs. AMAT score

tasks. These statistics related primarily to motion of the torso and elbow. Across tasks, all of the recorded torso displacement statistics correlated negatively with functional score. These correlations were consistently significant for both the mean and total torso displacements computed by either measurement device. A similar but less statistically significant negative correlation was found to exist between the variance in torso displacement and functional score.

Mean torso displacement for all tasks is illustrated in Figure 4.3. In the graphs, red corresponds to data that was generated by the motion capture based system and blue corresponds to the measurements produced by the system of our design. Every data point is numbered with the identification of the subject to which the data belongs. Data points have also been mean centered. This has been done to facilitate visual matching between measurements from our system and those of the motion capture device. For three of these tasks, the p value relating mean displacement as determined by motion capture to AMAT score indicates significance at the .1 level, and for the remaining task the p value is .16. P values for our system indicate significance for two of the tasks.

Elbow variance was also found to relate strongly to functional score for two of the four tasks. These variances, computed for all tasks, are illustrated in Figure 4.4. For two of the tasks, the motion capture and video based statistics indicate a significant relationship to functional score at the .1 level. For the comb task, this p value is slightly higher, at 0.14. In all cases, a greater variance in the angle recorded at the elbow was associated with a stronger functional score. Mean elbow flexion, by contrast, was frequently negatively associated with functional score. The more acute the recorded angle at the elbow, on average, the lower the AMAT score. This relationship, however, was not found to be significant at the .1 level.

Elbow variance is negatively correlated with functional score for the jar task. The range of recorded variances for this task, however, was more restricted than for others in the battery. Variances in elbow flexion typically ranged between +/- 10 degrees while, for the jar task, this range was reduced by roughly 50%. Mean elbow flexion was still negatively correlated with functional score for both the jar task as well as the others. The p value relating mean elbow flexion and



(a) Sandwich (b) Spoon (c) Comb (d) Jar

Figure 4.4: Variance in elbow flexion by AMAT score.

Statistic	P-value (correlation)
Mean force	0.35 (0.42)
Max force	0.76 (-0.12)
Variance force	0.48 (-0.33)
Jerk	0.95 (0.04)
Peaks	0.12 (-0.77)
MAPR	0.25 (0.53)
force entropy	0.83 (-0.12)
force ratio	0.89 (0.088)
Ratio Difference	0.56 (0.29)

Table 4.4: Force Statistics from the Spoon Task.

AMAT score for the jar task is 0.14, and the correlation is strongly negative, at -0.65.

Finally, the mutual information recorded between abduction at the shoulder and elbow flexion proved to be strongly related to functional score in many tasks. For the comb task, this mutual information is very strongly and negatively associated with functional score. The stronger the link between flexion and and abduction for this task, the lower the functional score. This same negative relationship exists for the jar task as well, while for the other tasks, the motion capture based statistic is positively correlated with score. Significant relationships, however, do not exist for any task other than the comb task.

4.4.2 Force Correlations with AMAT scores

Examples of force profiles from each subject that were recorded with the force sensing spoon are illustrated in Figure 4.5. A table of computed force statistics, and the p values relating these statistics to functional scores, is found in Table 4.4.

The statistic that was found to be most strongly related to functional status was the number of peaks in the recorded force profiles. The more the number of peaks in profiles, on average, the lower the subject's functional score. The p value relating this particular statistic to AMAT score is 0.12, and the correlation is -0.77. The number of peaks as related to AMAT scores is illustrated in Figure 4.6. The statistic for jerk does not correlate with AMAT score.

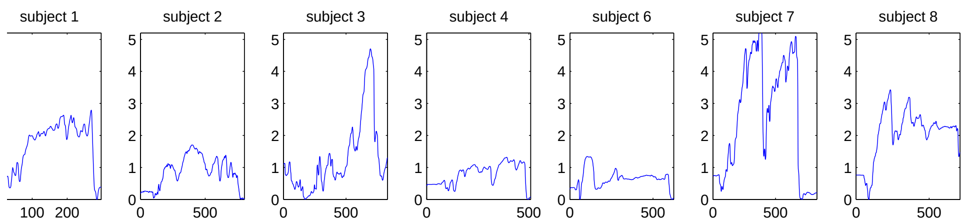


Figure 4.5: Examples of Force Profiles

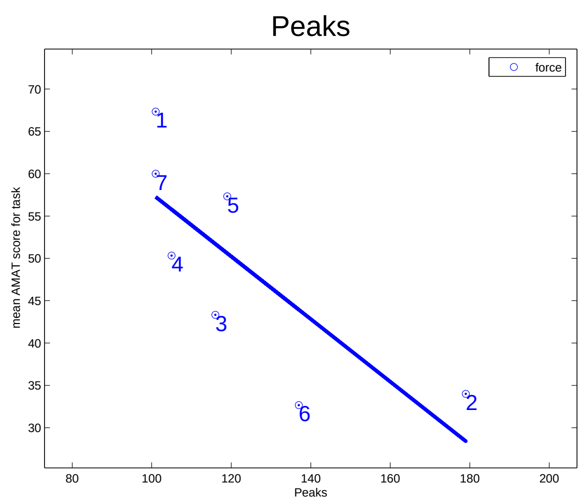


Figure 4.6: Peaks in force profile recorded during the ”Spoon” Task.

AMAT Task/Subtask	Category of Motion
"Sandwich" Task	
1. Pick up foam "sandwich" 2. "Sandwich" to mouth	Grasp Lift
"Spoon" Task	
3. Pick up spoon 4. Pick up dried kidney bean with spoon 5. Spoon to mouth	Grasp Manipulation Lift
"Comb" Task	
6. Pick up comb 7. Comb hair	Gasp Lift
"Jar" Task	
8. Grasp jar top 9. Screw jar top open	Grasp Manipulation

Table 4.5: AMAT segments and the Movement Categories.

The MAPR computed for force was also loosely, and positively, associated with functional score, as was the mean force. None of these associations, however, were significant. Those with higher functional scores therefore had a tendency to produce consistently larger forces than those with lower functional scores.

4.5 Analysis of Task Segments

In an effort to better understand task-specific statistics of motion, elements of the AMAT were decomposed into pieces. Pieces of tasks that shared similar movement requirements were grouped together for this portion of the analysis. All grasps for objects were grouped into one category, all manipulations of objects were grouped into a second category, and lifts of the hand from the table to the mouth or head were grouped into a third category. Task components and the categories of movement to which they were assigned are given in Table 4.5. All data segmentations were executed by hand and based on a review of the video. Results which follow relate to these segments of data and to the identified categories of motion to which they correspond.

4.5.1 Correlations between AMAT scores and grasping statistics

	Sandwich Mocap	Video		Spoon Mocap	Video	
Torso Statistics						
mean displacement	**0.24** (-0.51)	**0.03** (-0.81)	*	**0.1** (-0.68)	**0.29** (-0.46)	
sum displacement	**0.1** (-0.6)	**0.01** (-0.89)	*	**0.026** (-0.83)	**0.14** (-0.62)	**
variance displacement	**0.42** (-0.37)	**0.016** (-0.86)	*	**0.25** (-0.51)	**0.72** (-0.16)	
Shoulder Statistics						
mean abduction	**0.47** (0.34)	**0.61** (0.24)		**0.67** (0.2)	**0.57** (0.26)	
variance abduction	**0.53** (0.3)	**0.3** (-0.46)		**0.094** (0.69)	**0.15** (0.63)	**
mean flexion	**0.28** (0.5)	**0.64** (-0.22)		**0.14** (-0.62)	**0.36** (-0.39)	
variance flexion	**0.39** (0.37)	**0.55** (-0.27)		**0.031** (0.82)	**0.058** (0.76)	***
Elbow Statistics						
mean flexion	**0.057** (-0.72)	**0.11** (-0.68)	**	**0.04** (-0.65)	**0.11** (-0.71)	**
variance flexion	**0.13** (0.65)	**0.4** (0.38)		**0.21** (0.61)	**0.14** (-0.61)	
Inter-joint Statistics						
abduction/elbow info	**0.22** (0.55)	**0.77** (-0.15)		**0.14** (0.64)	**0.091** (-0.7)	*
flexion/elbow info	**0.78** (0.13)	**0.99** (-0.017)		**0.33** (0.44)	**0.97** (0.036)	
Speed Statistics						
mean speed	**0.9** (-0.069)	**0.39** (-0.4)		**0.51** (0.3)	**0.11** (-0.67)	
max speed	**0.74** (0.15)	**0.23** (-0.54)		**0.74** (0.16)	**0.19** (-0.57)	
MAPR	**0.0012** (-0.95)	**0.042** (-0.79)	***	**0.43** (-0.35)	**0.067** (-0.74)	*
length of segment	**0.69** (-0.18)	**0.69** (-0.18)		**0.059** (-0.74)	**0.059** (-0.74)	***
Smoothness Statistics						
jerk	**0.6** (0.25)	**0.073** (-0.74)	*	**0.7** (0.19)	**0.25** (-0.54)	
peaks	**0.69** (0.19)	**0.91** (-0.048)		**0.069** (-0.72)	**0.056** (-0.75)	***
Unimpaired Statistics						
mean other side	**0.74** (-0.16)	**0.43** (-0.38)		**0.22** (-0.55)	**0.6** (-0.25)	
sum other side	**0.74** (-0.16)	**0.41** (-0.39)		**0.018** (-0.85)	**0.41** (-0.37)	**

Table 4.6: Measured kinematics for the sandwich and spoon tasks. These statistics relate to segments of data corresponding to grasps.

P-values that relate computed kinematic statistics for grasping motions to the mean of expert assigned AMAT scores are reported in Table 4.6 and 4.7. As before, correlations between statistics and AMAT scores are reported next to each p value in parentheses. A single asterisk in a column indicates a significant p value at the .1 level for the video statistic, a double asterisk indicates the same for the motion capture statistic and a triple asterisk indicates significance for statistics from both devices.

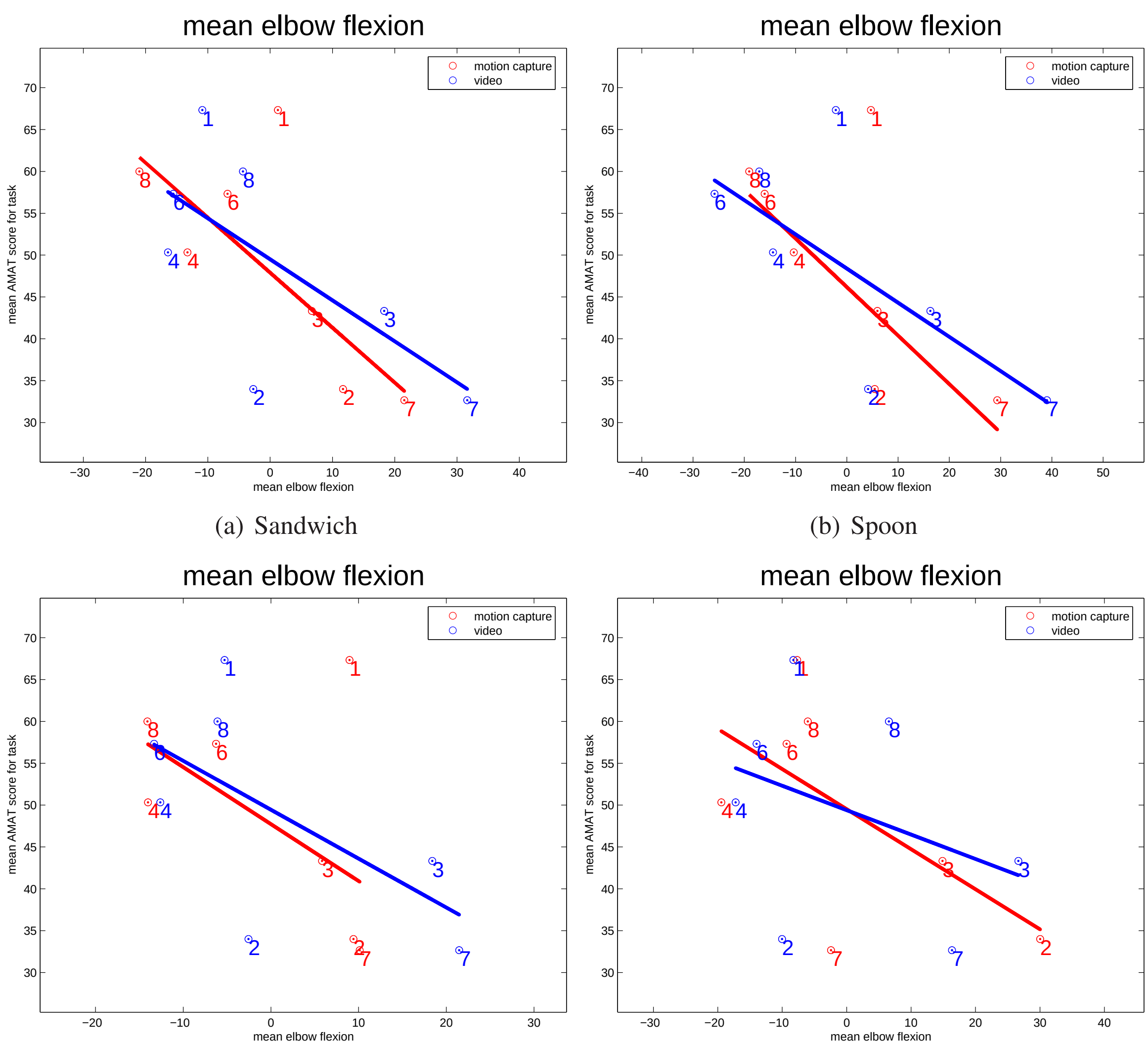
Torso displacement again correlated consistently and negatively was associated with func-

	Comb Mocap	Video		Jar Mocap	Video	
Torso Statistics						
mean displacement	**0.3** (-0.48)	**0.37** (-0.42)		**0.011** (-0.56)	**0.092** (-0.7)	***
sum displacement	**0.25** (-0.5)	**0.28** (-0.45)		**0.024** (-0.82)	**0.056** (-0.77)	***
variance displacement	**0.25** (-0.52)	**0.38** (-0.42)		**0.47** (-0.32)	**0.33** (-0.45)	
Shoulder Statistics						
mean abduction	**0.73** (0.16)	**0.78** (0.13)		**0.69** (0.19)	**0.66** (0.21)	
variance abduction	**0.39** (0.39)	**0.57** (0.26)		**0.18** (0.58)	**0.15** (0.61)	
mean flexion	**0.34** (-0.38)	**0.37** (-0.43)		**0.65** (-0.22)	**0.51** (-0.31)	
variance flexion	**0.68** (0.22)	**0.61** (0.24)		**0.015** (0.86)	**0.08** (0.72)	***
Elbow Statistics						
mean flexion	**0.21** (-0.45)	**0.16** (-0.63)		**0.17** (-0.6)	**0.44** (-0.37)	
variance flexion	**0.89** (0.064)	**0.44** (0.36)		**0.43** (-0.37)	**0.55** (-0.28)	
Inter-joint Statistics						
abduction/elbow info	**0.23** (0.5)	**0.28** (0.5)		**0.52** (0.3)	**0.29** (0.48)	
flexion/elbow info	**0.84** (0.08)	**0.53** (-0.31)		**0.4** (0.39)	**0.92** (-0.04)	
Speed Statistics						
mean speed	**0.39** (-0.35)	**0.88** (0.077)		**0.35** (-0.43)	**0.96** (-0.026)	
max speed	**0.78** (-0.14)	**0.77** (0.13)		**0.26** (-0.5)	**0.37** (-0.4)	
MAPR	**0.47** (0.31)	**0.66** (0.2)		**0.018** (0.32)	**0.22** (-0.49)	**
length of segment	**0.65** (-0.22)	**0.65** (-0.22)		**0.078** (-0.72)	**0.078** (-0.72)	***
Smoothness Statistics						
jerk	**0.58** (0.25)	**0.54** (0.28)		**0.26** (-0.51)	**0.03** (-0.81)	*
peaks	**0.87** (0.066)	**0.45** (-0.35)		**0.57** (-0.26)	**0.046** (-0.78)	*
Unimpaired Statistics						
mean other side	**0.37** (-0.42)	**0.19** (-0.58)		**0.75** (-0.15)	**0.59** (-0.25)	
sum other side	**0.52** (-0.3)	**0.17** (-0.61)		**0.078** (-0.71)	**0.11** (-0.59)	**

Table 4.7: Measured kinematics for the comb and jar tasks. These statistics relate to segments of data corresponding to grasps.

tional score. This significance of this negative correlation was strongest as it related to the total torso displacement measured by both systems. Negative correlations are also found for both the variance and mean displacement. Total displacement as measured by motion capture during grasps was significantly related to AMAT score at the .1 level for three tasks, and this level of significance was achieved for the video based statistics for two tasks.

More interestingly, perhaps, the mean elbow flexion recorded during grasps emerged as a



(a) Sandwich (b) Spoon (c) Comb (d) Jar

Figure 4.7: Mean elbow flexion by AMAT score during Grasps.

strong indicator of functional health in this subset of data. Significant correlations between the average flexion of the elbow and score were found to exist at the .1 level for two of the four tasks. No such significant correlations were detected during the analysis of entire task data. Figure 4.7 illustrates the mean flexion, as recorded by both devices, recorded during grasps. As before, red corresponds to motion capture data and blue corresponds to the data from our system. Data points are once again numbered to facilitate visual matching between measurements.

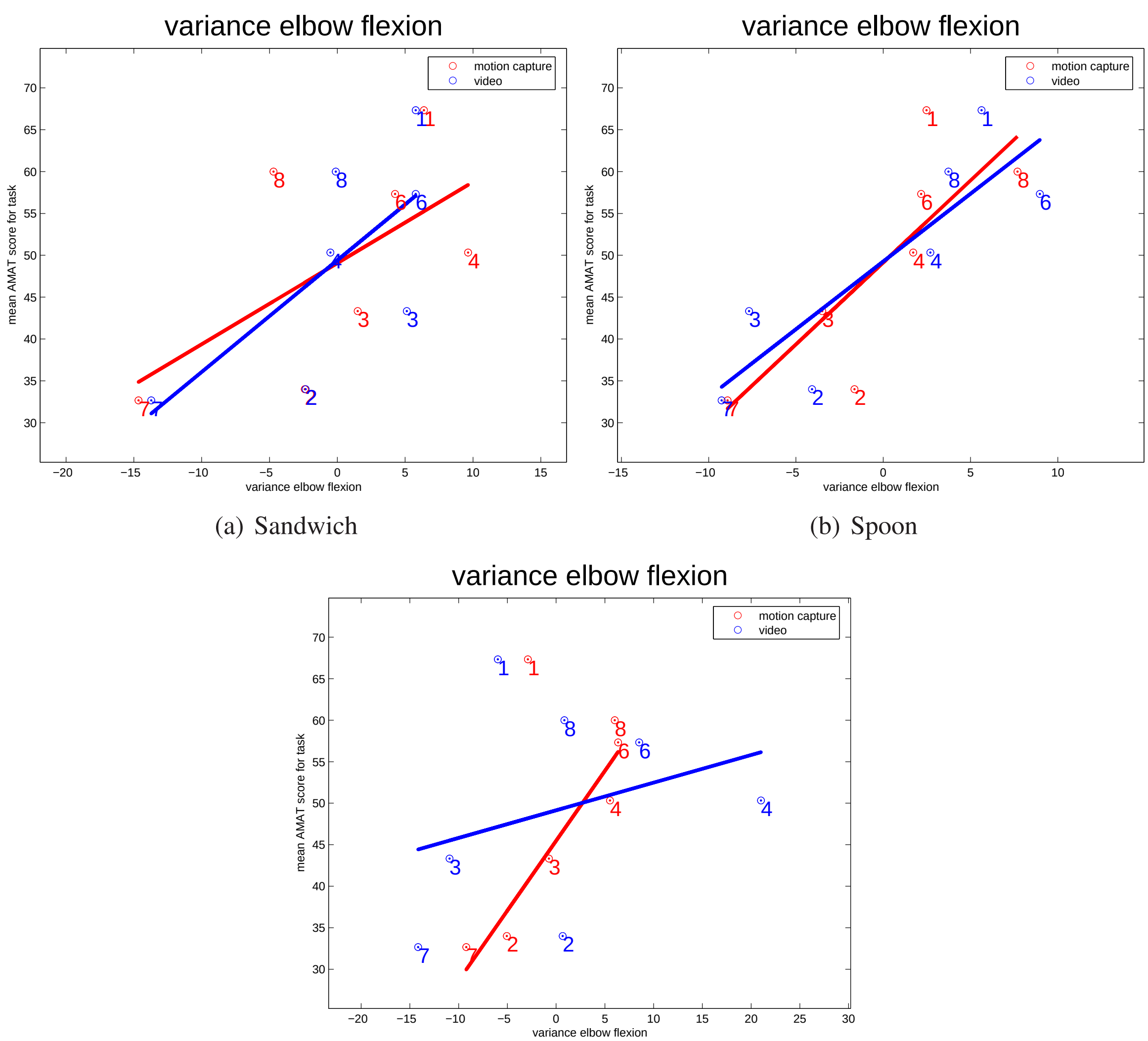
4.5.2 Correlations with AMAT scores for Periods of Lifting

	Sandwich **Mocap**	**Video**		**Spoon** **Mocap**	**Video**	
Torso Statistics						
mean displacement	**0.24** (-0.45)	**0.38** (-0.4)		**0.012** (-0.7)	**0.22** (-0.54)	**
sum displacement	**0.33** (-0.42)	**0.62** (-0.23)		**0.26** (-0.5)	**0.33** (-0.44)	
variance displacement	**0.52** (-0.3)	**0.67** (-0.2)		**0.57** (-0.26)	**0.36** (-0.43)	
Shoulder Statistics						
mean abduction	**0.52** (0.3)	**0.56** (0.28)		**0.61** (0.24)	**0.58** (0.26)	
variance abduction	**0.41** (0.38)	**0.23** (0.54)		**0.59** (0.25)	**0.44** (0.36)	
mean flexion	**0.036** (0.56)	**0.98** (-0.018)	**	**0.83** (0.1)	**0.8** (-0.1)	
variance flexion	**0.36** (0.4)	**0.31** (0.46)		**0.53** (0.3)	**0.76** (0.15)	
Elbow Statistics						
mean flexion	**0.66** (-0.21)	**0.51** (-0.29)		**1** (-0.0017)	**0.44** (-0.36)	
variance flexion	**0.2** (0.59)	**0.096** (0.7)	*	**0.039** (0.81)	**0.017** (0.86)	***
Interjoint Statistics						
abduction/elbow info	**0.95** (0.038)	**0.72** (-0.17)		**0.75** (0.17)	**0.093** (0.68)	*
flexion/elbow info	**0.73** (0.15)	**0.59** (0.26)		**0.079** (0.73)	**0.051** (0.76)	***
Speed Statistics						
mean speed	**0.28** (0.49)	**0.32** (0.45)		**0.22** (0.57)	**0.2** (0.56)	
max speed	**0.72** (0.21)	**0.56** (0.29)		**0.38** (0.43)	**0.87** (0.082)	
mapr	**0.27** (0.5)	**0.37** (-0.41)		**0.65** (0.22)	**0.84** (-0.097)	
length of segment	**0.89** (-0.053)	**0.89** (-0.053)		**0.56** (-0.27)	**0.56** (-0.27)	
Smoothness Statistics						
jerk	**0.55** (-0.28)	**0.91** (-0.036)		**0.94** (-0.039)	**0.25** (-0.52)	
peaks	**0.22** (-0.5)	**0.48** (-0.31)		**0.18** (-0.58)	**0.23** (-0.52)	
Unimpaired Statistics						
mean other side	**0.12** (-0.68)	**0.96** (0.029)		**0.045** (-0.79)	**0.14** (-0.63)	**
sum other side	**0.1** (-0.69)	**0.99** (-0.0033)		**0.13** (-0.66)	**0.18** (-0.58)	

Table 4.8: Measured kinematics for the sandwich and spoon tasks. These statistics all relate to data recorded during segments corresponding to lifts.

	Comb Mocap	Video	
Torso Statistics			
mean displacement	**0.14** (-0.64)	**0.3** (-0.47)	
sum displacement	**0.45** (-0.28)	**0.7** (-0.18)	
variance displacement	**0.13** (-0.65)	**0.16** (-0.62)	
Shoulder Statistics			
mean abduction	**0.45** (0.34)	**0.78** (0.13)	
variance abduction	**0.4** (0.39)	**0.68** (-0.2)	
mean flexion	**0.2** (0.41)	**0.41** (-0.39)	
variance flexion	**0.95** (0.032)	**0.83** (0.1)	
Elbow Statistics			
mean flexion	**0.37** (0.41)	**0.36** (-0.45)	
variance flexion	**0.022** (0.62)	**0.53** (0.29)	**
Interjoint Statistics			
abduction/elbow info	**0.038** (-0.79)	**0.45** (0.32)	**
flexion/elbow info	**0.58** (-0.27)	**0.43** (0.32)	
Speed Statistics			
mean speed	**0.63** (0.22)	**0.34** (0.43)	
max speed	**0.32** (0.43)	**0.33** (0.45)	
mapr	**0.22** (0.54)	**0.59** (0.27)	
length of segment	**0.64** (0.22)	**0.64** (0.22)	
Smoothness Statistics			
jerk	**0.46** (0.34)	**0.66** (0.22)	
peaks	**0.84** (0.1)	**0.71** (0.19)	
Unimpaired Statistics			
mean other side	**1** (-0.0085)	**0.78** (-0.066)	
sum other side	**0.57** (0.25)	**0.93** (0.082)	

Table 4.9: Measured kinematics for the comb task. These statistics all relate to data recorded during segments corresponding to lifts.

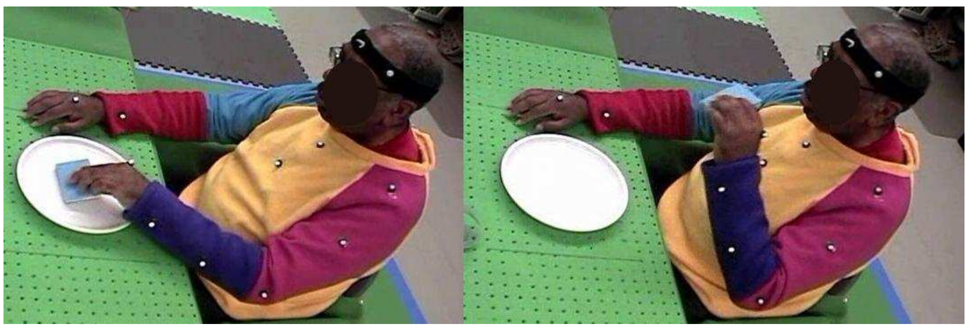


(a) Sandwich

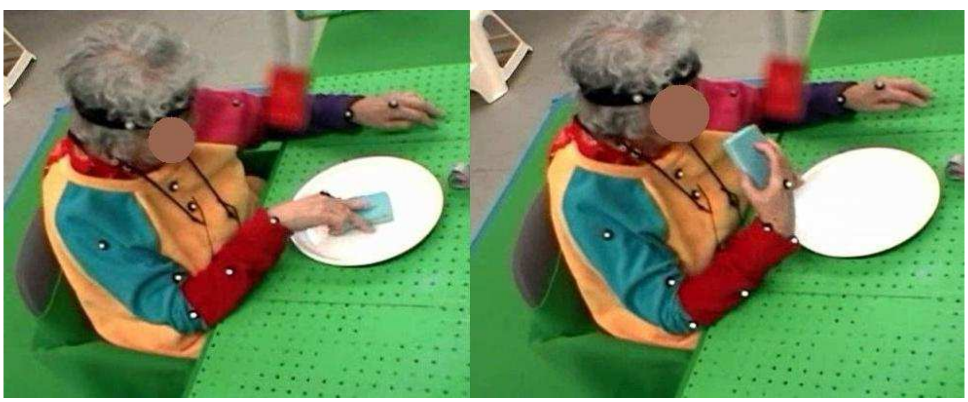
(b) Spoon

(c) Comb

Figure 4.8: Variance in elbow flexion by AMAT score during Lifts.



(a) High scoring subject



(b) Low scoring subject

Figure 4.9: Lifts performed by two subjects, one with a high AMAT score and one a low one.

P-values for statistics recorded during periods of lifting are reported in Tables 4.8 and 4.9.

During segments that correspond to lifts of the hand to the mouth or head, variances in motion about joints were positively correlated with functional score. The most significant of these correlations related to the range of motion measured about the elbow. Motion capture based p-values reflecting the strength of the relationship between elbow range and AMAT score were below 0.2 for all tasks and below 0.1 for two of the three tasks in the category. The same is true for statistics recorded by the device of our design. Elbow range of motion, as measured by both systems, is reported across all tasks in Figure 4.8. Figure 4.9 illustrates an individual with a high AMAT score performing one of the recorded lifts as well as a low scoring individual performing the same lift. The difference in variance around the elbow between these subjects is visible.

Positive correlations also were found to exist between variances in shoulder abduction and functional score. Correlations between this statistic and AMAT score were consistently found to be positive for data from the motion capture system, but the association is not significant.

Smoothness was also a more significant feature during lifts. Recorded jerk based on motion capture and our system was negatively and significantly related to functional score for the spoon task. The peaks statistic was also negatively and strongly related to functional score for this task. Strong negative correlations with smoothness metrics also exist for the sandwich task.

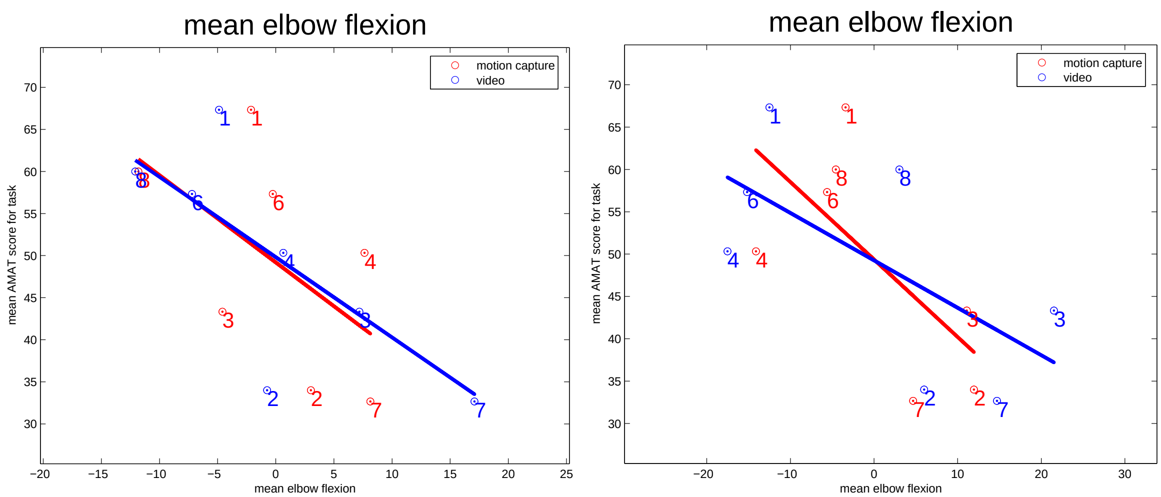
4.5.3 Correlations with AMAT scores for Manipulations of Objects

During the two manipulation segments, statistics relating to a static body posture once again came into play. Individuals with higher functional scores tended to lean in less towards the objects that were being manipulated, and they tended to extend at the elbow and abduct at the shoulder more greatly. Recorded flexion, by contrast, was found to be negatively correlated with functional status for periods of manipulation. The strength of this correlation is especially great for the spoon task. Computed statistics for this portion of tasks are given in Table 4.10.

The motion capture based p value relating mean elbow flexion to functional score is .06 for the

	Spoon Mocap	Video		Jar Mocap	Video	
Torso Statistics						
mean displacement	**0.18** (-0.59)	**0.67** (-0.22)		**0.035** (-0.8)	**0.051** (-0.77)	***
sum displacement	**0.089** (-0.7)	**0.54** (-0.29)	**	**0.019** (-0.85)	**0.071** (-0.74)	***
variance displacement	**0.51** (-0.32)	**0.71** (0.17)		**0.088** (-0.72)	**0.055** (-0.77)	***
Shoulder Statistics						
mean abduction	**0.71** (0.17)	**0.71** (0.17)		**0.64** (0.22)	**0.4** (0.41)	
variance abduction	**0.16** (0.6)	**0.13** (0.64)		**0.4** (0.38)	**0.26** (0.53)	
mean flexion	**0.91** (-0.044)	**0.65** (-0.21)		**0.34** (-0.44)	**0.055** (-0.46)	*
variance flexion	**0.24** (0.54)	**0.084** (0.46)	*	**0.42** (-0.37)	**0.57** (-0.28)	
Elbow Statistics						
mean flexion	**0.055** (-0.76)	**0.001** (-0.88)	***	**0.1** (-0.68)	**0.12** (-0.67)	
variance flexion	**0.22** (0.55)	**0.28** (0.5)		**0.7** (0.15)	**0.83** (-0.098)	
Inter-joint Statistics						
abduction/elbow info	**0.5** (0.33)	**0.68** (0.19)		**0.84** (0.073)	**0.38** (0.4)	
flexion/elbow info	**0.71** (0.19)	**0.9** (0.067)		**0.43** (-0.38)	**0.66** (0.21)	
Speed Statistics						
mean speed	**0.79** (-0.12)	**0.011** (0.65)	*	**0.019** (0.48)	**0.43** (0.39)	**
max speed	**0.12** (0.29)	**0.36** (0.47)		**0.046** (-0.78)	**0.47** (-0.36)	**
MAPR	**0.38** (-0.41)	**0.17** (0.62)		**0.1** (0.69)	**0.57** (0.29)	
length of segment						
Smoothness Statistics						
jerk	**0.011** (-0.88)	**0.16** (-0.6)	**	**0.9** (-0.049)	**0.44** (-0.37)	
peaks	**0.087** (-0.69)	**0.065** (-0.75)	***	**0.29** (-0.48)	**0.18** (-0.59)	
Unimpaired Statistics						
mean other side	**0.058** (0.76)	**0.48** (0.36)	**	**0.5** (-0.33)	**0.57** (0.27)	
sum other side	**0.37** (0.32)	**0.16** (-0.61)		**0.16** (-0.67)	**0.79** (0.14)	

Table 4.10: Measured kinematics for manipulations (i.e. getting bean, opening jar).



(a) Spoon (b) Jar

Figure 4.10: Mean elbow flexion by AMAT score during Manipulations.

	Grasping		Manipulation		Lifting	
Statistic	P-value		P-value		P-value	
Mean force	0.43 (0.37)		0.91 (-0.041)		0.35 (0.42)	
Max force	0.78 (0.13)		0.86 (-0.08)		0.76 (-0.12)	
Variance force	0.062 (0.77)	*	0.57 (0.23)		0.48 (-0.35)	
Jerk	0.77 (0.14)		0.96 (-0.03)		0.95 (-0.04)	
Peaks	0.01 (-0.76)	*	0.47 (-0.32)		0.12 (-0.77)	
MAPR	0.037 (-0.73)	*	0.19 (-0.59)		0.25 (0.53)	
Ratio Difference	0.79 (0.12)		0.19 (-0.55)		0.92 (-0.037)	

Table 4.11: Force Statistics from the Spoon Task.

spoon task and .1 for the jar task. Elbow flexion statistics from both systems are illustrated in Figure 4.10.

Finally, both the jerk statistic and the number of peaks in the velocity profile proved to be a strong indicator of functional health for the spoon task. The peaks statistic proved to be a stronger indicator of health than the jerk statistic for the jar task. All correlations are negative; the more jerk the lower the functional score.

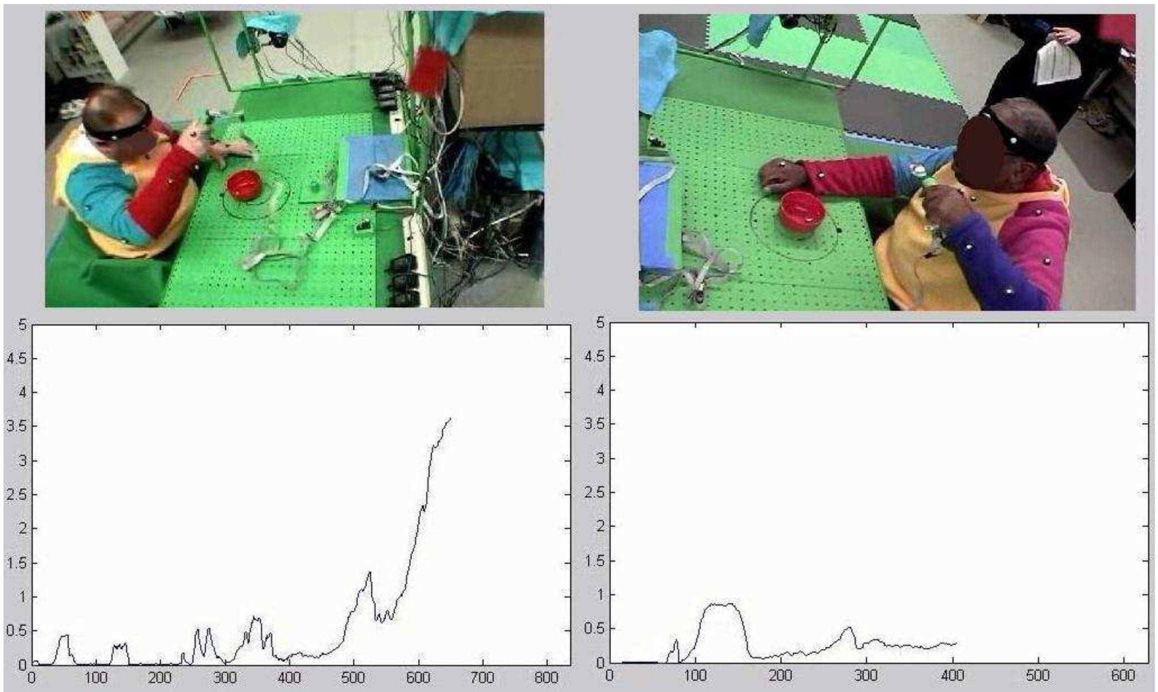
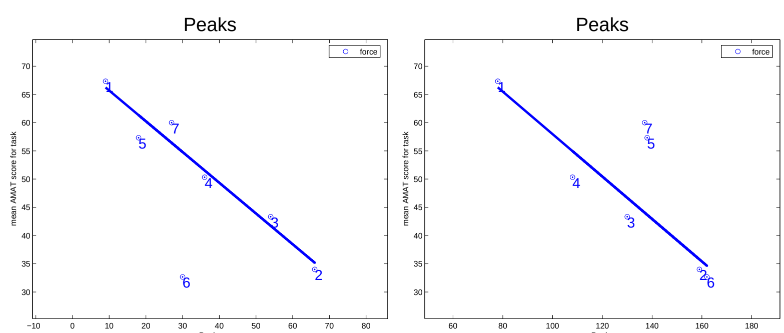
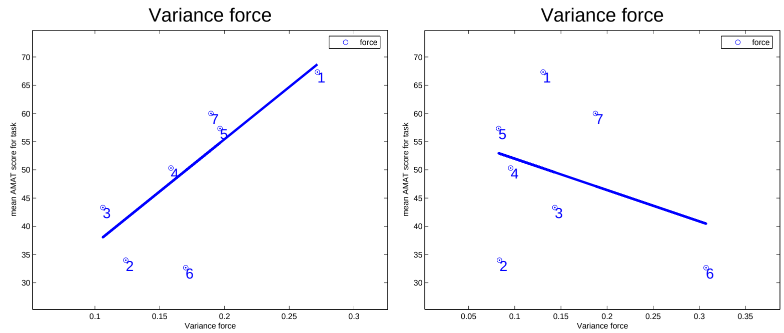


Figure 4.11: A low and high scoring individual using the force sensing spoon.



(a) Grasp (b) Lift

Figure 4.12: Peaks in Force Profile by AMAT score during Grasp and Lift.



(a) Grasp (b) Lift

Figure 4.13: Variance in Force Profile by AMAT score during Grasp and Lift.

4.5.4 Force Correlations with AMAT scores

Force statistics that were associated with functional score during the spoon task also varied from segment to segment. During several segments the number of peaks in the force profile was a strong indicator of functional health. The strength of the association, however, varied from segment to segment. During the grasp as well as the lift portions, p values for the association were below 0.15. In all cases, the correlation with score is negative. The variance in force was also an indicator of functional health during the grasping portion of the movement. The direction of the correlation, however, changes between the grasp and the lift. During the lift, variance was greater in individuals with lower functional scores than in those with higher functional scores. This is not the case during the grasp. The complete list of p values for computed statistics and associated correlations is given in Table 4.11.

Examples of force profiles recorded from individuals are illustrated in Figure 4.11. At the left is an individual with a low functional score and at the right an individual with a high functional score. The variation in force produced by the low functioning individual is visibly larger during the lift portion of the task than that for the high functioning one. Force gets large as this low scoring individual raises the arm to the head; for the high scoring individual it remains relatively constant. Almost the reverse situation, however, exists during the grasping portion of the task.

The number of peaks in the force profiles during two task segments is illustrated in Figure

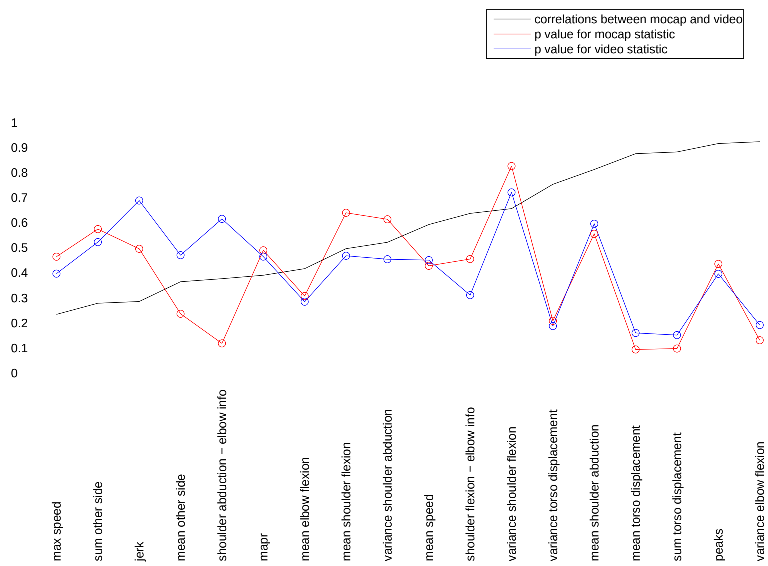
4.12. For two of these task segments, the association between peaks and score is below .15; the more peaks the lower the AMAT score. The variation in force during these same task segments is illustrated in Figure 4.13. There is a different correlation between variation and functional score during either task segment.

4.5.5 Kinematic Accuracy

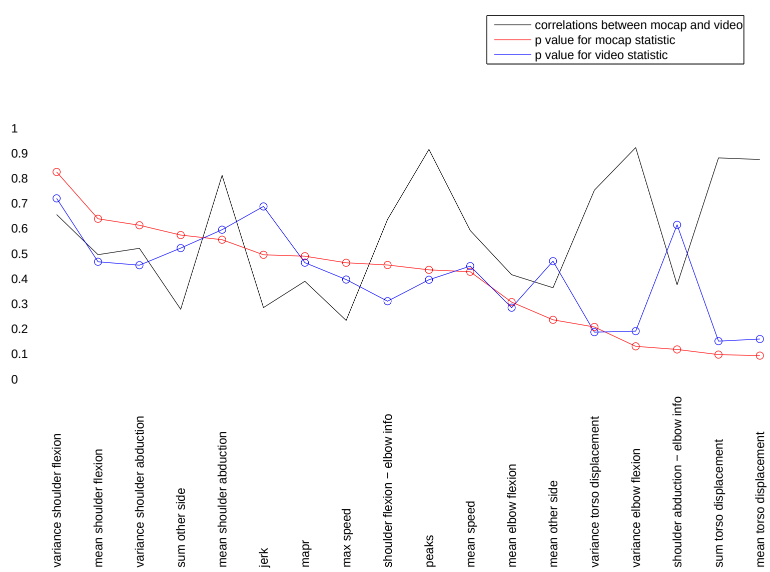
When we optimize the accuracy of our measurement devices we focus on statistics that prove to be most salient indicators of functional health. Figure 4.14 illustrates the relationship between the information each computed statistic carries about functional status and our ability to measure this statistic accurately relative to a motion capture device. In the figure, black lines reflect correlations between the statistics as measured by the different kinematic tracking devices. Red lines represent the p-values which relate statistics, as determined by motion capture, to expert assigned functional scores. Blue lines are the p-values from our kinematic measurement system. All correlations as well as p-values have been averaged across the four tasks in the battery.

At the top portion of the figure, computed statistics are organized by increasing correlation between the measurement devices. At the bottom, statistics are organized according to decreasing p-values for motion capture based statistics. We would ideally like to see, when p-values for the motion capture system are low, that correlations between our system and motion capture are high. Conversely, when p-values are not significant, we can more easily afford to find correlations between measurement devices that are low.

Figure 4.15 points to several informative statistics that prove to be difficult to capture with the device we have designed. Perhaps the most notable of these is the computed mutual information relating elbow flexion and shoulder abduction. The average motion capture-based p-value for this statistic is below .2, yet the correlation between motion capture and our system is less than .5. It should be noted that this mutual information statistic combines two kinematic signals, each of which carries its own noise. One signal is the abduction computed at the shoulder and the



(a) By ascending correlation



(b) By descending p-value

Figure 4.14: Correlations between devices and associated p-values for kinematic statistics.

other is flexion computed at the elbow. Abduction is captured poorly relative to motion capture, as the correlation here is just over .5. Correlations between elbow variances computed by either system are comparatively strong, and over .9 consistently. Shoulder abduction suffers in that it is not computed in a body centered coordinate frame, and is coupled with shoulder flexion.

Similarly, the mean speed of the hand and wrist computed on the less impaired side of the body was found to carry diagnostic information and yet was measured relatively poorly with our device. P values for this statistic are below .3 on average, and yet measurement correlations are roughly .5. There is, however, not much motion on the unimpaired side of the body, on average, during the performance of tasks.

Finally, the computed smoothness statistics also proved to be difficult to capture accurately, relative to motion capture. The number of peaks proved to carry more diagnostic information than computed jerk for the particular tasks.

4.5.6 Discussion

In this section, we first discuss force and kinematic statistics and their associations with AMAT score. We then look at the task specificity of these statistics, and finally we evaluate the accuracy of the kinematic measurement system of our design relative to motion capture.

1. Kinematic statistics and their relationship to functional scores. Some kinematic statistics are strong indicators of functional status across entire tasks, while others relate better to functional score when associated with task segments. Many segments, moreover, exhibit kinematics that are comparable across tasks, while the kinematics of other segments are less comparable across tasks.

The motion of the torso and elbow both proved to correlate consistently and strongly with functional health across all tasks. Prior research has indicated the movement of the torso to be strongly related both to functional health and stroke severity [74, 89, 90]. In [89], for example, stroke survivors were found to recruit more torso motion during reaches to objects than controls, while several more distal features of movement, like the size and timing of grip aperture, remained relatively unaffected. Our results are consistent with these prior results, and illustrate that torso motion can serve as a robust indicator of functional across a wide variety of tasks.

Stroke survivors also commonly exhibit excess elbow flexion and flexion that is more strongly coupled with shoulder abduction than people who have never had a stroke. Research with robot arms has provided data demonstrating abnormal coupling of elbow flexion and shoulder abduction, which results in a work area for stroke survivors that is reduced relative to people that have never had a stroke [**?**]. Research with robots has also measured constraint forces that are consistent with the flexor synergy, which couples elbow flexion and shoulder abduction during reaching tasks [108]. The precise statistics used to measure joint coupling, however, varies from paper to paper. In [74], for example, inter-joint coordination is measured in terms of the smoothness of motion in an inter-joint space. Mutual information, by contrast, is used to quantify joint synergies for the purpose of naturalistic human animation in [110].

The salience of both elbow and torso motion, however, was found to depend on the particular portion of tasks being examined. In general, during periods of grasping and manipulation, statistics relating to a static gross posture were more indicative of functional health than during periods where the gross upper body was set into motion. During both grasps and manipulations, for example, individuals who were more impaired tended to lean more extensively into the table and flex more at the elbow. During lifts, by contrast, variances about the joints were diagnostic. Elbow variance was the most significantly informative of measured statistics during lifts, although variances in shoulder statistics also proved to be consistently and positively correlated with AMAT score.

Smoothness statistics were only found to be of significance during periods of object manipulation; a weaker association was found during periods of object lifting. Prior research has laid much emphasis on smoothness as a means by which to measure the recovery of stroke patients. In [113], for example, five statistics of smoothness were measured with a robotic device in the hemiparetic arm of 31 individuals recovering from stroke. Four of the five metrics showed increases as recovery occurred. The authors postulate that increases in smoothness are the result of better control over sub-movement blending. Krebs et al. [63] also report that movements made by patients recovering from stroke become smoother as recovery proceeds. In our study, however, smoothness disturbances are most manifest during periods where individuals are required to hold their bodies relatively steady.

Our data, however, may suffer in that it is relatively low fidelity. High fidelity measurements of measurements made of Huntington's patients, for example, have shown these individuals to exhibit jerk manifesting 200 to 300 ms after a movement's onset [125]. 200ms, however, represents only 5 data samples when recording at our system's rate of 30 Hz. These points are reduced to 4 when computing the number of peaks in a profile and 2 when computing jerk. Our ability to detect subtle tremors is therefore very limited.

Another kinematic statistics that revealed a strong dependence on task constraints is the mutual information computed between joints. Mutual information was only diagnostic during the comb task, and most significantly during the period where the comb was lifted to the back of the head. Other lifts in the "lifting" category did not elicit informative amounts of shoulder abduction. The lift in the comb task, then, differs significantly in its kinematics relating to the shoulder from the other tasks in the battery.

2. Force statistics and their relationship to functional scores. Like the kinematic statistics, several force measurements proved to be strong indicators of functional status across entire tasks, while others related better to functional score when associated with task segments.

Smoothness of force proved to be the strongest indicator of functional health in our study. This is consistent with results reported by [66], who demonstrated that, in the context of a force tracking task, decreases in force smoothness could be associated with motor recovery due to rehabilitation. In our work, smoothness is best measured by counting the number of peaks in the recorded force profile. The jerk statistic, by comparison, produces associations that are less consistently significant.

This smoothness statistic, however, is not an equally strong indicator of functional health for all task segments. During grasps and lifts, for example, the number of peaks in the force profile is a better indicator of health than during the period of manipulation .

Prior research has indicated stroke survivors to exhibit lower peak grip forces, on average, than healthy controls [16, 75]. These same individuals, however, may exert more force than is necessary during object manipulations [96]. Grip strength, moreover, has been shown to be a very sensitive indicator of initial limb recovery in individuals who are acute stroke survivors [5] and it predicts later recovery fairly accurately as well [6]. Our results, however, failed to show a statistically consistent and significant relationship between mean and maximum forces produced by subjects and functional score. The mean force applied to the spoon handle is an indicator of health during the grasps of the object, but not at any other time. Moreover, mean force here is positively correlated with score, not negatively.

In general it appears that, in our study, individuals who were less impaired tended to produce more force on the handle of the object, and they maintained this force fairly consistently throughout the task. More impaired individuals, by contrast, use less force on average, but there were moments of time during the task where they applied much more force than may have been required. The range of force that is produced varies substantially among the more impaired individuals during manipulation and lift. This lack of control of the force profile is not only manifest in the variance of forces recorded by the handle, but also in the number of peaks in the resulting force profile.

3. Task Specificity of Measured Statistics. Results illustrate the task-specificity of discriminating kinematics. In the three tasks that require lifts to the mouth with the hand, for example, elbow variance positively relates to functional health. For the fourth task, this variance negatively relates to functional health. This fourth task, however, which involves opening a jar on the desktop, does not require the same extent of elbow motion as do the other tasks. Similarly, coupling between elbow flexion and shoulder abduction was shown to carry diagnostic information for the comb task in a way that it is not for the others. This task is the only task that requires individuals to lift their arms over the head and to touch the back of their heads with a hand. It necessarily elicits abduction in a way the other tasks do not.

The influence of task constraints on the diagnostic utility of kinematics after stroke is well represented in the literature. In [89], for example, limitations in range about the shoulder and elbow were more likely to discriminate stroke survivors from healthy controls when measured during reaches to targets that were far from the body. Reaches to locations at half an arm's length from the body were comparatively less likely to yield discriminating statistics. Task and location constraints have been similarly emphasized in [149], wherein motor impaired individuals were shown to reach more smoothly to real coins than to a non-coins. Several other studies have shown motor impaired individuals to use decidedly different movement strategies in contexts where they are pantomiming functional tasks than in environments where they execute tasks for real [55, 85]. It is not a surprise that diagnostic kinematics depend on the tasks to which they relate.

Despite the variations in tasks, however, many task elements were found to share comparable kinematics. All of the lifts in the task, for example, produced elbow variances that discriminated between subjects. Even though one lift is of a slightly different nature than the others, as it requires abduction, it is still comparable to the other lifts in the battery. These common kinematics give us reason to believe that it may be possible to determine functional health based on the perception of task segments, like lifts, irrespective of the tasks to which they belong. This is an idea we will explore further when we predict scores.

4. Accuracy of Kinematic Statistics Relative to Motion Capture. Not only do our results point to salient kinematics, they identify statistics that are relatively difficult or easy to approximate with a low cost tracking device like that which we have designed. Our device performs accurately relative to a VICON when it comes to capturing the salience of torso motion and elbow flexion. Capturing shoulder abduction and smoothness statistics, by comparison, prove to be difficult given our low cost devices.

Because our measurements of shoulder flexion and abduction are not in body centered coordinates, there is some redundancy between the estimates of shoulder flexion and abduction. Moreover, for at least one subject (number 7), the fit of the jersey was loose at the shoulder. The proximal ends of cylinders approximating the upper arms of this subject were biased towards midline as a result. Finally, for the task involving the comb in particular, the field of view of the cameras is too small to capture shoulder abduction sufficiently. This task requires individuals to reach above and behind their heads, yet there no cameras were placed behind subjects. At maximum abduction for many subjects, then, the arm disappears from view. Video camera based reconstructions suffer as a result.

4.6 Predicting Scores

In this section, we evaluate the ability of statistics with low p-values to predict scores of stroke survivors in the absence of a clinician. We test several different paradigms.

First, we train one regression for each task with data that was generated across the entire task performance. We call this Predictor Number 1.

Second, we train a regression for each part of each task. One is trained to predict scores on the lift segment of the sandwich task, for example, and a second is trained on data from the grasp segment for this same task. Automatic scores are defined as the sum of predicted scores across all segments for a given task. We call this Predictor Number 2.

Third, we explore the ability to create regressions that generalize across tasks. We train one regression on all the kinematic data from the lifts, a second on data from all grasps, and a third on the data from manipulations. We then use resulting functions to predict the scores for lifts on all tasks, as well as scores for grasps and manipulations. As before, automatic scores are defined as the sum of predicted scores across all segments for a given task. We call this Predictor Number 3.

4.6.1 Prediction with Complete Task Data

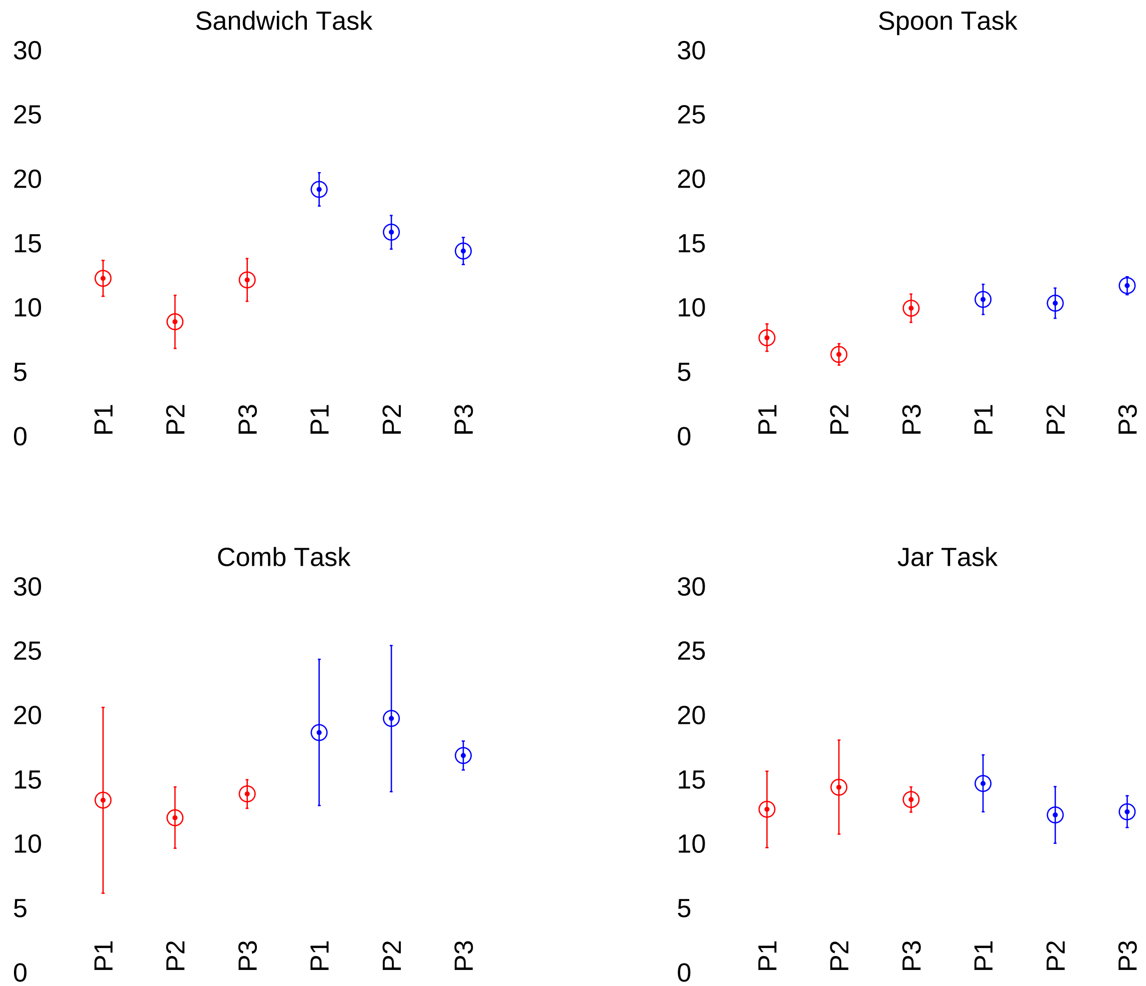


Figure 4.15: Prediction errors for all predictive schemes. Errors are reported as the mean percentage of error for subjects. A 10% error for a 10 point task, then, is an error of 1 point. For a 15 point task, it is an error of 1.5 points. Results reported are error rates as computed over multiple values of p and λ.

For all regressions, a range of parameters for the p-value threshold and λ were tested. Values for λ were varied between .5 and 15, and values for the p-value threshold were varied between .01 and .61. In Figure 4.15 we report the mean and variances in errors for all regressions. Errors

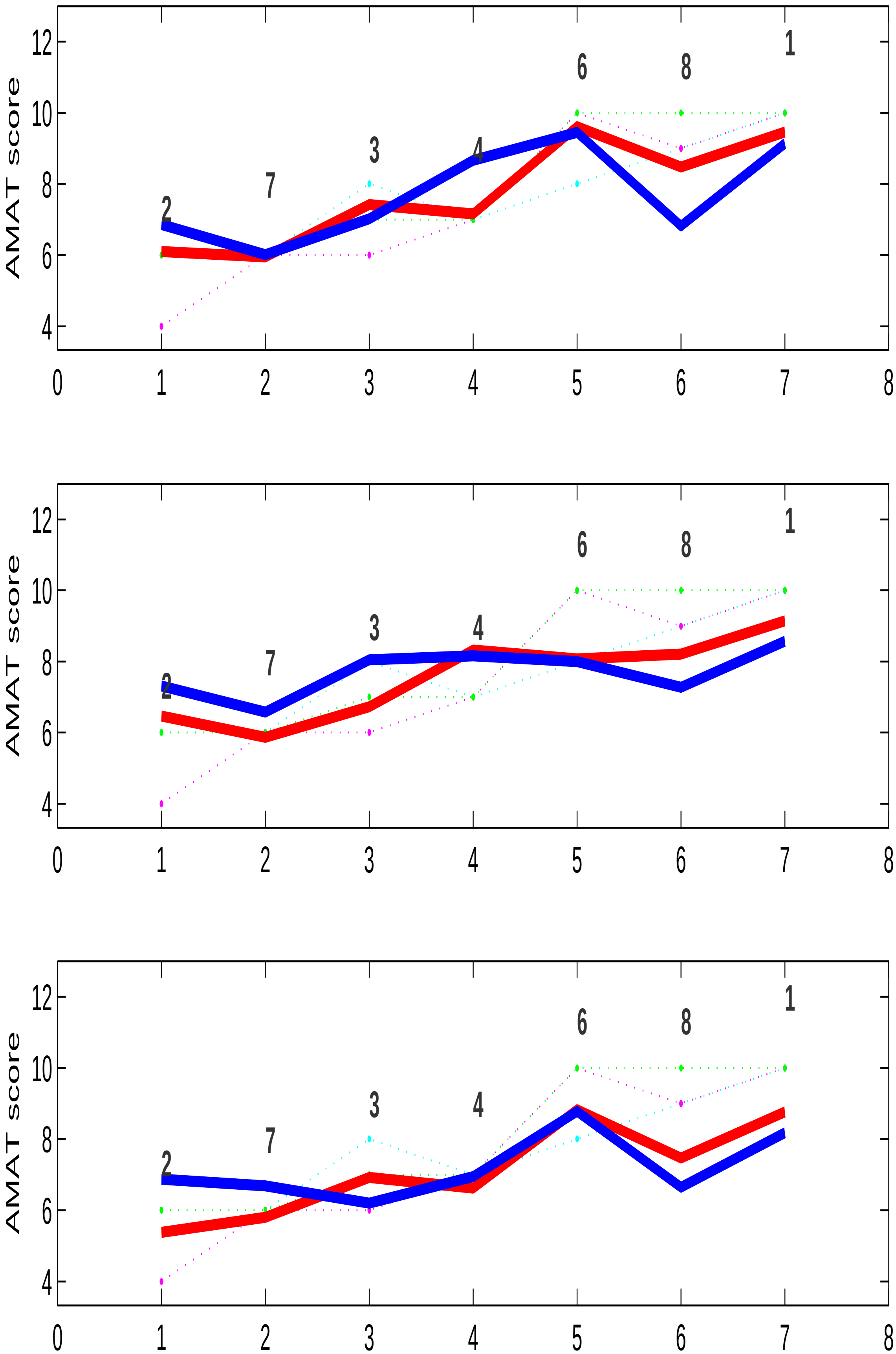


Figure 4.16: Predictions for the sandwich task. At the top are predictions made using all task data (i.e., using Predictor 1). In the middle are predictions made using task segments (i.e., using Predictor 2). At the bottom are predictions made using a generalized prediction scheme (i.e., using Predictor 3). In dotted lines are corresponding assessments from human experts.

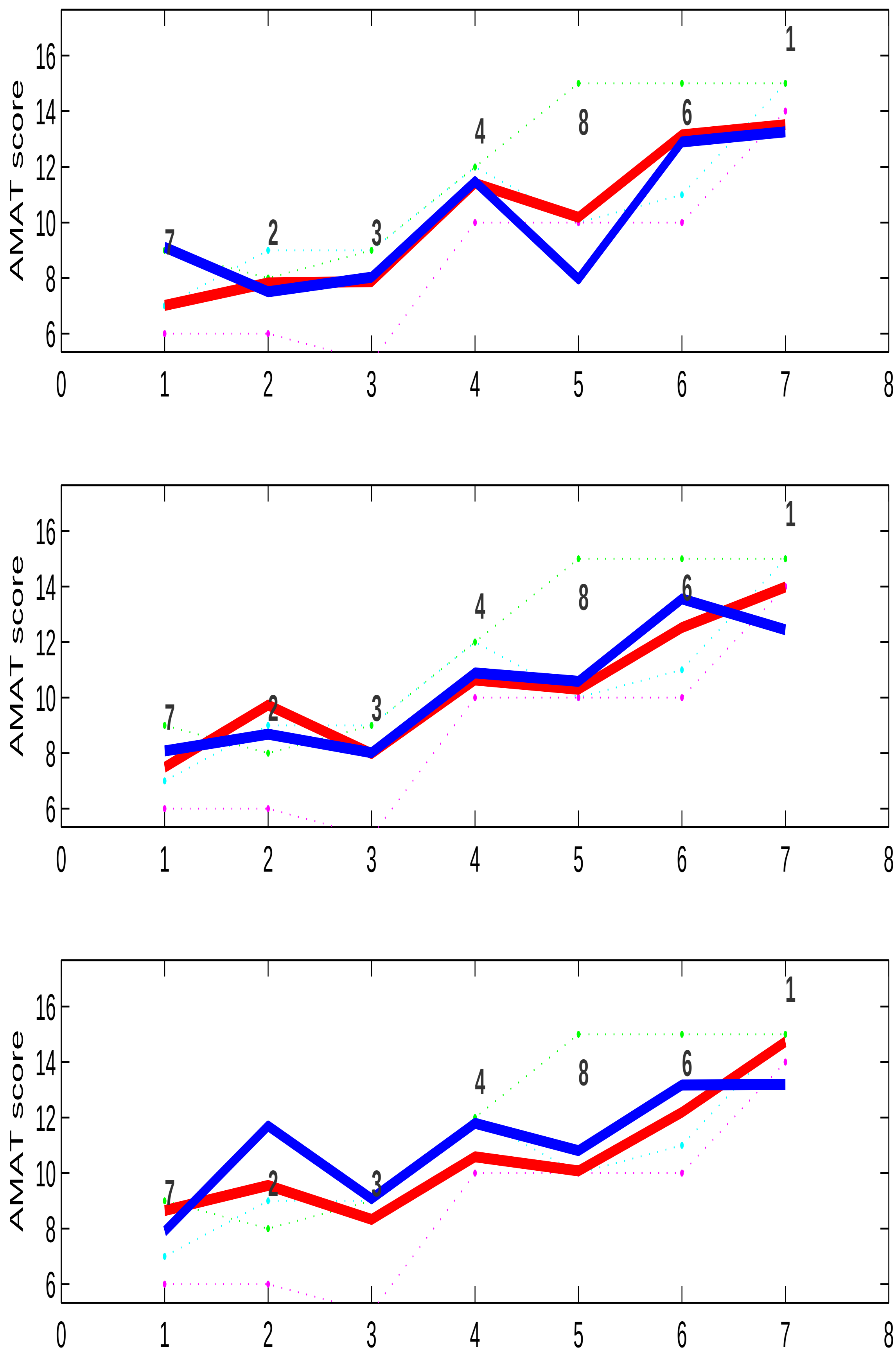


Figure 4.17: Predictions for the spoon task. At the top are predictions made using all task data (i.e., using Predictor 1). In the middle are predictions made using task segments (i.e., using Predictor 2). At the bottom are predictions made using a generalized prediction scheme (i.e., using Predictor 3). In dotted lines are corresponding assessments from human experts.

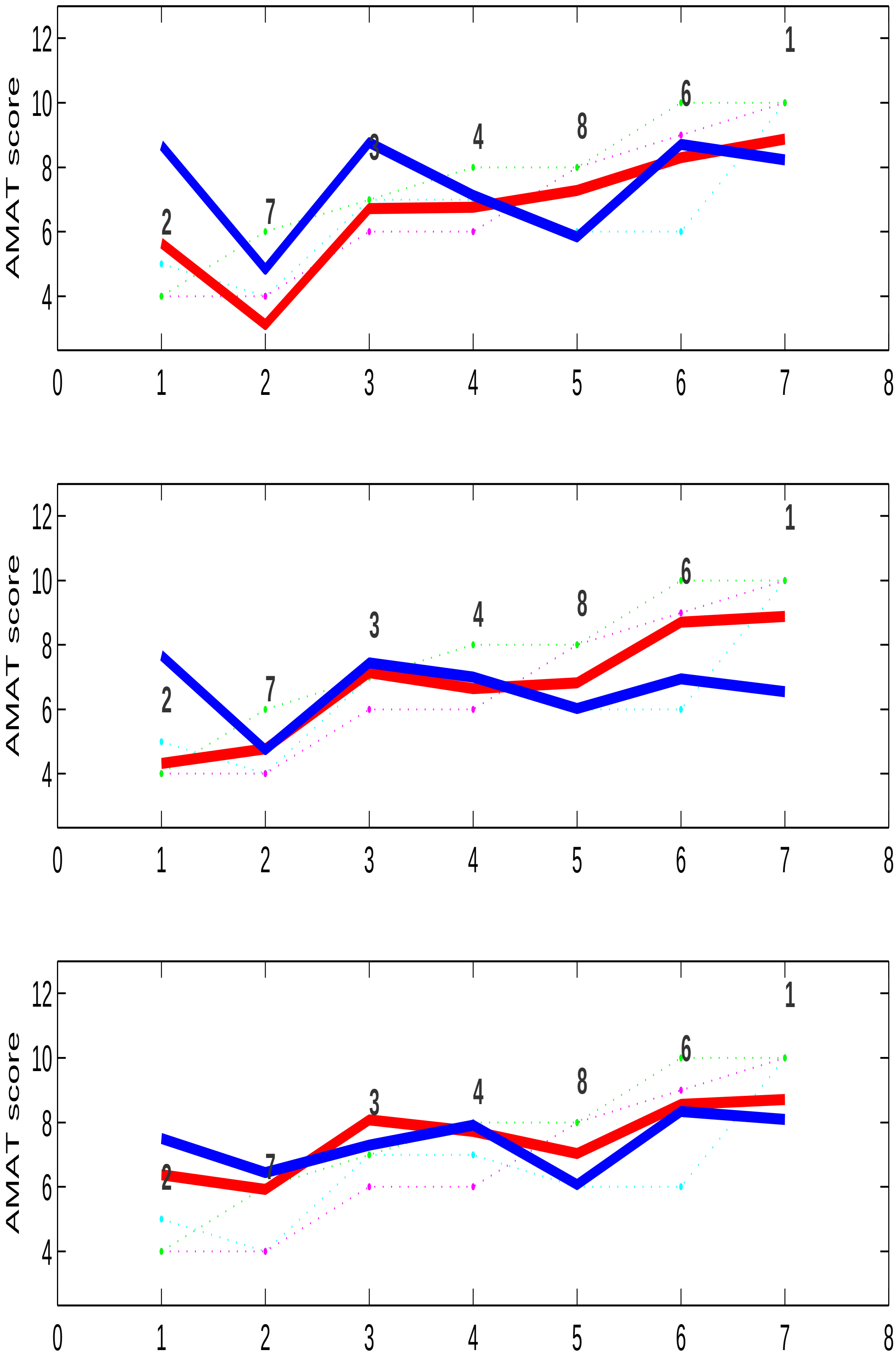


Figure 4.18: Predictions for the comb task. At the top are predictions made using all task data (i.e., using Predictor 1). In the middle are predictions made using task segments (i.e., using Predictor 2). At the bottom are predictions made using a generalized prediction scheme (i.e., using Predictor 3). In dotted lines are corresponding assessments from human experts.

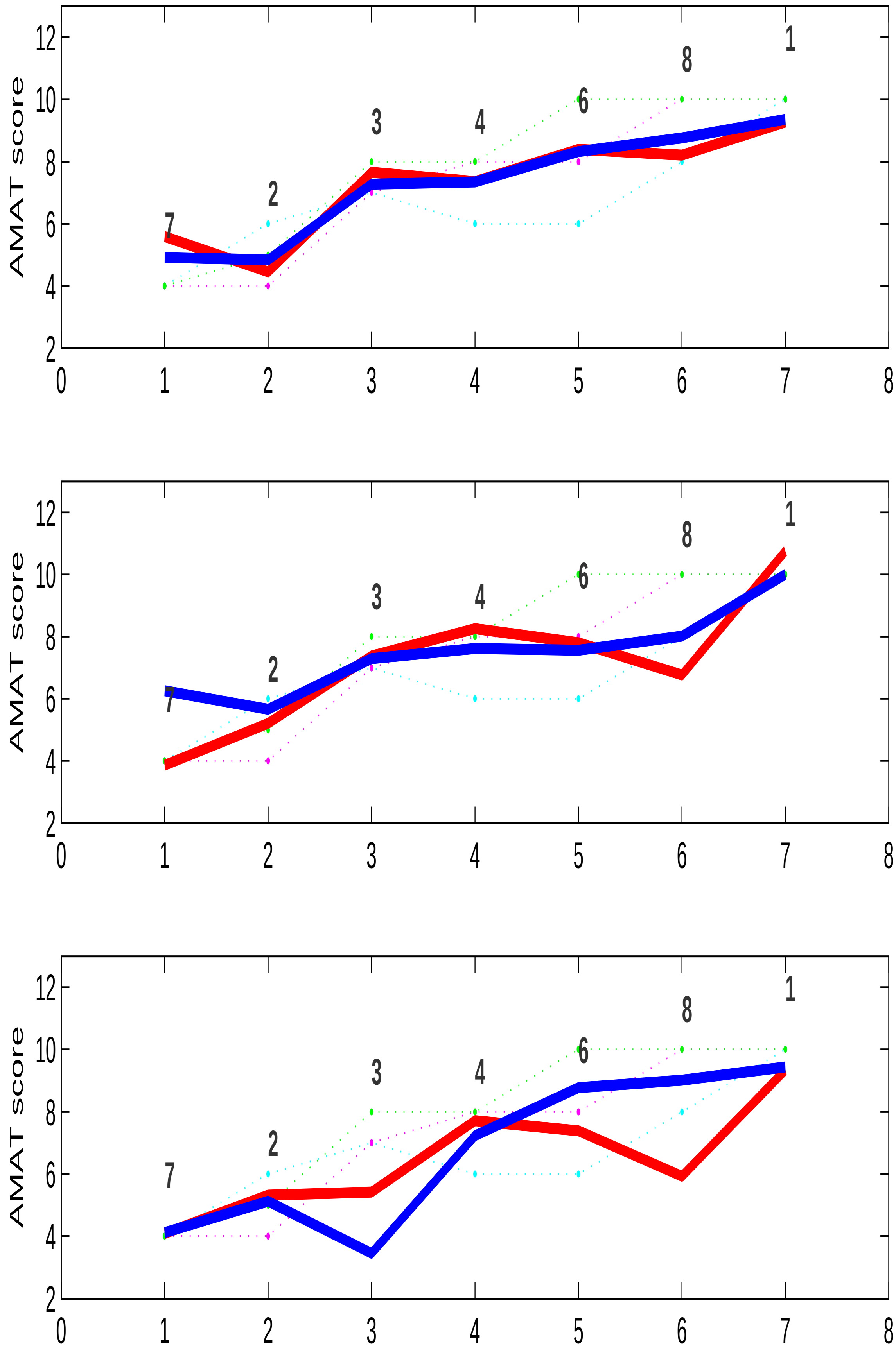


Figure 4.19: Predictions for the jar task. At the top are predictions made using all task data (i.e., using Predictor 1). In the middle are predictions made using task segments (i.e., using Predictor 2). At the bottom are predictions made using a generalized prediction scheme (i.e., using Predictor 3). In dotted lines are corresponding assessments from human experts.

are reported as the mean percentage of predicted scores that are in error for subjects. A 10% error for a 10 point task, then, is an error, on average, of 1 point. For a 15 point task, it is an error of 1.5 points. Results for the video and force based system of our design are reported in blue, while results for the motion capture based systems are reported in red.

Figures 4.16 through 4.19 illustrate results for the best regressions, as parameterized by the optimal p-value threshold and λ coefficients. In these figures, the red lines are scores predicted by the motion capture based system and the blue lines are predictions from the system of our design. The dotted lines correspond to scores from each of the three human experts. At the top of each figure are predictions made with regressions trained on specific parts of each task (i.e. Predictors Numbered 2). In the middle are results from regressions trained on data from entire tasks (i.e. Predictors Numbered 1). At the bottom of each figure are predictions from regressions trained on groups of motions sharing similar kinematic characteristics (i.e. Predictors Numbered 3). To generate this last figure, one regression was trained to predict the scores of all lifts across tasks, another on grasps and a third on manipulations.

In general, the motion capture system outperformed the device of our design. For the many of the tasks, motion capture reduced error by roughly 30%. The exception is the jar task, where our system slightly out-performed motion capture.

It is debatable as to whether Predictors Numbered 1 outperform Predictors Numbered 2. For the sandwich, comb and spoon task, motion capture regressions trained on individual parts out-performed the regressions trained on data from entire tasks. For the jar task, the reverse is true. In general, it seems that Predictors Numbered 2 are at least as good or better than Predictors Numbered 1 more often, yet the strength of each paradigm depends on the task.

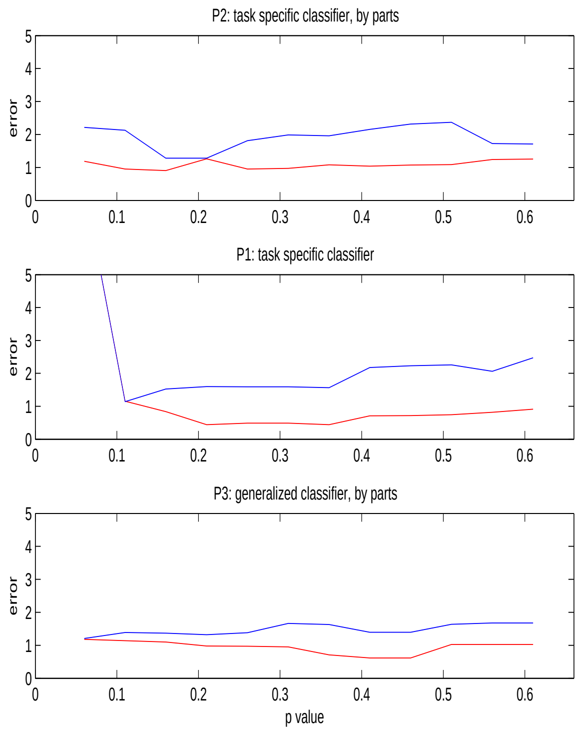
The generalized motion capture regressions (i.e. Predictors Numbered 3) perform comparably to either of other prediction paradigms for tasks when looking at motion capture data. In addition, the generalized regressions tend to produce results that are more consistent across values of λ and p. The variance in error is significantly reduced relative to that of other regressions for all tasks.

The worst results relate to the comb task. This is perhaps not surprising given our system, as we have a limited range of view and video based reconstructions suffer as a result. Error here is frequently over 15% of the AMAT range for this task, while it is under 15%, on average, for all other tasks. For both systems, some of the best prediction results relate to the spoon task. This is surprising in a way in that the spoon task has the most sub-components and therefore admits more possible error than the other tasks. In general, slight over-estimation of the scores at the low and of the scale is a frequent source of error, as is slight under-estimation of scores at the high end of the scale. The subjects with scores in the middle of the range, by contrast, are modelled fairly accurately.

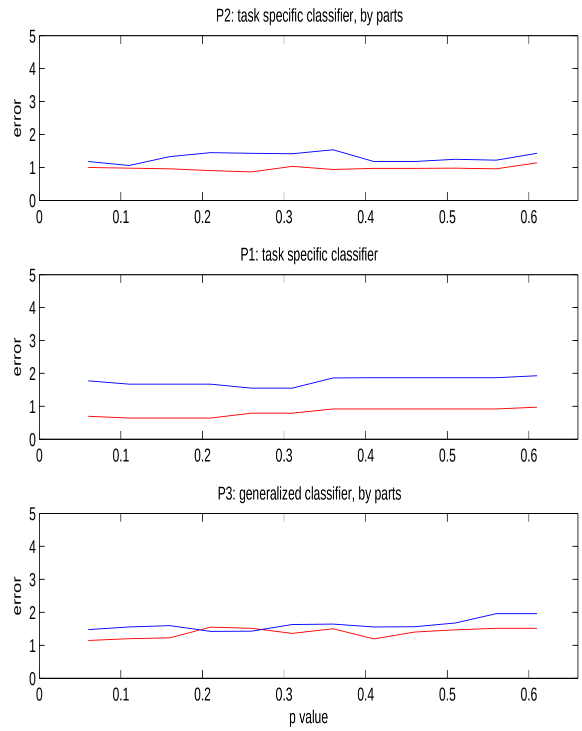
Figures 4.20 and 4.21 illustrate the impact of the selection of p-value and λon the different regressions. Of interest here is the comparison between task specific regressions (Predictors Numbered 1 and 2) and the generalized regressions (Predictors Numbered 3). The generalized regressions tend to perform best when p values used to filter features are extremely low. When p values are varied, however, performance remains reasonably stable. With the task specific regressions, by contrast, slight changes in the threshold on p values can have a very dramatic effect on performance. This is especially well illustrated for the sandwich and comb task.

Predictors Numbered 3 have an advantage over the others in that they were trained on more data. Data from all four tasks were assembled during training, meaning that p values reflect significance in linear fits over four times the number of data points. The relative stability and smoothness of error variation that relates to changes in p-values is most likely a result of the volume of data.

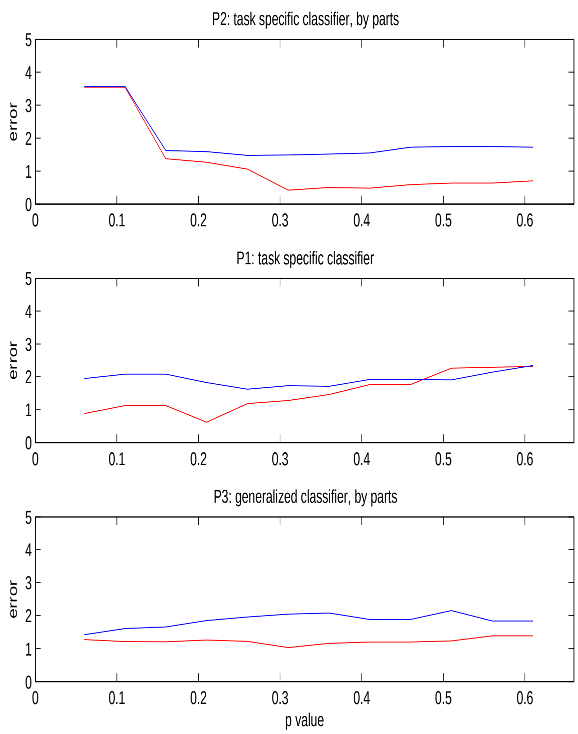
Changes in λ, by contrast, tended to have a much smoother and more predictable impact on results. For many of the regressions, the optimal value of λ is around 2 or lower, and performance becomes worse as values for λ increase. This is very much the case for both the spoon and jar task. The other tasks are less sensitive to changes in the λ parameter, on average.



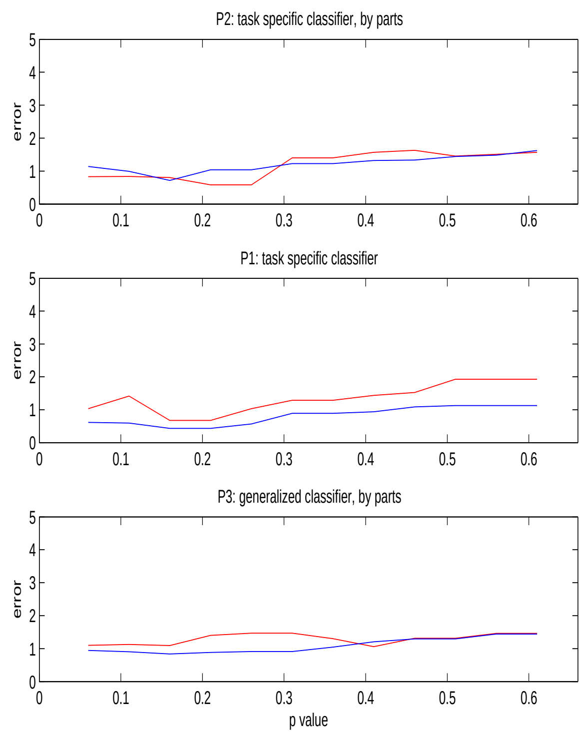
(a) Sandwich



(b) Spoon

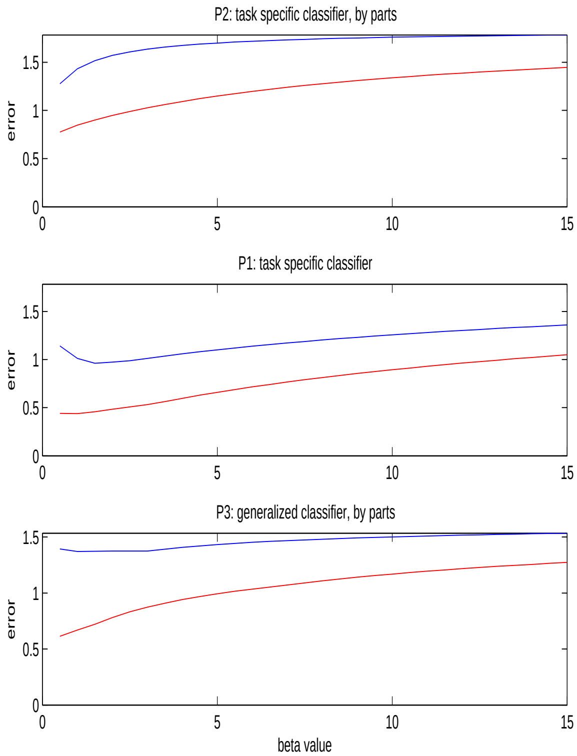


(c) Comb

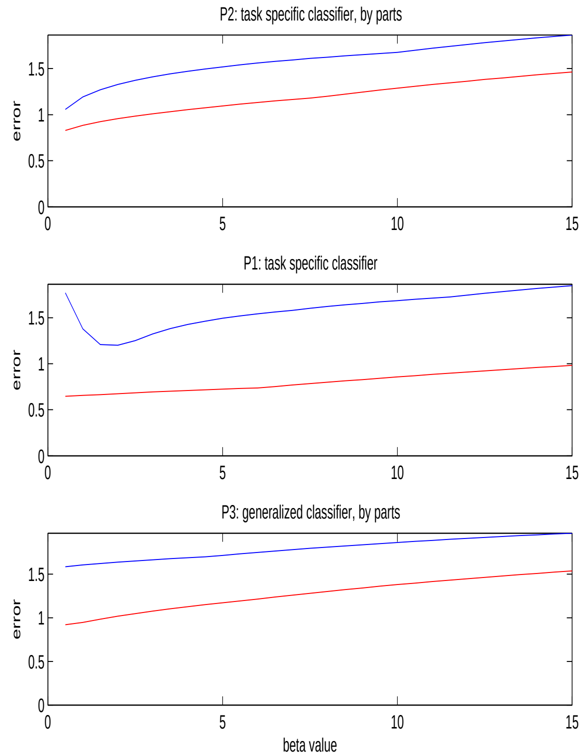


(d) Jar

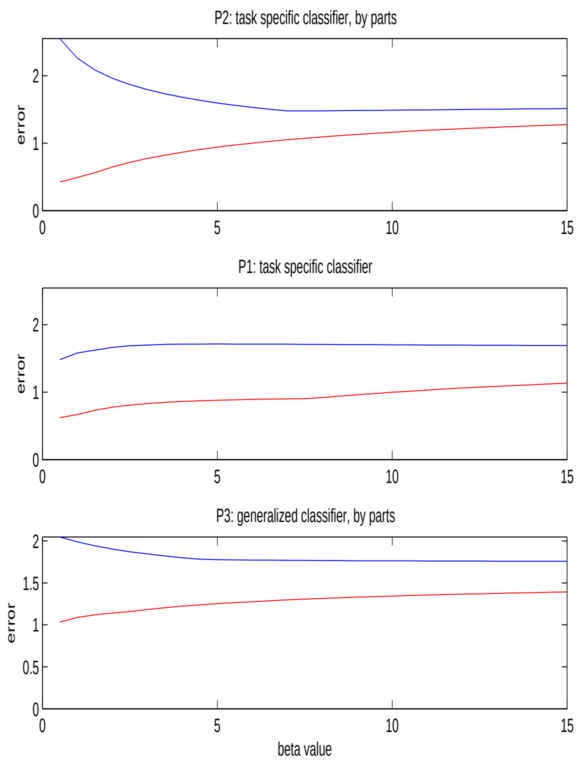
Figure 4.20: The impact of varying the p value threshold. The x axes are varying p values thresholds and the y axes are the mean error percentage across all subjects.



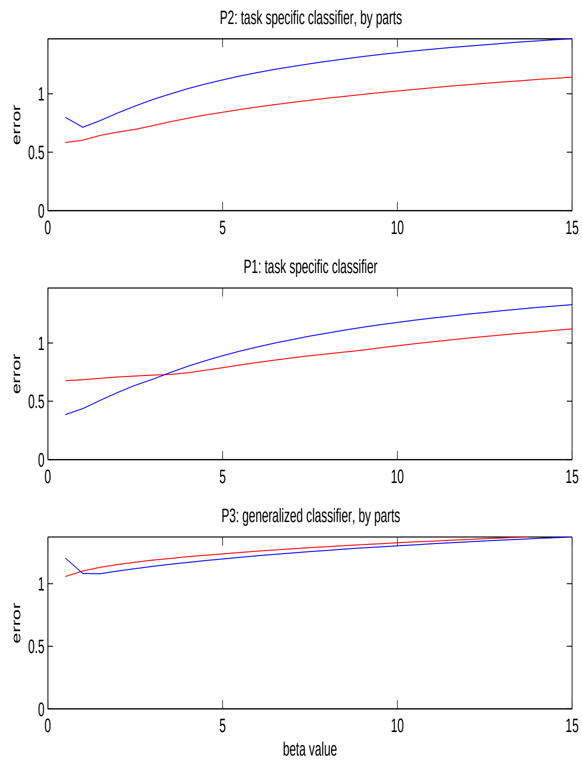
(a) Sandwich



(b) Spoon



(c) Comb



(d) Jar

Figure 4.21: The impact of varying λ. The x axes are varying λ values thresholds and the y axes are the mean error percentage across all subjects.

4.6.2 Prediction with Force Data Alone

In this section, we evaluate the ability of statistics based on recorded force data to augment or impair the predictive capacity of the kinematic data. We focus on the spoon and bean task, as this is the only analyzed task for which force data was recorded. To test the value of the force data we proceed as follows:

First, we train regressions with only the kinematic data from the motion capture device.

Second, we train regressions with only the kinematic data from the device of our design.

Third, we train regressions with only the force data from the force sensing spoon.

Finally, we train regressions with both the force and kinematic data from the tools that we developed.

In all cases, we design regressions based on total data as well as parts of the spoon task (Predictors Numbered 1 and 2).

In Figure 4.22 we illustrate error rates for all regressions. Errors are once again reported as the mean percentage of error across subjects. A 10% error for a 15 point task is an error, on average, of 1.5 points. As before, each set of regressions was built using several different values for λ and p-value parameters. P values ranged from .01 to .21, and the λ values ranged from .5 to 15. Each point on the graph, then, represents the mean and standard deviation in error across all of these parameter values. In Figure 4.23 are example predictions made with force data alone, kinematic data alone, as well as both streams of data.

In Figure 4.22, red are results from the motion capture device. Next, are results from the tools of our design using kinematic data alone. To the right of this are results from the force data alone. Finally, are results using both the force and kinematic data.

The motion capture device outperforms all of the tools of our design consistently. When looking out our device, force data alone generates predictions that are comparable to those based on kinematic data alone. The force data produces less average error when computed over the entire task, and more when computed over parts. The reverse is true when looking at kinematic

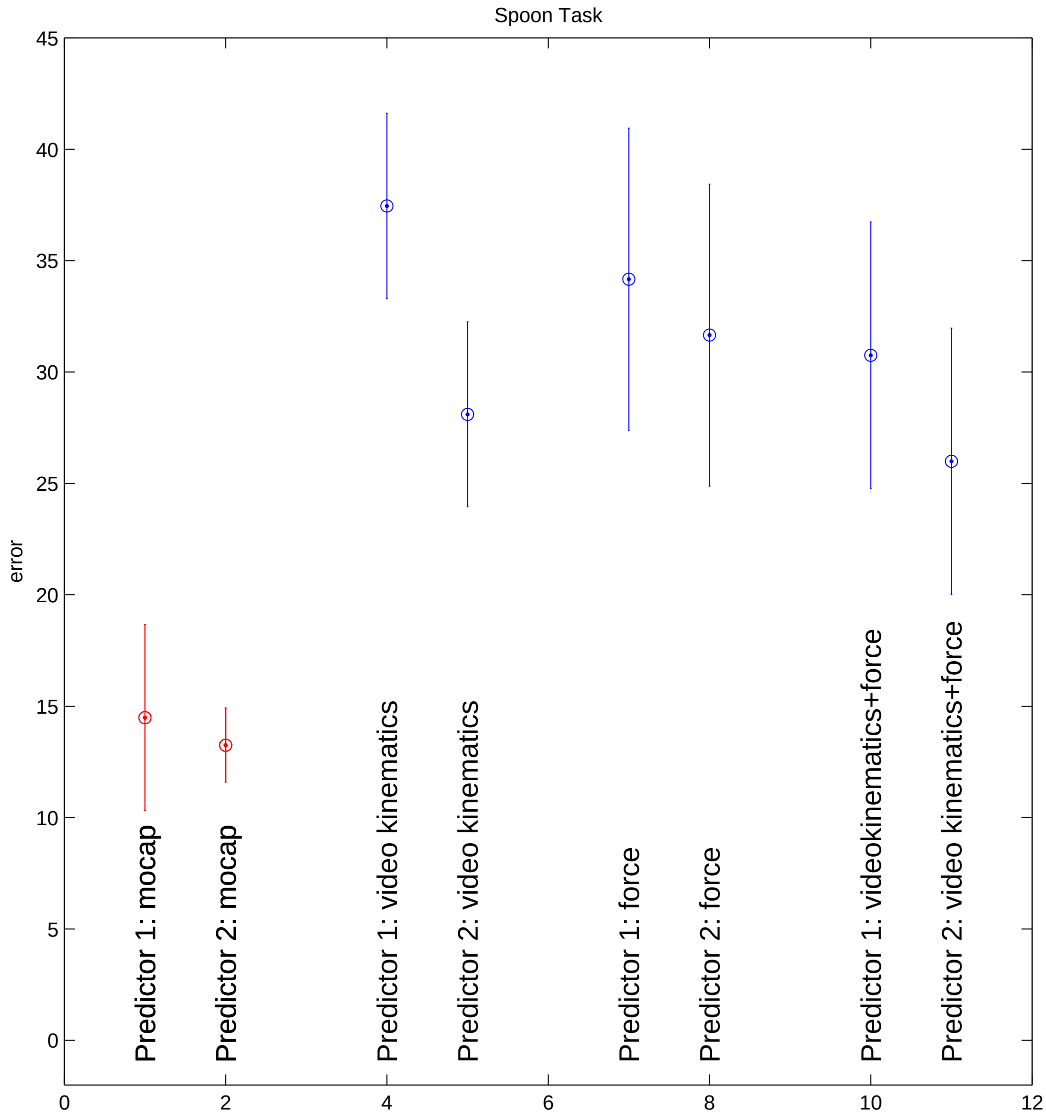


Figure 4.22: A comparison of predictions made with force data and kinematic data. Errors are reported as the mean percentage of error for subjects. A 10% error for a 15 point task is an error of 1.5 points. Results reported are error rates as computed over multiple values of p and λ.

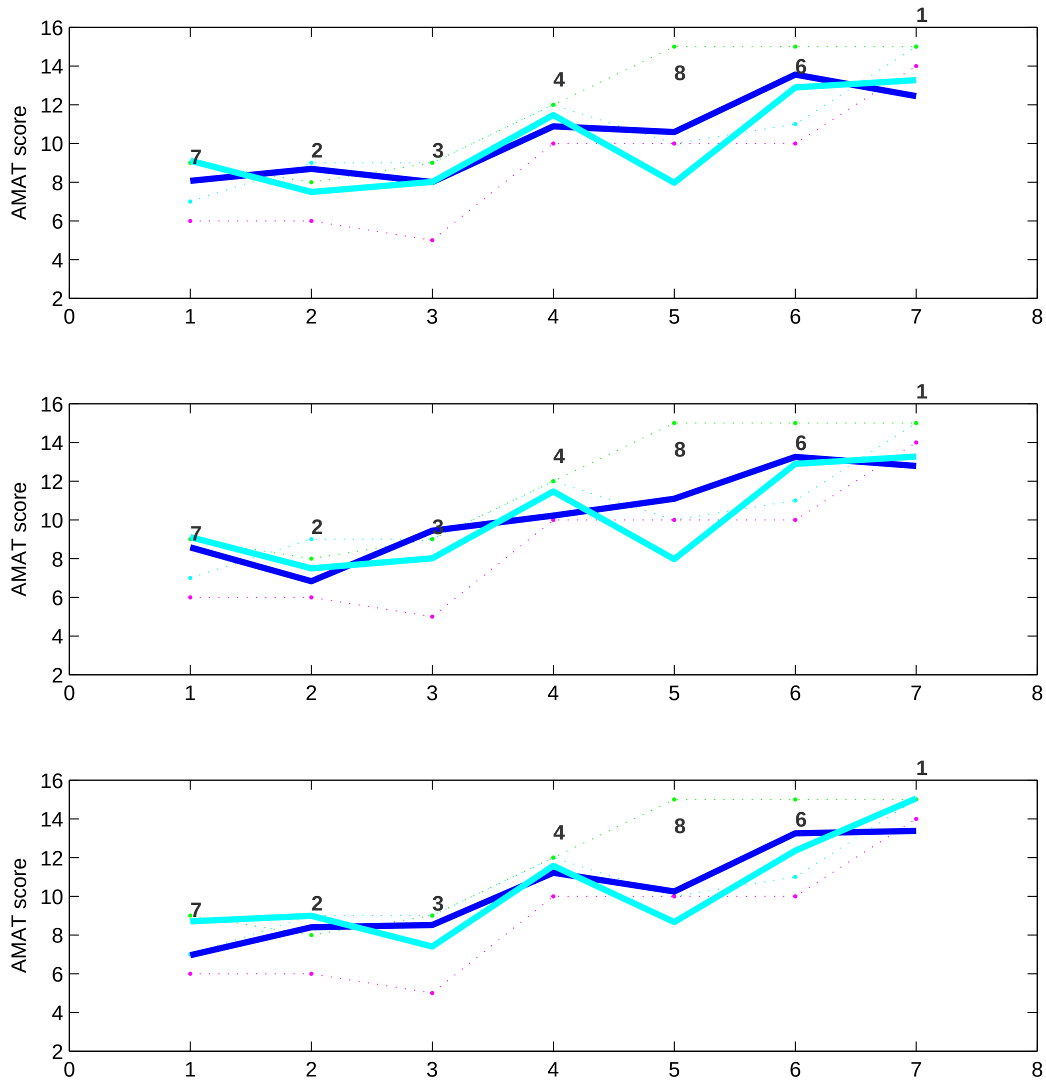


Figure 4.23: Predictions made with force and kinematic data separately. At the top are predictions made using kinematic data alone. In the middle are predictions made using force data alone. At the bottom are predictions made with both sources of data. In dotted lines are corresponding assessments from human experts. In light blue are predictions based on all task data (i.e., using Predictor 1). In dark blue are predictions using task segments (i.e., using Predictor 2)

data alone. Moreover, when the force data is combined with the kinematic data, the overall error of predictions is smaller than for kinematic or force data alone. But, predictions become slightly more sensitive to variations in p-value and λ.

4.6.3 Discussion

The key contribution of this chapter is to demonstrate that robust predictors of functional health can be produced using low cost and home appropriate devices. Prior work has demonstrated several kinematic or force statistics that correlate strongly with functional health [5, 6]; what we demonstrate here is that many of these correlations can be detected across a wide variety of tasks and used to predict scores in the absence of a clinician. We additionally show that these predictions can be made with relatively cheap technologies, like cameras and force sensing resistors.

Several motor statistics in particular seem to strongly relate to functional status irrespective of the constraints of our particular task set. The most notable of these is the displacement of the torso during reaches and manipulations of objects on the desktop. Other kinematic and force statistics, by contrast, contain information about functional status only during select portions of the tasks to which they belong. Such statistics include abduction at the shoulder, variance in flexion about the elbow, or the number of peaks in the recorded force profile on the surface of objects that are manipulated.

Interestingly, the measured features that most strongly related to functional health do not directly relate to those features that were emphasized by humans in the last chapter. While the system relied heavily on perception of the elbow and torso, the human experts focused predominantly on movement of the hands and fingers during their verbalizations. Moreover, the humans focused heavily on smoothness. Our system, however, has a very difficult time capturing smoothness statistics reliably and consistently. These differences in human, as opposed to machine, perception may complicate the provision of feedback.

In this chapter, we also illustrate several modelling options options for automatic score prediction that yield reasonable results. We can, for example, model each task independently or we can model each task as the sum of its parts. We can also create generalized regressions that operate across tasks.

The use of generalized regressions to predict functional scores opens the doors to interesting future research. A general predictor of lifts, for example, will not work on all lifts but it may work on lifts that share some common features, like start and end position of the hand. Scores for lifts in the context of the comb task, for example, were not well predicted by regressions in our study. So how broadly can a general tool be used to score the functional health of lifts? Which lifts are comparable, and which are not?

In the future, we would like to see stroke survivors monitored continuously by low cost and ubiquitous monitoring devices at home. We would like to see discriminatory task elements, like lifts to the mouth or reaches to targets, automatically detected and evaluated as they take place during real, every day routines. here, we prototype a system that is consistent with our long term vision, but which works in the context of a laboratory environment and with several simplifications. Although the current system makes use of markers and objects that have wires, we are optimistic that the state of the art is such that markerless algorithms and wireless devices can easily be integrated into the measurement framework we have outlined. Our current work, then, serves as a baseline for development and technical optimization.

For the clinical community, we seek to provide food for thought relating to the kinds of therapy, and therapeutic activity, that might take place in the home. For the technical community, by contrast, we provide a set of application oriented evaluation criteria with which to determine the utility of algorithms, particularly those which relate to kinematic tracking and activity recognition.

Chapter 5

Future Work: Moving to the Home

The results from the last chapter demonstrate a reasonable capacity to predict functional scores on the AMAT automatically, robustly, and with inexpensive devices. Moreover, they illustrate the potential to predict functional health based on the perception of lifts, grabs and object manipulations that take place in the context of several different tasks.

In the future, we want to deploy automatic assessment tools in a range of stroke survivors' homes and workplaces. There, we expect to make functional assessments continuously, over days or weeks or months, and based on the perception of activity as it actually takes place in real functional environments. We expect to perceive lifts, grabs and manipulations as they relate to a wide variety of functional activity, and to extract those lifts and grabs that carry the most information about functional health. We will then summarize our findings for review by therapists or clients of therapy, thereby providing a record of functional change as it is situated in the real world and over long stretches of time.

In order to get a feel for the kinds of problems that are specific to deployment in homes and workplaces, we made a preliminary effort to instrument a single stroke survivor's home. This involved transporting all of the cameras and force sensing objects to a private home, and installing them in his basement recreation room. Once installed, we measured a variety of desktop activity on the part of this stroke survivor over a two week period.

The exercise 200 GB of data that will fuel future research. The information we acquired promises to provide insight into how robust our assessment technologies are to new environments and how accurately we can track performance quality relating to activity outside of the clinic. The installation, moreover, has clearly illustrated some of the challenges and difficulties that are specific to home deployment.

In this chapter, we report a preliminary set of results from this initial home deployment. The results that are quantitative relate to the force information that was acquired. The kinematic data, by contrast, required time consuming daily calibrations of cameras; calibration and reconstruction, then, remains a work in progress. Nevertheless, we illustrate some of the challenges encountered; much work, however, is left for future research.

5.1 Methods

A single stroke survivor was recruited for the study in the home. This individual was a single 83 year old male, 13 years post-stroke and he was recruited through a local area stroke support group. As in prior studies, the participant was screened for both cognitive as well as basic physical and functional health. The preliminary assessments he was asked to perform included the Semmes-Weinstein, the upper body and speed portion of the Fugl-Meyer Assessment (FMA) and the Visual Analog Scale (VAS). He was also assessed using the Mini-Mental Exam (MME) and the Ashworth scale at the elbow and wrist. Results from preliminary assessments are given in Table 5.1. All assessments were conducted in the home and by a licensed Occupational Therapist with more than 10 years experience working with stroke survivors.

After initial screening, the investigator and an Occupational Therapist visited the home of the stroke survivor and met both with him and his wife. This visit was made to determine the most appropriate site for measurement equipment, as well as to determine the activities to be measured in the home. The visit was also made so as to connect with others who were living in the home and to inform them of the intention, purpose and duration of the study.

ID	Age	Gender	Years Post-Stroke	M.M.E.	Lesion Site	Dominance	Aphasia
1	83	Male	13	25	Right	Right	None
ID	Ashworth Elbow	Ashworth Wrist	Touch	Vision	Audition	Pain (rest)	Pain (motion)
1	0	0	Intact	Impaired	Intact	0	0
ID	FMA Wrist	FMA Hand	FMA Arm	FMA Speed	FMA Total		
1	7	13	26	6	52		

Table 5.1: Basic demographics of the participating stroke survivor.

AMAT Task/Subtask
Task 1: *Cut ”Meat”*
1. Pick up knife and fork 2. Cut ”meat” (Play-Doh) 3. Fork to mouth
Task 2: *Foam ”Sandwich”*
4. Pick up foam ”sandwich” 5. ”Sandwich” to mouth
Task 3: *Eat With Spoon*
6. Pick up spoon 7. Pick up dried kidney bean with spoon 8. Spoon to mouth
Task 4: *Comb Hair*
9. Pick up comb 10. Comb hair
Task 5: *Open Jar*
11. Grasp jar top 12. Screw jar top open
Task 6: *Use Telephone*
13. Phone received to ear 14. Press phone number
ARAT Task
Task 1: *Lift 10 cm block*
Task 2: *Lift 2.5 cm block*
Task 3: *Lift 5 cm block*

Table 5.2: AMAT and ARAT tasks chosen for home recording.

1. Designing the Activity Regimen. The activity regimen for home practice was designed to span no more than two hours in length. It was also designed to be performed at an instrumented desktop situated in the subject's home, every week day over a two week period. The activities that were chosen for measurement were of three varieties:

First, the same elements of the AMAT that had been measured in the laboratory were selected for performance. These activities were chosen so as to easily and directly permit use of those assessment devices that were trained in the laboratory. These activities are listed in Table

5.2.

Second, three elements of the Action Research Arm Test (ARAT) were selected for daily performance. These activities required the participant to grasp and lift blocks of three different dimensions, and to place these blocks on top of a small box. ARAT activities were selected because they involved simple, constrained lifts and grasps confined to the same work area as the AMAT. Diagnostic kinematic and dynamic statistics relating to AMAT, then, were expected to generalize well to these tasks. ARAT activities are also listed in Table 5.2.

Finally, we selected activities that were less constrained, but which contained easily detectable lifts, grasps, and manipulations confined to the desktop. This last category of activity was designed primarily to be motivating to the participant, to be easily decomposable into task segments, and to elicit movement on the more impaired side of the body that was challenging but not exhausting. After discussion with the subject and therapist, daily games of checkers were selected as the activity of interest. These games were to take place on a larger than normal board, however, so as to promote reaches spanning the whole table. The board that was agreed upon was 18 x 18 inches in size; the farthest squares were located roughly a full arm's length from the subject during game play.

2. Instrumenting the Home. The site that was chosen for the location of our system was the participant's basement recreation room. This was a room roughly 200 square feet in size, with low overhead ceilings, roughly 8 feet high; it was the regular site for family card games. There were no windows in the room and hence no ambient light, making the room easy to control for the purpose of measurement with cameras. It was also adjacent to the garage, making the transportation of large or heavy equipment relatively simple.

The installation spanned an area that was roughly 64 square feet in size. It consisted of eight cameras, as before, all of which were installed on poles that were designed to hold commercial lights. These poles, which were cast iron, extended from the floor to the ceiling and were stably affixed to large cast iron bases. Four poles were erected roughly one foot from each of the four corners of a card table; this table was 3 feet by 3 feet in size and its surface was 2 and a half feet from the floor.

Video from the eight cameras was sent to a set of AXIS MPEG2 severs, which converted the video to MPEG2 format in real time. These MPEG2 streams were then sent via Ethernet through a desktop computer; data was subsequently stored on an external hard drive. The desktop was a 3 GHz PC with a front side bus. Force data from the objects was acquired through a Data Translation DT 9800 analog to digital converter. Wires ran from the objects to this converter box, which was connected to the same desktop machine via a USB connection. All of the devices were synchronized at the onset and offset of recording with a button press that came from the investigator. This button press would simultaneously send a flash of light to each recording camera and a pulse to the analog to digital converter.

Two challenging features of the installation were the negotiation of electricity requirements and the safe storage of wire and cable. The MPEG2 boxes required 8 amps in total; to avoid blowing fuses that might be rated only for 15 amperes we therefore used multiple circuits. Accessing these circuits, however, required the use of extension cables. The extension cables and the wires related to the setup all combined to create a potentially hazardous situation in the house, as they were easy to trip over. To alleviate this risk, the investigator agreed to store cables securely in the home at the end of each day and re-cable instruments before daily recordings. This proved to be an acceptable and safe arrangement, but it added a created a substantial amount of additional work. Not only was work created as a result of the cabling itself, but work was created when cameras were bumped or touched during the re-wiring. Bumping and touching of cameras meant that calibrations for each day of recording are each somewhat unique.

Finally, to facilitate easy visual tracking, we once again asked the subject to wear a brightly colored jersey. This jersey was custom tailored for the subject, to insure a proper fit. It was also made of a slightly lighter fabric, as recording took place when the weather was warm. Prior

Figure 5.1: The cotton jersey.

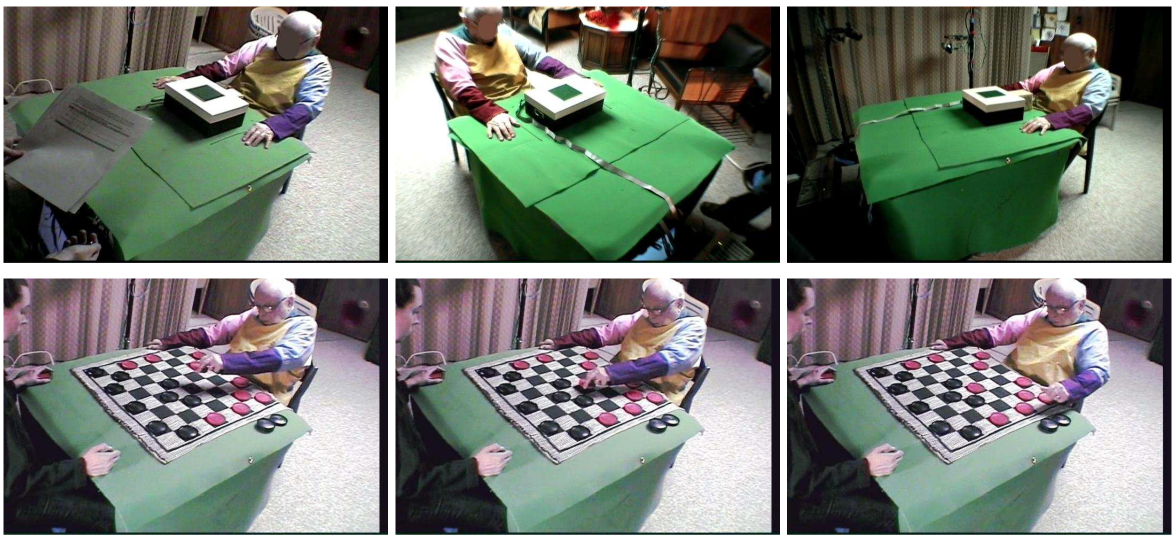


Figure 5.2: Views of the subject performing activities in the home.

jerseys were made of fleece which, while providing a very nice and matte surface with consistent colors, were hot. The cotton jersey is illustrated in Figure 5.1.

3. Performing the Activity Regimen. Recordings of activity took place every weekday over a two week period. The protocol for every day of activity remained the same. The investigator would arrive at the home of the participant, put cables for the recording devices in place, adjust cameras, and turn the system on. The subject would then be outfitted with the colorful jersey and seated at the desktop in the basement recreation room. Once seated, he would perform the six elements of the AMAT and the three elements of the ARAT. Finally, the day's activity would close with extended games of checkers. During these games, the investigator sat opposite the participant at the card table, and served as his opponent. Before any activity took place, the subject was informed that he was permitted to take rests whenever he needed, or to terminate recording at any point.

During the performance of the AMAT and ARAT elements, the starting locations for objects were standardized according to either assessment's instructions [132]. These starting positions required the subject to have his back touch the back of the chair, his elbows at roughly 90 degree angles, and hands on designated spots on the counter top. The starting location for the checkerboard was also standardized from day to day; this location was such that the farthest squares on the board were roughly a full arm's length from the subject's sternum.

AMAT and ARAT tasks were performed at a comfortable movement speed and at the prompting of an investigator. Once the subject was completed with these tasks, he was asked to return his hands to their starting locations. During game play, starting and stopping locations were more flexible. The subject was asked to keep his hands visible during game play, and not to rest them under the table and on top of his lap.

An illustration of activities as they took place in the home is found in Figure 5.2. In the top row, the subject is performing an element of the ARAT with a force sensitive block. In the bottom row, the investigator and the subject are engaged in a game of checkers on an 18 x 18 inch board.

During the performance of activities, data was captured with video cameras and force sensing objects, if they existed. All measurement devices were synchronized with a button press by the investigator.

3. "Ground Truth" Data. In order to generate "ground truth" data relating to the functional health of the individual in the home, we asked a licensed Occupational Therapist to make visits to the house periodically. During these visits, the therapist watched performances of the AMAT and ARAT, and assessed the quality of the perceived motions. The first visit was made on Day 3 of the recording, the second visit on Day 5, and the last two visits were on Day 8 and 10.

5.2 Data Analysis

The first two days of recording primarily served to provide an opportunity to debug the camera setup and the activity regimen. By day three, a rhythm had been established and recordings proceeded in a more smooth and organized fashion. As a result, we discard the first two days of recording and focus our results on the data generated after Day 2.

We report results relating to the force data here. Before computing any statistics on the force profiles, we clipped them so that the first sample represents the first moment when the computed force was over 5% of the maximum force, and the last sample represents the last moment when the recorded force is over this threshold. The spoon task is slightly complicated by the fact that the subject remains holding the spoon at the end of this task, in many instances.

After clipping, the following statistics were computed across recorded force profiles:

1. The mean, variance and maximum recorded force. To capture potential force disturbances or changes in force distribution across time, we measured the peak, average and variances of forces produced during object manipulations. All forces were measured in volts, converted to Newtons, and then summed across all the sensors placed on an object. These summed forces were then median filtered with a filter that was a sixth of a second in width. They were also passed through a fourth order Butterworth filter, to remove high frequency noise.

2. Smoothness of force. To capture jerks in the force profiles, we measured smoothness of force production. As in prior studies, smoothness in force profiles was measured as the sum of squared jerk as well as and the number of force profile peaks.

3. Force MAPR. Finally, we measured the proportion of time that the recorded force exceeded a given percentage of the peak force. This we define a force related "Movement Arrest Period Ratio" (MAPR). It may how ever be more accurately termed a "Force Arrest Period Ratio" (FAPR).

In order to determine the relative significance of any of these features, we tested the strength of their association with the AMAT scores assigned by the human expert. AMAT scores, however, were only assigned at 4 time points, while force recordings were made over 8 complete days. In order to fill in the blanks, AMAT scores from the human were linearly interpolated between days. The expected score for a given day between the therapist's visits was therefore assumed to be a linear interpolation of the two scores assigned both before and after that day.

To test for the significance of the various force features, we correlated the interpolated AMAT scores from the human and the computed force features. We then looked for correlations that strongly departed from zero. To quantify the deviation of each computed correlation from zero, we computed the following test statistic, t, for each correlation:

$$t = \frac{R(i) - 0}{SE(R(i))} \tag{5.1}$$

Where i is the index of a particular feature (i.e. mean or maximum force, for example) and $SE(R(i))$ is the estimate of the standard error in this features feature's correlation with AMAT scores over time. Computed t statistics are distributed according to t distributions. Those t statistics that lie at the tails of these distributions indicate significant departures from zero, and can be identified by computing associated p-values. Smaller p-values indicate t statistics that are located farther from the center of the distribution.

5.3 Preliminary Results

Figure 5.3 illustrates the AMAT scores that were assigned to the participant by the therapy expert over the two week interval. Each of these scores is the sum of scores that were produced on the six tasks in the battery. The scores for each task are shown in Table 5.3.

Evaluation	Cut Meat	Sandwich	Spoon	Comb	Jar	Telephone	Total
1	12	9	13	6	8	7	55
2	12	9	12	7	9	8	57
3	15	10	15	8	10	9	67
4	12	10	15	8	10	8	63

Table 5.3: AMAT scores over a two week period.

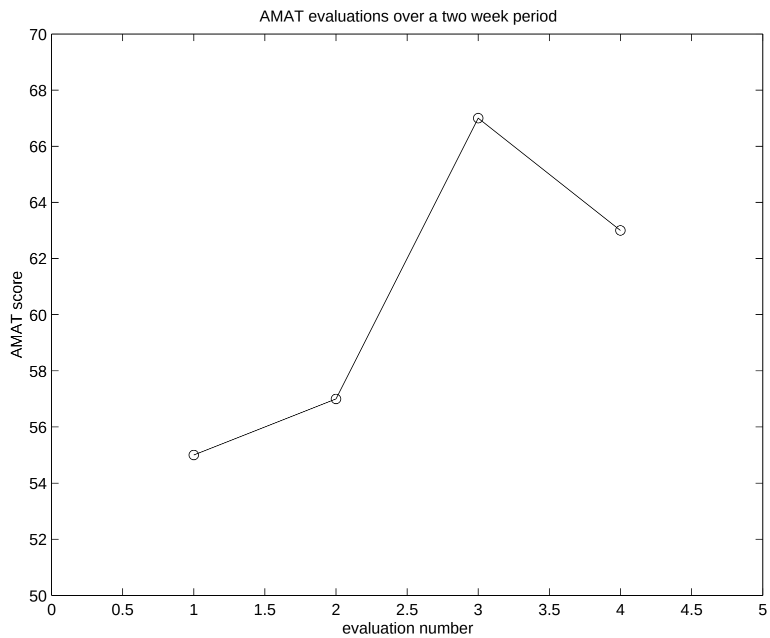


Figure 5.3: AMAT scores across a two week period.

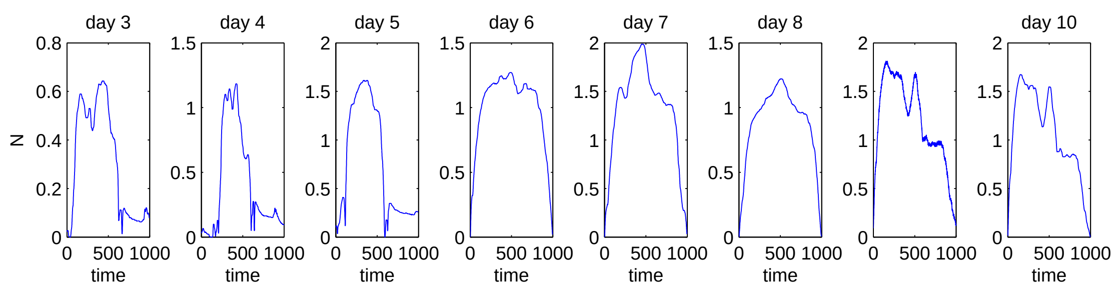


Figure 5.4: Cube force profiles across a two week period.

Figure 5.4 illustrates the sum of force measurements as measured across all the sides of the 7.5 cm cube during the performances of the ARAT. The Y axes on these graphs is measured in Newtons and the X axes span the number of samples acquired during the task. In Figure 5.5 are the sum of force profiles recorded from the handle of the force sensing spoon. The X axes span, as before, the number of recorded samples.

In Tables 5.4 and 5.5 are results from the correlations between the computed force statistics and the interpolated AMAT scores from the expert. Numbers in each column correspond to the p-values that test for the strength of each correlation and, in parentheses next to these p values, are the correlations themselves. An asterisk indicates significant association with functional score at

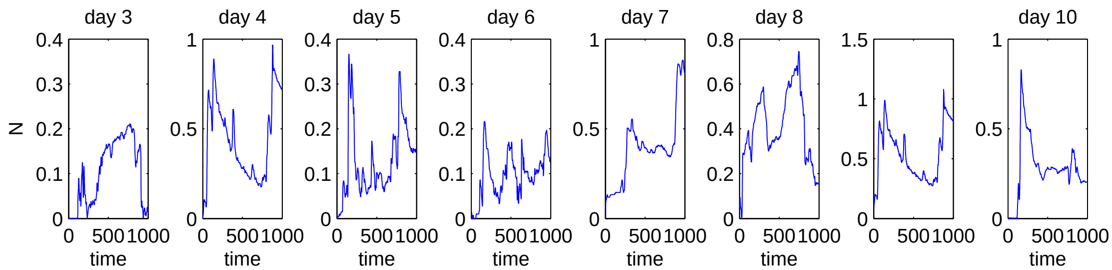


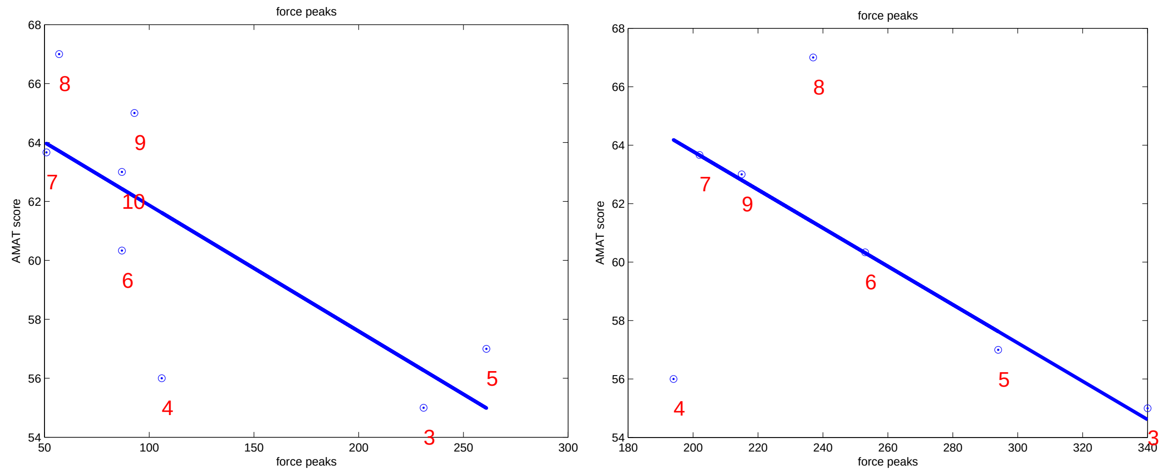
Figure 5.5: Spoon force profiles across a two week period.

Statistic	P-value (corr)	
Mean	0.049 (0.77)	*
Maximum	0.24 (0.49)	
Variance	0.013 (-0.86)	*
Jerk	0.22 (-0.55)	
Peaks	0.057 (-0.76)	*
MAPR	0.13 (0.66)	

Table 5.4: Force statistics from the cube.

Statistic	P-value (corr)	
Mean	0.73 (0.18)	
Maximum	0.33 (0.43)	
Variance	0.17 (-0.61)	
Jerk	0.23 (-0.55)	
Peaks	0.022 (-0.51)	*
MAPR	0.27 (0.5)	

Table 5.5: Force statistics from the spoon.



(a) Cube (b) Spoon

Figure 5.6: Peaks in force profiles, from two objects, across a two week period.

the .1 level.

Finally, in Figure 5.6 are graphs of the number of peaks in force profiles as related to functional scores. One the left of the figure are peaks in profiles recorded by the cube, and on the right are peaks recorded by the spoon. Each point corresponds to a different day of measurement, and the days have been indicated by numbers on the graphs.

Both objects record increases in maximum forces across the various days of use that are associated with increases in functional score, but these associations are not significant at the .1 level. The cube does record significant associations between functional score and the variance in force, the mean force produced, as well as the number of peaks in the force profile. The only significant association between force statistics and functional score for the spoon relates to the number of peaks in the force profile. In all cases, the variance in force is negatively associated with functional score. The is the same for the recorded jerk as well as the number of peaks in the force profile. The mean and maximum forces, by contrast, are positively correlated with score, as is the force related MAPR.

The same directions in correlations, for the most part, are found in the associations between force statistics and functional score in the laboratory environment. In the lab study, the mean force recorded by the force sensing spoon was positively correlated with functional score, as was the MAPR. The most significant association between a force statistic and AMAT score in the lab related to the peaks statistic, and the same is true here.

5.4 Discussion

It is encouraging so see that results in the home relate to the results from the laboratory study. Where the results in the laboratory were used to discriminate between functional health across many individuals, however, here the results discriminate between the functional status of the same individual across many days. This gives us reason to believe that predictive regressions trained in the lab with data from many individuals can potentially be used to track the performance change of a single individual in the home.

Moreover, it is encouraging to see that the performance trends that are captured by one object can generalize across objects. Trends captured by the force sensing cube were also well reflected in the data from the force sensing spoon. The individual force profiles from the spoon, however, are decidedly more complicated than those from the cube. Nevertheless, the objects obviously share some similarities when it comes to the manner in which they are manipulated across days.

The reasons for the change in the home, however, remains somewhat unclear. Some confounding variables that may have impacted motor performance include:

1. Familiarity with the Protocol. The recording setup as it currently stands does not integrate seamlessly into the home. It requires a fair amount of setup time and a fair amount of learning on the part of the subject. The objects have wires and the individual who interacts with the system must wear a jersey. All of these elements make for an unusual situation that requires at least a little bit of time to become fully comfortable with. The improvements in the recorded scores, then, could be the product of increased familiarity with the recording setup and with the measurement protocol.

2. Better Preparation for Recording. Coupled with the increased familiarity with the recording setup came an increased ability on the part of the subject to prepare for recordings. The recording setup was located in the subject's basement recreation room. Getting to the setup required the participant to climb down a full flight of stairs. During the first few days of recording, this climb down the stairs occurred immediately before recordings began. The climb, moreover, was by far the most tiring activity of the day as may very well have influenced performance at the desktop for some time afterwards. After a few days of familiarity with the recording protocol, however, the subject would prepare for sessions by entering the recreation room perhaps twenty minutes before the sessions began. He would work in his office and recover strength; by the time recording began on these later days, then, he was relatively well rested. This difference may also have impacted the quality of later, versus earlier, scores.

Irrespective of the reason for the increase in scores, we demonstrate that statistics which capture this increase can be captured with low cost devices.

5.5 Conclusion and Future Work

The purpose of this portion of the work is really to demonstrate that it is feasible to instrument the home of stroke survivors with low-cost sensing devices, and to record clinically meaningful motion in this environment. Some of our preliminary results tell us that:

1. Low cost measurements made in the home can, in fact, capture clinically meaningful change. Force data yields several statistics that capture functional change over a two week period. Many of these statistics, moreover, seem to directly relate to meaningful statistics in the laboratory. Such statistics include the number of peaks in the force profile as well as the mean force recorded on objects surfaces.

2. Manipulations of different objects share characteristic force features. In the laboratory, we found that kinematic data from different tasks were comparable. Torso displacement, for example, related strongly to functional score across the tasks that were measured. Here, we find that force statistics also share similarities across tasks. This is encouraging, and tells us that task sub-components may possibly be defined not only as they relate to kinematics, but as they relate to force data as well.

3. Information rich activity regimens can be both designed and measured in the home. One of the interesting features about our work in the home is that it involved partnering the capacities of a particular individual with the strengths of our measurement devices. Walking into the home, we knew we needed to restrict the activities we recorded to those that would be likely to emulate the laboratory study. A therapist and a stroke survivor worked within these constraints and in the home to select activities that would be motivating, fun, and clinically meaningful. This way of interacting with clients in the home emulates what we expect to have happen in the future. We expect to build devices that are capable of measuring functional change as it relates to key kinds of movements, and we expect that clients of therapy and therapists will design home rehabilitation programs that keep this in mind.

Although we present encouraging initial results here, we still must work to recover kinematic data. Kinematic data proved to be more challenging to process in the home environment than it did in the lab. There are several reasons for this:

First, the calibrations from data to day changed slightly in the home. This adds a considerable amount of work to the data processing problem, as calibrations are currently done largely by hand. It is therefore important that we work on some alternative, automatic calibration tools for home deployments in the future.

Second, the lighting, although it was fairly controlled, made for difficult processing of colors. In the home situation, light came from low overhead ceilings. These lights, moreover, were almost directly over the subject. As a result, the subject's head in images is often extremely white in color. The harsh overhead lighting also created very extreme shadows on the body of the subject. These shadows mean that individual colors on the jersey may have many manifestations in a given image. To ameliorate this situation, we must engineer color detectors that are conscious of shadows. We must also strengthen our kinematic tracker by engineering stronger constraints on the shapes and sizes of filtered limbs.

These problems, however, also present new opportunities. The difficulty of the tracking situation in the home challenges the limits of the tracking tools we have already developed, and inspires the integration of new tools and technologies into our system. Difficulties, then, outline our next iteration of software development, and ensure that we will continue to lie at the edge of what is possible.

Chapter 6

Conclusion

6.1 Summary

The main purpose of this research is to explore inexpensive and robust measurement devices that are sensitive to motor pathologies typical to stroke survivors, and which can theoretically operate invisibly and continuously in the home. We have offered three pieces of research in this dissertation, all of which provide incremental movement toward our long term goal.

First, we interviewed therapists to determine the ways in which they typically assess clients of therapy in the context of functional activity. We did this by executing a protocol analysis of therapists, which asked them to verbalize their decision making process as it was taking place. From this exercise, we learned that therapists tend to emphasize the hands and fingers of their clients more than any other body part. We also learned that they heavily emphasize smoothness in perceived motion when they are making their functional determinations.

Next, we looked to automate assessments with low cost devices and in a fashion that is consistent with therapists. Our technologies of choice were video cameras as well as objects instrumented with small arrays of force sensing resistors. Our results indicate that automation using low cost technologies is possible in a laboratory context. Moreover, we show that automatic tools can be built to predict functional scores based on the observation of lifts, manipulations and grasps that come from many different tasks. The measured statistics that were found to correlate most strongly with functional score included variance about the elbow and motion of the torso. The smoothness of force applied to the surface of objects, as measured in terms of the number of peaks in the force profiles, also strongly related to functional health.

Finally, we explored the possibility of making automated assessment in the home. To do this, we transported all of our measurement devices to the home of a stroke survivor and measured his functional desktop activity over the course of two weeks. Preliminary results indicate our measurement devices are capable of measuring performance change in this context in a way that is consistent with the opinions of human experts.

Future Work

The results we have produced thus far are encouraging, yet there is still a substantial amount of work left before we achieve our goal. Some of the work that must take place includes:

1. Refinement of the system. The system we have developed thus far is an interesting prototype, but it is far from being really useful. For one, it is not real time. Tracking results currently take slightly less than a second per frame to produce, which means that a second of video may take 30 seconds of processing time to analyze. The algorithms that we use, however, are all fairly simple and should be able to operate at 10 frames per second, conservatively. To develop a real time system, however, a substantial software engineering thrust is required. Moreover, calibration is still largely done by hand both across the cameras and objects. And finally, the system still requires markers and wires, which inhibits the "naturalness" of the environment in which it is situated. These elements must be removed for a mature instantiation of our system.

2. Testing with more subjects. In Appendix 1 are results from power analysis. These indicate that, in order to confidently claim that our system can produce accurate functional scores for stroke survivors, we need data from at least 25 stroke survivors. We must therefore test with more individuals. Moreover, preliminary work illustrates that our system has a tendency to under-estimate the scores of high functioning individuals and to over-estimate scores of low functioning individuals. It may be the case that three separate classifiers will produce the most accurate results. One would be for individuals at the high end of the AMAT range, one for those in the middle, and one at the low end.

3. The development of feedback paradigms. Finally, none of the work presented here is really useful unless it is used to provide feedback about motor performance to therapists or clients of therapy. What we have learned through protocol analysis, however, is that therapists may prioritize features like smoothness or motion of the hands. These features, however, are very difficult to measure with the devices that we have designed. There is an open question, then, as to how useful feedback that we can provide about the gross motion of the upper body will be. The utility of this information must be tested in the context of a biofeedback intervention, once our technologies are real-time and robust.

Bibliography

[1] URL http://www.immersion.com/3d/products/cyber_grasp.php. 1.2.3

[2] Heart disease and stroke statistics - 2004 update. www.americanheart.org. 1.1.1

[3] Medicare benefit policy manual, ch. 7. Centers for Medicare and Medicaid Services. http://www.cms.hhs.gov/manuals/. (document), 1.2, 4.1

[4] Medicare modernization act. Centers for Medicare and Medicaid Services, December 2003. http://www.cms.hhs.gov/medicarereform/. (document), 1.2, 4.1

[5] Heller A, Wade D, Wood V, Sunderland A, Hewer R, and Ward E. Arm function after stroke: measurement and recovery over the first three months. 13(4):354–62, 1999. 4.5.6, 4.6.3

[6] Sunderland A, Tinson D, Bradley L, and Hewer R. Arm function after stroke. an evaluation of grip strength as a measure of recovery and a prognostic indicator. *J Neurol Neurosurg Psychiatry*, 50, 1987. 4.5.6, 4.6.3

[7] A Agarwal and B Triggs. Learning to track 3d human motion from silhouettes. 2004. 1.2.2

[8] C Anderson, S Rubenach, C Mhurchu, M Clark, C Spencer, and A Winsor. Home or hospital for stroke rehabilitation? results of a randomized controlled trial : I: Health outcomes at 6 months. *Stroke*, 31:1024–1031, 2000. 1.1.2

[9] P Bach-y Rita. Theoretical and practical considerations in the restoration of function after stroke. *Top Stroke Rehabil.*, 8(3):1–15, 2001. 1.1.3

[10] P Bach-y Rita. Volume transmission and brain plasticity. *Evolution and Cognition*, 8: 115–22, 2002. 1.1.2

[11] P Bach-y Rita. Late post-acute neurologic rehabilitation: neuroscience, engineering and clinical programs. *Arch Phys Med Rehab*, 84(3):1100–1108, 2003. 1.1.2

[12] X Baptiste, X Law, X Pollock, H Polatajko, X McColl, and A Carswell. The canadian occupational performance measure. *World Federation of Occupational Therapy Bulletin.*, 28:47–51, 1993. 1.1.2, 1.1.3

[13] N Bennis and A Roby-Brami. Coupling between reaching movement direction and hand orientation for grasping. *Brain Res.*, 952(2):257–67, 2002. 1.1.1

[14] H Beppu, M Suda, and R Tanaka. Analysis of cerebellar motor disorders by visually guided elbow tracking movement. *Brain*, 107:787809, 1984. 4.3.1

[15] N Bernstein. The co-ordination and regulation of movements. *Pergamon Press*, 1967. 1.1.1, 4.3.1

[16] P Boissy, D Bourbonnais, MM Carlotti, D Gravel, and BA Arsenault. Maximal grip force in chronic stroke subjects and its relationship to global upper extremity function. *Clin Rehabil.*, 13(4):354–62, 1999. 4.5.6

[17] C Bregler and J Malik. Tracking people with twists and exponential maps. *Computer Vision and Pattern Recognition (CVPR)*, pages 8–15, 1998. 1.2.2

[18] B Brewer, R Klatzky, and Y Matsuoka. Effects of visual feedback distortion for the elderly and the motor-impaired in a robotic rehabilitation environment. *Proceedings of the 2004 IEEE International Conference on Robotics and Automation (ICRA '04)*, 2004. 1.2, 1.2.1, 1.2.3

[19] VB Brooks, JD Cooke, and JS Thomas. *Control of posture and locomotion*. 1973. 4.3.1

[20] S Brunnstrom. *Movement therapy in hemiplegia: a neurophysiological approach.* Harper and Row, New York, New York, 1st edition, 1970. 1.1.1, 2, 4.3.1

[21] C Burgar, P Lum, P Shor, and H Van der Loos. Development of robots for rehabilitation therapy: The palo alto va/stanford experience. *Journal of Rehabilitation Research and Development*, 37(6), 2000. 1.2.1

[22] MH Cameron. *Physical Agents in Rehabilitation: From Research to Practice*. W. B. Saunders, Philadelphia, PA, 2nd edition, 2003. 1.1.1, 1.1.3

[23] JR Carey. Manual stretch: effect on finger movement control and force control in stroke subjects with spastic extrinsic finger flexor muscles. *Arch Phys Med Rehabil*, 71(11): 888–94, 1990. 1.1.1, 1.2.3, 4.3.1

[24] A Cheung, S Baker, and T Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. *Computer Vision and Pattern Recognition (CVPR)*, pages 77–84, June 2003. 1.2.2, 2.3.1, 4.1

[25] J Cohen. *Statistical Power Analysis for the Behavioral Sciences.* Academic Press, New York, New York, 2nd edition, 1969. 3.6.2

[26] C Collin, D Wade, S Davies, and V Horne. The barthel adl index: a reliability study. *Int Disabil Stud.*, 10(2):61–3, 1988. 1.1.3

[27] Robert Collins, Omead Amidi, and Takeo Kanade. An active camera system for acquiring multi-view video. pages 517 – 520, September 2002. 2.2.1

[28] AE Colman. A companion encyclopedia of psychology. 1994. 3.2

[29] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. *Conf. on Comp. Vis. and Pat. Rec.*, pages 142–151, 2000. URL `citeseer.ist.psu.edu/comaniciu00realtime.html`. 2.4.1, 2.4.1

[30] L Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3): 297–33, 1951. 3.1

[31] Jack D, A andTremaine M Boian, R andMerians, G Burdea, S Adamovich, M Recce, and H Poizner. Virtual reality-enhanced stroke rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 9(3):308–318, 2001. 1.2.3

[32] WJG DeWeerdt and MA Harrison. Measuring recovery of arm - hand function in stroke patients: a comparison of the brunnstrom-fugl-meyer test and the action research arm test. *Physiother Can*, 37:65–70, 1985. 3.1

[33] D Druckman and R Bjork. 1.2

[34] DG Embrey, MR Guthrie, OR White, and J Dietz. Clinical decision making by experienced and inexperienced pediatric physical therapists for children with diplegic cerebral palsy. *Phys Ther*. 3.2

[35] A Ericsson and H Simon. Protocol analysis: Verbal reports as data. 1993. 3.2

[36] SE Fasoli, HI Krebs, M Ferraro, N Hogan, and BT Volpe. Does shorter rehabilitation limit potential recovery poststroke? *Neurorehabil Neural Repair.*, 18(2):88–94, 2004. 1.1.2, 4.1

[37] SE Fasoli, HI Krebs, R Hughes, J Stein, and N Hogan. Functionally-based rehabilitation: Benefits or buzzwrod? 1.2.1

[38] SE Fasoli, HI Krebs, J Stein, WR Frontera, and N Hogan. Effects of robotic therapy on motor impairment and recovery in chronic stroke. *Arch Phys Med Rehabil*, 84:477–82, 2003. 1.1.2, 1.2.1, 4.1

[39] P Felzenszwalb and D Huttenlocher. Efficient matching of pictorial structures. *International Conference on Pattern recognition (ICPR)*, 2000. 1.2.2, 2.3.1

[40] T Flash and N Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. 1985. 4.3.1

[41] D Forsyth. Workshop on human tracking. *Computer Vision and Pattern Recognition*, 2005. 1.2.2

[42] AR Fugl-Meyer, L Jaasko, I Leyman, S Olsson, and S Steglind. The post-stroke hemiplegic patient, a method for evaluation of physical performance. *Scand J Rehabil Med.*, 7 (1):13–31, 1975. 1.1.1, 1.1.1, 1.1.3, 3.3

[43] KH Funkesson, EM Anbcken, and Ek. Nurses' reasoning process during care planning taking pressure ulcer prevention as an example. a think-aloud study. *Int J Nurs Stud*, June 2006. 3.2

[44] J Gladman and N Lincoln. Follow-up of a controlled trial of domiciliary stroke rehabilitation (domino study). *Age and Ageing*, 23:9–13, 1994. 1.1.2

[45] J Gladman, D Whynes, and N Lincoln. Cost comparison of domiciliary and hospital-based stroke rehabilitation. *Age and Ageing*, 23:241–245, 1994. 1.1.2

[46] M Glanz, S Klawansky, and T Chalmers. Biofeedback therapy in stroke rehabilitation: a review. *J R Soc Med*, 90:33–39, 1997. 1.2.1

[47] I Guyon and A Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003. 4.3.2

[48] J Hermsdorfer, E Hagla, D Nowakb, and M Marquardta. Grip force control during object manipulation in cereberal stroke. *Clinical Neurology*, 114(3):915–929., 2003. 1.1.1, 1.2.3, 4.3.1

[49] N Hogan. Interactive robot therapy. Technical Report USA patent 5,466,213, 1995. 1.1.2

[50] M Holden. Virtual environments for motor rehabilitation: Review. *Cyberpsychology and Behavior*, 8(3), 2005. 1.2.1

[51] CL Hsieh, IP Hsueh, FM Chiang, and PH Lin. Inter-rater reliability and validity of the action research arm test in stroke patients. *Age and Ageing*, 27(2):107–13, 1998. 3.1, 3.5

[52] C Jadhav and V Krovi. A low-cost framework for individualized telerehabilitation. *Proceedings of IEEE Engineering in Medicine and Biology International Conference*, 2004. 1.2.1

[53] DL Jaffe, DA Brown, and et al Pierson-Carey, CD. Stepping over obstacles to improve walking in individuals with post-stroke hemiplegia. *Journal of Rehabilitation Research and Development*, 41(3):283292, 2004. 1.2.1

[54] MW Jaspers, T Steen, C van den Bos, and M Geenen. The think aloud method: a guide to user interface design. *Int J Med Inform.*, 73(11-12):781–95, 2004. 3.2

[55] M Jeannerod. *The Neural and Behavioural Organization of Goal-Directed Movements*. Oxford Science Publications, 1988. 1.1.2, 4.5.6

[56] M Johnston, B Pollard, V Morrison, and R MacWalter. Functional limitations and survival following stroke: psychological and clinical predictors of 3-year outcome. *Int J Behav Med.*, 11(4):187–196, 2004. 1.1.1

[57] M Johnston, KJ Wisneski, J Anderson, D Nathan, and R Smith. Development of adler: The activities of daily living exercise robot. *IEEE-EMBS Biomedical Robotics (BioRob 2006)*, page 881886, 2006. 1.2.1

[58] A Ju, M Black, and Y Yacoob. Cardboard people: A parameterized model of articulated motion. *Int. Conf. on Automatic Face and Gesture Recognition*, pages 38–44, 1996. 1.2.2

[59] LE Kahn, WZ Rymer, and DJ Reinkensmeyer. Adaptive assistance for guided force training in chronic stroke. *Engineering in Medicine and Biology Society*, 4:2722 – 2725, 2004. 1.2.1

[60] LE Kahn, ML Zygman, WZ Rymer, and DJ Reinkensmeyer. Effect of robot-assisted and unassisted exercise on functional reaching in chronic hemiparesis. *Engineering in Medicine and Biology Society*, 2001. 4.3.1

[61] I Kakadiaris and D Mataxas. Model-based estimation of 3d human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1453 – 1459, December 2000. 1.2.2

[62] B Kopp, A Kunkel, H Flor, T Platz, U Rose, R Mauritz, K Gresser, K McCulloch, and

E Taub. The arm motor ability test: Reliability, validity and sensitivity to change. *Arch. Phys. Med. Rehab.*, 78:615–620, 1997. 1.1.1, 1.1.3, 1.1.3, 1.1, 3.1, 3.5

[63] H Krebs, M Aisen, B Volpe, and N Hogan. Quantization of continuous arm movements in humans with brain injury. *Proc Natl Acad Sci USA*, 96:4645–4649, 1999. 1.1.1, 4.3.1, 4.5.6

[64] H Krebs, N Hogan, B Volpe, M Aisen, L Edelstein, and C Diels. Overview of clinical trials with mit-manus: a robot-aided neuro-rehabilitation facility. *Technology and Health Care*, 7(6):419–423, 1999. (document), 1.1.2, 1.2.1

[65] G Kriz, J Hermsdorfer, C Marquardt, and N Mai. Feedback-based training of grip force control in patients with brain damage. *Arch Phys Med Rehabil*, 76:653–9, 1995. 1.1.1, 4.3.1

[66] G Kurillo, M Gregoric, N Goljar, and T Bajd. Grip force tracking system for assessment and rehabilitation of hand function. *Technology and Health Care*, 13:113, 2005. 4.5.6

[67] G Kwakkel, R Wagenaar, Koelman, G T Lankhorst, and J Koetsier. Effects of intensity of rehabilitation after stroke: A research synthesis. *Stroke*, 28:1550–1556, 1997. 1.1.2

[68] JW Lance. *Spasticity Disordered Motor Control*. Year Book Publishers, Chicago, 1980. 1.1.1

[69] E Lawrence, C Coshall, R Dundas, J Stewart, A Rudd, R Howard, and C Wolfe. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. *Stroke*, 32:1279–1284, 2001. 1.1.1, 1.1.1, 1.1.1

[70] H LeBreck and D Holzemer. The effect of verbalization of cognitive processes on clinical decisionmaking. *Res Nurs Health*, 12:187–193, 1989. 3.2

[71] KC Lee, L Carson, E Kinnin, and V Patterson. The ashworth scale: a reliable and reproducible method of measuring spasticity. *Journal of Neurological Rehabilitation*, 3(4): 205–9, 1989. 3.3

[72] M Lee and R Nevatia. Dynamic human pose estimation using markov chain monte carlo approach. *In IEEE Workshop on Motion and Video Computing*. 1.2.2

[73] M Levin. Interjoint coordination during pointing movements is disrupted in spastic hemiparesis. *Brain*, 119:281–293, 1996. 1.1.1

[74] M Levin, J Desrosiers, D Beauchemin, N Bergeron, and A Rochette. Development and validation of a scale for rating motor compensations used for reaching in patients with hemiparesis: the reaching performance scale. *Phys Ther.*, 84(1):8–22, Jan 2004. 1.1.1, 4.5.6

[75] S Li, ML Latash, GH Yue, V Siemionow, and V Sahgal. The effects of stroke and age on finger interaction in multi-finger force production tasks. *Clin Neurophysiol*, 114:1646–55, 2003. 1.1.1, 4.5.6

[76] J Liepert, H Bauder, W Miltner, E Taub, and C Weiller. Treatment-induced massive cortical reorganization after stroke in humans. *Stroke*, 31, 2000. 1.1.2

[77] J Liepert, H Bauder, W Miltner, E Taub, and C Weiller. Treatment-induced massive cortical reorganization after stroke in humans. *Stroke*, 31:12101216, 2002. 1.1.2

[78] J Liepert, W Miltner, H Bauder, M Sommer, C Dettmers, E Taub, and C Weiller. Motor cortex plasticity during constraint-induced movement therapy in stroke patients. *Neurosci Lett*, 250:5–8, 1998. (document), 1.1.2

[79] K Lin, C Wu, L Tickle-Degnen, and W Coster. Enhancing occupational performance through occupationally embedded exercise: A meta-analytic review. *The Occupational Therapy Journal of Research*, 17:2547, 1997. (document), 1.1.2

[80] JM Linacre, AW Heinemann, BD Wright, CV Granger, and BB Hamilton. The structure and staility of the functional independence measure. *Arch Phys Rehabil*, 75:127–132, 1994. 1.1.3

[81] JW Little, P Micklesen, R Umlauf, and Brittell C. Lower extremity manifestations of

spasticity in chronic spinal cord injury. *Am J Phys Med Rehabil*, 68:3236, 1989. 1.1.1, 1.1.3

[82] P Lum, D Reinkensmeyer, R Mahoney, WZ Rymer, and C Burgar. Robotic devices for movement therapy after stroke: Current status and challenges to clinical acceptance. *Top Stroke Rehabil*, 8:4053, 2002. 1.2.1

[83] PS Lum, C Burgar, and P Schor. Evidence for strength imbalances as a significant contributor to abnormal synergies in hemiparetic subjects. *Muscle and Nerve*, 27(2):211–221, 2002. 1.1.1

[84] X Luo, R Kenyon, and D Kamper. Vr post-stroke hand opening rehabilitation: An approach utilizing virtual reality, body orthosis and pneumatic device. *International Conference on Aging, Disability and Independence*, 2006. 1.2.3

[85] H Ma, C Trombly, and C Robinson-Podolski. The effect of context on skill acquisition and transfer. *Am J Occup Ther.*, 53(2):138–44, Mar-Apr 1999. (document), 1.1.2, 1.2, 4.5.6

[86] L MacClellan, D Bradham, J Whitall, B Volpe, D Wilson, J Ohlhoff, C Meister, N Hogan, H Krebs, and C Bever. Robotic upper-limb neurorehabilitation in chronic stroke patients. *J Rehab. Res. and Dev.*, 42(6):717722, 2005. 1.1.2

[87] N Maclean, P Pound, C Wolfe, and A Rudd. The concept of patient motivation: a qualitative analysis of stroke professionals' attitudes. *Stroke.*, 33(2):444449, 2002. 1.1.2

[88] AL Melin, S Hakansson, and LO Bygren. The cost-effectiveness of rehabilitation in the home: a study of swedish elderly. *Am J Public Health*, 83(3):356362, 1993. 1.1.2

[89] S Michaelsen, S Jacobs, A Roby-Brami, and M Levin. Effect of trunk restraint on the recovery of reaching movements in hemiparetic patients. *Stroke*, 32:1875–1883, 2001. 4.5.6

[90] S Michaelsen and M Levin. Short-term effects of practice with trunk restraint on reaching

movements in patients with chronic stroke: A controlled trial. *Stroke*, 35:1914, 2004. 1.1.1, 4.5.6

[91] E Michels. Synergies in hemiplegia. *Clin. Management, vol. 1, pp.*, 1:9–16, 1981. 1.1.1, 4.3.1

[92] WH Miltner, H Bauder, M Sommer, C Dettmers, and E Taub. Effects of constraint induced movement therapy on patients with chronic motor deficits after stroke: a replication. *Stroke*, 30:586–592, 1999. 1.1.2, 4.1

[93] HW Mims. Electromyography in clinical practice. *South Med J*, 49:804, 1956. 1.2.1

[94] T Mitchell. *Machine Learning*. McGraw Hill, 1997. 4.3.1

[95] G Mori and J Malik. Recovering 3d human body configurations using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1052–1062, 2006. 1.2.2

[96] D Nowak, J Hermsdorfer, and H Topka. Deficits of predictive grip force control during object manipulation in acute stroke. *Journal of Neurology*, 250(7):85–60, 2003. 1.1.1, 4.3.1, 4.5.6

[97] SJ Page, Levine P Sisto, S, and RE McGrath. Efficacy of modified constraint-induced movement therapy in chronic stroke: a single-blinded randomized controlled trial. *Arch Phys Med Rehabil*, 85:1418, 2004. 1.1.2

[98] D. Pai, VanDerLoo, A. Sadhukhan, and P. Kry. The tango: a tangible tangoreceptive whole-hand human interface. *World Haptics Symposium*, 2005. 1.2.3

[99] C Painter. Web based motion control for physical rehabilitation. *Master Thesis, University of California at Irvine*, 2000. 1.2.1, 4.1

[100] Polhemus. http://www.polhemus.com/. Technical report. 1.2.1

[101] N Pollock. Client-centred assessment. *American Journal of Occupational Therapy*, 47: 298–301, 1993. (document), 1.1.2

[102] D. Ramanan. Learning to parse images of articulated objects. *Neural Info. Proc. Systems (NIPS)*, 2006. 1.2.2

[103] D. Ramanan and DA. Forsyth. Finding and tracking people from the bottom up. *Computer Vision and Pattern Recognition (CVPR)*, June 2003. 1.2.2, 2.2.1, 2.3.2, 4.1

[104] D. Ramanan, DA Forsyth, and A Zisserman. Strike a pose: Tracking people by finding stylized poses. *Computer Vision and Pattern Recognition (CVPR)*, 2005. 1.2.2, 2.3.2

[105] G Rankin and M Stokes. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clinical Rehabilitation*, 12(3):187–199, 1998. 3.4

[106] D Reinkensmeyer, L Kahn, M Averbuch, A McKenna-Cole, B Schmit, and Z Rymer. Understanding and treating arm movement impairment after chronic brain injury: Progress with the arm guide. *Journal of Rehabilitation Research and Development*, 37(6), 2000. 1.2.1

[107] D Reinkensmeyer, C Pang, J Nessler, and C Chris C. Painter. *Integration of Assistive Technolgy in the Information Age*. IOS Press, 2001. 1.2, 1.2.1, 4.1

[108] DJ Reinkensmeyer, JPA Dewald, and WZ Rymer. Guidance based quantification of arm impairment following brain injury: A pilot study. *IEEE Trans Rehabil Eng*, 7:1–1, 1999. 1.1.1, 4.5.6

[109] DJ Reinkensmeyer, BD Schmit, and WZ Rymer. Assessment of active and passive restraint during guided reaching after chronic brain injury. *Ann Biomed Eng*, 27:805–14, 1999. 1.1.1

[110] Liu Ren, Alton Patrick, Alexei A. Efros, Jessica K. Hodgins, and James M. Rehg. A data-driven approach to quantifying natural human motion. pages 1090–1097, 2005. 4.3.1, 4.5.6

[111] A Roby-Brami, A Feydy, M Combeaud, E Biryukova, B Bussel, and M Levin. Motor compensation and recovery for reaching in stroke patients. *Acta Neurol Scand.*, 107(5):

369–81, May 2003. 1.1.1, 4.3.1

[112] A Roby-Brami, S Jacobs, N Bennis, and M Levin. Hand orientation for grasping and arm joint rotation patterns in healthy subjects and hemiparetic stroke patients. *Brain Res.*, 969 (1-2):217–29, Apr 2003. 1.1.1

[113] B Rohrer, S Fasoli, H Krebs, R Hughes, B Volpe, W Frontera, J Stein, and N Hogan. Movement smoothness changes during stroke recovery. *The Journal of Neuroscience*, 22 (18), 2002. 4.3.1, 4.3.1, 4.5.6

[114] FD Rose, EA Attree, and BM et al Brooks. Training in virtual environments: transfer to real world tasks and equivalence to real task training. *Ergonomics*, 43:494–511, 2003. 1.2.1

[115] K Salter, N Foley, S Bhogal, and R Teasell. The most frequently cited outcome measures among randomized controlled evaluating stroke rehabilitation therapies. *LHRI Rehabilitation and Geriatric Care Research Day*, November 2003. 1.1.3

[116] R Sanchez, J Liu, S Rao, P Shah, R Smith, T Rahman, S Cramer, J Bobrow, and D Reinkensmeyer. Automating arm movement training following severe stroke: Functional exerciseswith quantitative feedback in a gravity-reduced environment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(3), September 2006. 1.2, 1.2.1

[117] RA Scheidt, DJ Reinkensmeyer, MA Conditt, WZ Rymer, and FA Mussa-Ivaldi. 1.2.1

[118] SV Scrimshaw and Maher C. Responsiveness of visual analogue and mcgill pain scale measures. *J Manipulative Physiol Ther.*, 24(8):501–4, 2001. 3.3

[119] H Sidenbladh and M Black. Learning the statistics of people in images and video. *International Journal of Computer Vision*, 2003. 1.2.2, 2.3.1, 2.3.2, 4.1

[120] H Sidenbladh, M Black, and D Fleet. Stochastic tracking of 3d human figures using 2d image motion. *ECCV*, 2:702–718, 2000. 1.2.2, 2.2.1

[121] L Sigal, S Bhatia, S Roth, M Black, and M Isard. Tracking loose-limbed people. *Computer Vision and Pattern Recognition (CVPR)*, 2004. 1.2.2, 4.1

[122] ER Skidmore. *Functional assessment after stroke: Examining outcome models using the International Classification of Functioning, Disability and Health*. PhD thesis, University of Pittsburgh, 2003. 1.1.1, 1.2.3

[123] N Smaby, ME Johanson, B Baker, DE Kenney, WM Murray, and VR Hentz. Identification of key pinch forces required to complete functional tasks. *J Rehabil Res Dev.*, 41(2):215–24, 2004. 1.1.1

[124] S Sminchicescu and B Triggs. 1.2.2

[125] Smith. *Nature*, 2000. 4.5.6

[126] C Spearman. The proof and measurement of association between two rings. *Amer. J. Psychol.*, 15:72–101, 1904. 3.1

[127] Tomas Svoboda, Daniel Martinec, and Tomas Pajdla. A convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4), 2005. 2.3.1

[128] Vicon Motion Caputer Systems. http://www.vicon.com/. Technical report. 1.2.1

[129] H Tan, Lu, and A Pentland. The chair as a novel haptic user interface. *Proceedings of the Workshop on Perceptual User Interfaces*, 1997. 1.2.3

[130] E Taub. Overcoming learned nonuse: A new approach to treatment in physical medicine. In A. R. Seifert J. G. Carlson and N. Birbaumer, editors, *Clinical Applied Psychophysiology*, pages 185–220. Plenum Press, 1994. 1.1.2, 4.3.1

[131] E Taub and J Crago. Overcoming learned nonuse: A new behavioral approach to physical medicine. In T Kikuchi, H Sakuma, I Saito, and K Tsubo, editors, *Biobehavioral self-regulation: Eastern and western perspectives*, pages 2–9. Springer Verlag, 1995. 1.1.2, 4.3.1

[132] E Taub, PS Lum, P Hardin, V Mark, and G Uswatte. Automated delivery of ci therapy with reduced effort by therapists. *Stroke*, 36:1301, 2005. 1.1.3, 1.2.1, 3.1, 4.2, 5.1

[133] E Taub, NE Miller, TA Novack, EW Cook, WC Fleming, CS Nepomuceno, JS Connell, and JE Crago. Techniques to improve chronic motor deficit after stroke. *Arch Phys Med Rehabil*, 74:347–54, 1993. 1.1.2, 4.1

[134] E Taub, G Uswatte, and DM Morris. Improved motor recovery after stroke and massive cortical reorganization following constraint-induced movement therapy. *Phys Med Rehabil Clin N Am.*, 14(1):277–91, 2003. (document), 1.1.2

[135] E Taub, G Uswatte, and R Pidikiti. Constraint-induced movement therapy: a new family of techniques with broad application to physical rehabilitation–a clinical review. *J Rehabil Res Dev*, 36:237–51, 1999. 1.1.2

[136] C Trombly. Occupational therapy of physical dysfunction. 1995. 1.1.2

[137] TE Twitchell. Restoration of motor function following hemiplegia in man. *Brain*, 74: 443–480, 1951. 1.1.1

[138] A Twycross and L. Powls. How do children's nurses make clinical decisions? two preliminary studies. *J Clin Nurs.*, 15(10):1324–35, 2006. 3.2

[139] R Urtasun, D Fleet, and P Fua. Temporal motion models for monocular and multiview 3d human body tracking. *Computer Vision and Image Understanding special issue Modeling People, (CVIU)*, 2006. 1.2.2, 2.3.2

[140] J van der Lee, H Beckerman, G Lankhorst, and L Bouter. The responsiveness of the action research arm test and the fugl-meyer assessment scale in chronic stroke patients. *J Rehabil Med*, . 2.1

[141] J van der Lee, V De Groot, H Beckerman, R Wagenaar, and G Lankhorst. The intra- and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke. *Arch Phys Med Rehabil*, . 3.1, 3.5

[142] F Van der Weel, A Van der Meer, and D Lee. Effect of task on movement control in cerebral palsy: Implications for assessment and therapy. *Developmental Medicine and Child Neurology*. 1.1.2

[143] B Volpe, H Krebs, and N Hogan. Is robot-aided sensorimotor training in stroke rehabilitation a realistic option? *Curr Opin Neurol*, 14:745–752, 2001. 1.1.2, 1.1.3, 1.2.1

[144] R Wagenaar, O Meijer, P van Wieringen, D Kuik, G Hazenberg, and J Lindeboom. The functional recovery of stroke: A comparison between neuro-developmental treatment and the brunnstrom method. *Scand J Rehabil Med*, 22:1–8, 1990. 1.1.1, 3.1

[145] NT Watts. 3.2

[146] GF Wittenberg, R Chen, RK Ishii, KO Bushara, E Taub, LH Gerber, M Hallett, and LG Cohen. Constraint-induced therapy in stroke: magnetic-stimulation motor maps and cerebral activation. *Neurorehabilitation and Neural Repair.*, 17:4857, 2003. 1.1.2

[147] S Wolf, P Catlin, M Ellis, A Archer, B Morgan, and A Piacentino. Assessing wolf motor function test as outcome meansure for research in patients after stroke. *Stroke*, 32:1635–1639, 2001. 1.1.1, 1.1.3

[148] C Wren, A Azarbayejani, T Darrell, and A Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 780–785, July 1997. 2.3.2

[149] C Wu, CA Trombly, K Lin, and L Ticke-Degnen. Effects of object affordances on reaching performance in persons with and without cerebrovascular accident. *Am J Occup Ther.*, 52 (3):179–87, Mar 1998. 1.1.2, 1.2, 4.5.6

[150] K Yin and D Pai. Footsee: an interactive animation system. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 329–338, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association. ISBN 1-58113-659-5. 1.2.3

[151] Z Zhang. Flexible camera calibration by viewing a plane from unknown orientations. *International Conference on Computer Vision (ICCV),*, 1999. 2.2.1