PROTOCOL ANALYSIS AS A TOOL TO EXPLORE THE FUNCTIONAL ASSESSMENT HEURISTICS OF EXPERTS

Sonya Allin[1], Ph.D. and Emily Eckel[2], O.T.D.
[1]*University of Toronto, Toronto, ON, Canada*
[2]*Chatham University, Pittsburgh, PA, USA*

ABSTRACT

This paper details an effort to determine heuristics used in practice by expert therapists to functionally assess the upper extremity of stroke survivors. The goal is to be able to compare these heuristics to heuristics used by a computer system that performs similar assessments, automatically. Three expert therapists were asked to verbalize their decision-making process as they made assessments of eight stroke survivors' upper extremity function using the Arm Motor Ability Test (AMAT). Transcripts were coded to determine which specific parts of the upper extremity and/or trunk were focused on in order to arrive at assessment scores. Results indicate that therapists were consistent in their heuristics, and prioritized distal features (like motion of the fingers) when making assessment decisions over proximal ones. The automated system, by contrast, prioritized proximal movement features (like motion of the torso) over distal ones to make the same assessments.

INTRODUCTION

Functional recovery after stroke can take place for many months, if not years [1], and long after an individual has been discharged from therapeutic services. Comprehensive monitoring of functional recovery to each survivor's maximal potential is not generally possible because of its cost. Much information about the sequence and character of long-term return of motor control is therefore hidden from clinical view, along with opportunities for clinicians to positively influence it over the long term.

In an effort to affordably enable long-term functional observations after stroke, recent research has explored the use of automated camera-based monitoring systems [2,3]. In order to ensure the system detailed in [2,3] is robust, it focuses on the motion of relatively large and easily observed parts of the body, like the torso. Motion of such proximal parts has, in fact, been shown to capture functional status after stroke in numerous studies with commercial motion capture [4,5]. Even when measured with less spatially precise but more affordable computer vision algorithms, torso motion has been found to correlate strongly with stroke survivors' Arm Motor Ability Test (AMAT) [3] and Fugl-Meyer Assessment (FMA) [2] scores.

In this paper, we explore the degree to which proximal statistics are being attended to by human therapists during the assessment act. Our purpose is to guide future iterations of automated assessment systems, and to ensure they are sensitive to those features clinicians actually use in practice.

To explore expert assessment heuristics, we have conducted a concurrent verbal protocol analysis [6]. This kind of analysis involves asking experts to verbalize their cognitive processes as they perform their assessment task. The resulting verbal reports are coded and analyzed as a source of evidence about the experts' cognition [7].

In the study we present here, three expert occupational therapists were asked to verbalize their clinical reasoning process as they made assessments using the AMAT. Transcripts were then coded to determine which parts of the body were focused on in order to arrive at a particular assessment score. The AMAT was chosen because it is an explicitly functional arm assessment that features high inter-rater and intra-rater reliability, as well as strong internal consistency [8]. It was also chosen so that direct comparisons with automated AMAT assessment heuristics [2] could be made.

THE AMAT

The Arm Motor Ability Test (AMAT) is an upper body assessment of stroke survivors developed in 1987 [8]. All 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. In the work presented here, six AMAT tasks were performed; these tasks were chosen so as to be consistent with [2]. A perfect score on these tasks would amount to 70.

METHODS

Table 1:Stroke Survivor Demographics. 'YPS' means 'years post stroke'. 'LES SITE' is the site of the subject's lesion, 'DOM' is his or her hand dominance. 'AMAT' are AMAT scores averaged across participating therapists and 'FMA' are Fugl-Meyer scores for the upper extremity (from one therapist).

ID	***Age***	***YPS***	***LES. SITE***	***DOM***	***AMAT***	***FMA***
1	75	2	Left	Right	68	65
2	60	2	Right	Right	35	44
3	47	22	Left	Right	42	46
4	82	12	Right	Right	50	53
5	64	4	Left	Right	11	13
6	58	13	Right	Right	56	64
7	78	7	Left	Right	34	39
8	63	35	Left	Right	60	64

In a laboratory environment, six tasks on the AMAT were administered to eight stroke survivors. Basic demographics of the stroke survivor are listed in Table 1. As the survivors performed the AMAT tasks, video of the activity was collected from eight synchronized cameras. Each camera was located approximately 1 meter from the seated stroke survivor, and all cameras focused on the table and upper body. Two cameras explicitly focused on hands. The positions of objects on the table were standardized according to AMAT instructions.

Three expert occupational therapists were then asked to assess the stroke survivors based on the video recordings. All participating therapists 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 tool.

Figure 1: Video of a stroke survivor performing a task on the AMAT. Therapists were allowed to view each task from eight camera angles.

To assess the stroke survivors' functional performance on the selected tasks, therapists made use of a computer interface; this is illustrated in Figure 1. Clicking on any one of the thumbnails at the top of the interface played video corresponding to that viewpoint in the main screen. Therapists were allowed to replay videos as frequently as they liked and could take as long as they liked when making their assessments.

As assessments were made, 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 an investigator to continue to speak. Audio recordings were made of all verbalizations; these recordings were subsequently transcribed for analysis.

DATA ANALYSIS

To determine inter-rater agreement between therapists, inter-class correlations (ICCs) between therapists' raw AMAT scores were computed [9].

Table 2: Codes assigned to the transcripts

Code	***Transcript Example***
Hands	"He doesn't extend all his *digits* into a full lumbrical *grip*."
Arm	"He is abducting his *shoulder*."
Head/Torso	"He is really moving his *head*."

To analyze therapists' verbalizations, codes representing various upper body parts were assigned to transcripts. These codes are shown in Table 2. In order to verify the consistency of the codes, two judges were asked to independently apply them to transcripts. Transcripts of eighteen assessments were used for coding verification; this represented slightly more than 10% of all the verbalizations. Cohen's Kappa was used to measure agreement between the two judges [10].

Histograms of codes were analyzed to determine features of interest across therapists as well as the consistency of therapists. ICCs between therapists' raw histograms were computed to determine agreement between assessors.

RESULTS

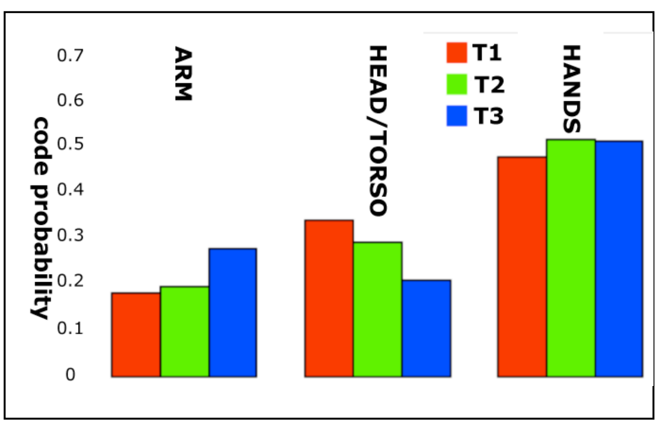

Figure 2: Histograms of codes assigned to transcripts. T1, T2 and T3 each represent a different therapist.

The ICC that was computed across therapists on AMAT scores was 0.96, indicating strong overall agreement as to assessment scores. This ICC is only slightly lower than the correlation reported in [8].

The Cohen's Kappa value relating codes from the two independent raters was 0.52. Disagreements resulted when, for example, discussion of "grasping" was interpreted as involving the arm by one judge and the hand by another. Despite these kinds of disagreements, the computed Kappas indicate moderate agreement between the judges in the use of codes.

The overall frequencies of codes applied to transcriptions are illustrated in Figure 2. The ICC value relating the therapists' distributions of body specific codes was 0.85.

Computed ICCs related to histograms indicate that therapists were, for the most part, relatively consistent in their emphasis on specific body parts during assessments. 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.

DISCUSSION

Protocol analysis proved to be a useful tool to explore the assessment heuristics employed by therapists. Results indicate experienced therapists to be relatively consistent not only in their AMAT assessment scores, but in their focus on prioritized parts of the upper body during assessments.

The emphasis of the therapists in this experiment seems to have been focused on stroke survivors' hands and fingers. The emphasis of the automated system, by comparison, is on the torso and shoulder. The fact that these two different assessment heuristics yield correlated assessment scores is by no means surprising. It is well known that the motion of the hand, torso and arm are deeply coupled during a reach [11]. Impairment of the hands and fingers correlates with disability after stroke [12]; excessive torso displacement during a reach similarly reflects low functional scores [13].

The fact that human therapists arrive at assessment scores using a different heuristic than our automated system, however, has implications related to feedback from the automated system. If therapists prioritize observations relating to the hands as they make assessments, perhaps an automated assessment device should do the same. Minimally, an automated assessment system should capture the information about features of interest to therapists, even if they are not used to make computerized assessment decisions.

The protocol analysis results reported here will guide future iterations of our computerized assessment system [2,3]. This system will be designed to provide meaningful feedback about long-term functional change to clinicians. Our hope is to provide these assessments affordably and automatically, and for the complete duration of time that functional recovery can take place. We seek to measure this recovery at least as accurately and precisely as do expert humans, using an instrument like the AMAT.

REFERENCES

[1] 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.

[2] S Allin, N Baker, E Eckel, D Ramanan. "Robust Tracking of the Upper Limb for Functional Stroke Assessment." IEEE Transactions on Neural Systems & Rehabilitation Engineering (NSRE). 5(18): 542-550, 2010.

[3] S Allin, D Ramanan. "Assessment of Post Stroke Functioning Using Machine Vision." IAPR Machine Vision and Applications (MVA), Tokyo, Japan, May 2007.

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

[5] 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, 2003.

[6] K A Ericsson. Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative task. In K A Ericsson, N Charness, P Feltovich, and R R Hoffman (Eds.). Cambridge handbook of expertise and expert performance. p. 223-242. Cambridge, UK: Cambridge University Press. 2006.

[7] K A Ericsson and H Simon. Protocol Analysis: Verbal Reports as Data. Lodon: MIT Press. 1984.

[8] 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.

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

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

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

[12] A Sunderland, D Tinson, L Bradley, and R Hewer. Arm function after stroke. An evaluation of grip strength as a measure of recovery and a prognostic indicator. J Neurol Neurosurg Psychiatry, 52(11):1267-72,1989.

[13] S Michaelsen, S Jacobs, A Roby-Brami, and M Levin. Compensation for distal impairments of grasping in adults with hemiparesis. Exp Brain Res. 57(2):162-73, 2004.