Automated Scoring of Rehabilitative Tests with Singular Spectrum Analysis

[1,2]Tracey K. M. Lee, [3]K. H. Leo,[4]Saeid Sanei and [5]Effie Chew

[1]School of Electrical and Electronic Engineering, Singapore Polytechnic, Singapore: tlee@sp.edu.sg
[2]School of Information Technology, Monash University, Malaysia
[3]School of Mechanical and Aeronautical Engineering, Singapore Polytechnic, Singapore: leokeehao@sp.edu.sg
[4]Faculty of Engineering and Physical Sciences, University of Surrey: s.sanei@surrey.ac.uk
[5]Division of Neurology, University Medicine Cluster, National University Hospital, Singapore: effie_chew@nuhs.edu.sg

ABSTRACT

In rehabilitation, continual assessment of those with disabilities is needed to determine the effectiveness of therapy and to prescribe the regimen and intensity of future treatment. Conducting assessments is challenging - there is a need to maintain objectivity and consistency across time. Also, repetitious tests can lull the assessor into lower levels of alertness. These motivate for automated scoring of rehabilitative tests.

In this paper, we describe our work in automating the widely used and accepted Action Research Arm Test. We focus on the grasp subtest which employs a cube into which we embed sensors. Previously we have used live patient simulators and now the full set of patient trials have been completed.

We employ Singular Spectrum Analysis on the signals, for which the resulting eigenvalues are then selected in a principled way to aid in signal filtering. The results show encouraging promise in our quest for automated scoring.

Index terms - *Singular spectrum analysis, subspace analysis stroke, rehabilitation, accelerometer, instrumented objects, automatic scoring*

1. INTRODUCTION

A patient who is unable to execute the activities of daily living (ADL) creates a challenging environment as there is the need for constant medical attention, resources and often a caregiver to administer whatever aid is necessary. The upper limbs determine the feeding and cleaning functions, the rehabilitation of which requires a customised regimen of exercises, tailored to the needs of the person. Also, progress needs to be monitored in order to assess the effectiveness of the treatment. But these tasks are labour intensive as trained therapists have to record and keep track of the outcomes of repetitive exercises. Besides, the lack of clinical skills in homes reduce the intensity of the rehabilitation process [1].

With the widespread use of information technology, a preferred solution is to use this to automate and monitor the tests and the exercises. Furthermore, by using tests that are widely accepted by the industry we take advantage of the ratification process ensuing from their widespread use. Another advantage is that this provides a point of focus in discussions with clinical staff familiar with the methodology and technologists seeking to automate the tests.

The Action Research Arm Test (ARAT) formulated by Lyle [2] is a performance test designed to assess recovery of upper limb function after damage to the cerebral cortex. It can be used to check on progress in treatment as well as evaluate the effectiveness of treatment. Additionally, it is reliable and quickly administered. Basically, it consists of various objects to be moved in a specified manner in assessing grasp, grip and pinch movements, which are used in the ADL. In trying to capture such fine movement, it is difficult to use methods which measure signals from sensors directly attached to the subject. These are intrusive, may impede motion, or if using video - which due to the nature of the signal, gives inherently noisy readings and is susceptible to the vagaries of lighting and occlusion effects.

In our approach, we embed sensors to the objects being handled in a rehabilitative setting for the following benefits: i) It is capable of sensing fine motion and pressure exerted by a person and ii) There is no need to mount sensors on the body of a person.

In Section 2 we present the motivation for our approach and cover background material. Section 3 describes our clinical setup. The theory for our signal analyses is covered in Section 4. Then the results of our experiments are presented in Section 5 before we summarize and conclude in Section 6.

2. ASSESSING LIMB FUNCTION AND MOTION

In this section we present the clinical motivation for our work, presenting the case for using instrumented objects used in standardised clinical tests.

In formulating tests of limb function and movement, enforcing a protocol for their administration provides for objective and quantitative measurements.

Currently, several of these tests use visual based scoring which introduces a degree of subjectivity and an inability to perceive subtle motions. Furthermore, the repetitious nature of the activities in the tests induce a measure of inattentiveness. This motivates for automating and monitoring these tests through electronic means by instrumenting the objects used in these tests, which is an ongoing field of research.

By using tests that are widely accepted by clinicians, which have been ratified through years of deployment, provides a point of focus and discussion.

Yozbatiran et al. [3] made further standardisations to the ARAT by specifying the placements of the objects and the dimensions of the furniture supporting these objects. They also quantified the scoring by taking note of the timing and quality of the movement performed by a person but the quality was rather descriptive in nature as well, befitting a visual based scoring system.

Lee et al. [4] described work done with the instrumented device described in this paper using healthy patient simulators. Portions of their paper have been reproduced here for the sake of continuity in discussion.

The analysis of biomedical signals benefit from decomposition into constituent parts to identify features of interest and recent frequency analyses using data driven decomposition processes have been employed successfully. Here, Singular Spectrum Analysis (SSA) has been used to analyse naturally occurring physical phenomena and only recently it has been applied to biological signals. The forms of the constituent signals it produces are not constrained to sinusoids, so it produces readily interpretable constituent signals such as trends, periodic data and noise from short

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

noisy signals.

In our work we combine two types of sensors not often used together, namely accelerometers and force sensors. This has the following benefits: i) it is capable of sensing fine motion and pressure exerted by a person and ii) there is no need to mount sensors on the body of a person. The next section describes our setup.

3. EXPERIMENTAL SETUP AND TRIAL CONDUCT

Here we describe how we implement the ARAT and show some preliminary results. In this paper, we focus on Test4 of the ARAT Grasp Subtest which involves the grasping of a wooden block shaped as a cube with a dimension of 7.5 cm. This object which we will call the Cube, is moved from a specified point directly to a target. The three main components of our instrumented object system as shown in Fig. 1 are:

i) A set of resistive sensors used for measuring forces exerted on the faces of the Cube.
ii) A tri-axial accelerometer for acceleration measurements.
iii) A microcontroller converting the force sensor and accelerometer readings, sending the data to a workstation.

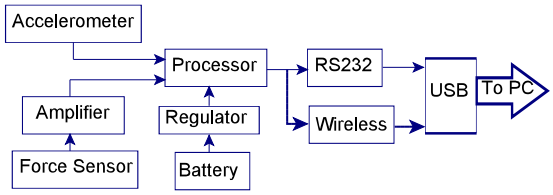

Fig. 1. Hardware block diagram of the embedded sensor system in the Cube. Dotted lines indicate optional portions.

The sensor readings are taken at a rate of 30 samples/sec so that a maximum frequency of 15 Hz can be reliably recorded. Unlike our previous work, we do not perform and pre-filtering of data to avoid missing important information.

3.1 ARAT scoring and test subjects

In Fig. 2 we see the Cube being grasped by a right handed person moving it from the lower, hand silhouette to the higher black target, the trajectory shown by a broken line. This action has to be completed in a given time. The Cube is held upright and the motion is to be what a healthy person would exert without undue duress. We would expect this task to be completed smoothly, with a minimum of energy. In Fig. 2, note that the non-grasping (left) hand is used as support, so the force exerted on the table can also provide useful data for assessment.

The ARAT scoring uses a four point scoring scale, from 3 for satisfactory completion to 0 which is non-completion. Yozbatiran et al. [3] attempted to make the rating more objective. We will only summarize comments from their paper in the interests of space. A score of 3 indicates completion of the task within 5 seconds with appropriate hand, arm and posture movements which are detailed in a table in the paper.

A score of 2 is given when the subject completes the task but does so "with great difficulty and/or takes abnormally long time" to fully complete the task, taking from 5 to 60 seconds.

For a score of 1 which indicates partial completion, the timing would be greater than 60 seconds. Also just being able to grasp, hold and lift the Cube would be sufficient to warrant this score.

However a score of 0 indicates that the subject is unable to perform any part of the task within 60 seconds. The inability of the hand to grasp the Cube within the time period would count towards this. Besides this, if the subject does not use the fingers to grasp the Cube or use another hand or mechanical support to manipulate the Cube would be considered for this score.

A total of 34 patients who have had a history of stroke and undergone rehabilitation were recruited for the trial which was conducted over a period of 60 days. The test was performed in a hospital setting. Data from patients with a score of zero could not be used as the Cube could not be lifted up properly. They participated in other subtests.

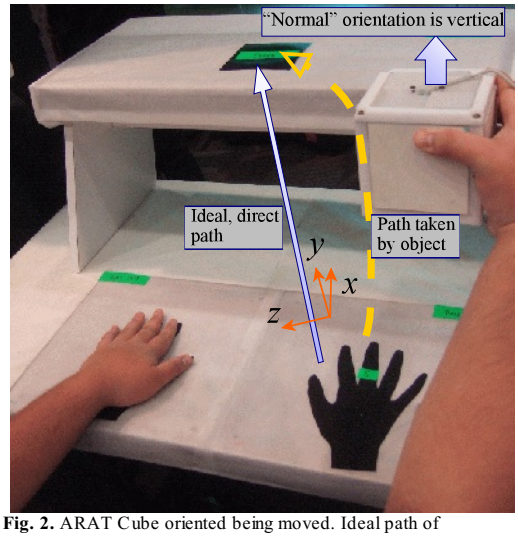

Fig. 2. ARAT Cube oriented being moved. Ideal path of object compared to actual path taken

Each patient would execute a series of ARAT motions in one session, up to 3 times per action, if possible. For each session, data is continuously recorded and manually segmented later into various trials. An important point is that the score is given to a patient on a *session* basis, and thus some sort of averaging is done on the trials. Furthermore, the sessions were run over a period of time and scored by different therapists, so there is some variability in the scores, even with briefings conducted.

In this case, we had a total of 225 signals from 26 patients who could participate in the ARAT Test4. Since we are only interested in the *z*-axis, we look at only 78 signals. It should be noted that three patients could only complete two trials. The distribution of the patients and their scores were: Score 3 - 11, score 2 - 13, score 1 - 2 and score 0 - 6.

3.2 Qualitative results

In Fig. 3 we show the force sensor signals obtained for two subjects with a score of 1 to demonstrate the ability of the sensors to detect nuanced movements.

The lines with *magenta* '.' and *blue* '◇' markers have values that are close to zero initially. These denote the surface for the hand to grasp. The line with the *red* '+' marker denotes the force on the bottom sensor exerted by the mass of the Cube when it rests on a surface. It goes to zero when the Cube is lifted and this acts as a cue to indicate the *start* and *end* of a movement. This allows automatic segmenting of signals yielding an accurate measure of the duration of the movement. From this signal, another observation from Fig. 3 is that the subject may incorrectly *drop* the Cube rather than placing it on the table and that a subject may graze the bottom of the

Cube against another object during a move.

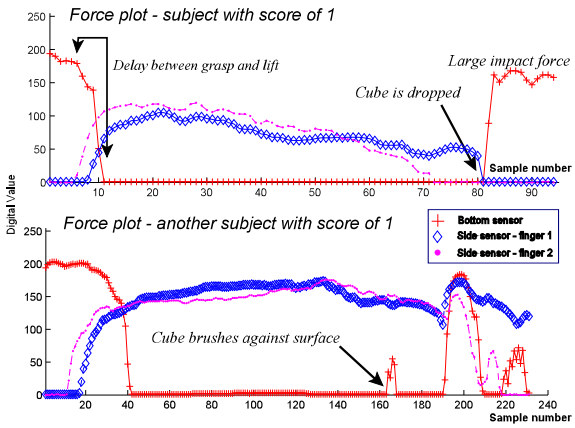

Fig. 3. Force sensor plots for score of 1 - top- Cube is dropped, not placed, bottom- Cube brushes against surface but is placed, not dropped. Marker with red + is bottom sensor plot, other markers are force exerted by fingers.

4. THEORY OF ANALYSIS

The theory for SSA is covered here, with the determination of the important eigenvalues of the system. We also briefly review some of our earlier approaches. In keeping with the relevant literature, we will substitute the term time series for a signal in this section.

4.1 Singular Spectrum Analysis

SSA is a subspace analysis method originally developed for single time variable analysis. In this section we describe the process, based on the work by Vautard et al.[5] where for a time series, at each time instant t the data is represented by a vector $\boldsymbol{x}(t) = \{x(t): t=1...N\}$ with N samples. A window of length $M < N$ is used to embed this series into a trajectory matrix $\mathbf{Y}$, of size $M\times(N- M)$ where for the first and second column vectors:

$$\boldsymbol{x}_1 = [x(1), x(2),..., x(M)]^T$$
$$\boldsymbol{x}_2 = [x(2), x(3),..., x_t(M+1)]^T \quad \text{and for column } M$$

$$\boldsymbol{x}_M = [x(N-M), x(N-M+1),..., x(N)]^T$$

where ^{T} denotes the transpose operator. By concatenating the vectors, the trajectory matrix is:

$$\mathbf{Y} = [\boldsymbol{x}_1 \boldsymbol{x}_2 ... \boldsymbol{x}_{(N-M)}]$$

and the covariance matrix $\mathbf{C}$ for the system is given by:

$$\mathbf{C} = \mathbf{Y}^T\mathbf{Y} / N \quad \text{of size } M\times M$$

Using Singular Value Decomposition (SVD) on $\mathbf{C}$ produces the sorted scalar eigenvalues λ and the eigenvectors $\boldsymbol{e}$, of length M. They are used to form principal components (PC), the k^{th} PC is a vector of length $N-M$, is given by:

$$\boldsymbol{a}^k = \sum_{j=1}^{M} \boldsymbol{x}(j)\boldsymbol{e}_j^k \qquad (1)$$

A useful step is to reconstruct a signal component (RC) corresponding to the k^{th} eigenvalue. This vector of length N has its components constructed differently, so that at a sample instance t we have:

$$RC(t) = \frac{1}{M}\sum_{j=1}^{M}\sum_{k\in K} a_{t-j}^k \, e_j^k \qquad \text{for } M \le t \le N - M + 1$$

$$= \frac{1}{t}\sum_{j=1}^{M}\sum_{k\in K} a_{t-j}^k \, e_j^k \qquad \text{for } 1 \le t \le M - 1$$

$$= \frac{1}{N-t+1}\sum_{j=t-N+M}^{M}\sum_{k\in K} a_{t-j}^k \, e_j^k \qquad \text{for } N - M \le t \le N$$

Where K is the set of PCs used for reconstruction, most often 1. The different equations are needed to cater for the beginning and end conditions of the embedding operation. More details are in the given reference.

4.2 Significant eigenvalues

The SVD returns a set of M sorted eigenvalues. The sorted rank of the eigenvalue shows the amount of contribution of that component to the variability of the data. By plotting the λs in terms of their magnitude we get a scree plot, so named as it resembles the scree which is the rubble at the foot of a mountain as seen in Fig. 4.

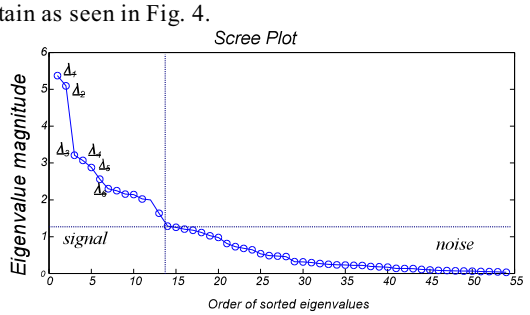

Fig. 4. Scree plot - first 6 eigenvalues indicated

It can be seen that the slope of the scree plot changes drastically in the first few λs, and settles on a gentle slope for the rest. This allows identification of the significant λs and eigenvectors ignoring the rest which can be attributable to noise, which reduces the data dimensionality.

The eigenvalue with the sequence number d at which the significant λs begin have been the object of much research. Early approaches worked on the linear *trends* of changes in λ as reviewed comprehensively by Raîche et al. [6]. In [5] this point has been referred to as the Statistical Dimension, however the computation seems geared towards nonlinear systems.

In this paper we introduce a new way of obtaining the significant eigenvalues based on Relevant Dimension Estimation (RDE) as introduced by Braun et al. [7]. They consider the problem of reducing the dimensionality of a classification problem that uses using kernel subspace methods. A kernel covariance matrix is computed and the λs (obtained by employing SVD) decrease in the same manner as in a scree plot. The lower valued λs are attributed to noise and their distribution is modelled by a zero-mean Gaussian distribution with standard deviation (SD) σ_2 while the higher valued λs explain the actual variation in data and modelled by another Gaussian distribution with SD σ_1 - or more concisely:

$$\lambda_i \sim \begin{cases} N(0,\sigma_1^2) & 1 \le i \le d \\ N(0,\sigma_2^2) & d < i \le M \end{cases} \qquad (2)$$

for a λ in ranked d in the series. Also, $\sigma_1 \gg \sigma_2$ in order to get meaningful results as the "noisy" λs will have a lower

variance as it is evenly distributed along smaller values.

The higher valued λs, as part of the signal proper, will have a higher variance as seen in the scree plot in Fig. 4. The optimal value of d, is one which will minimize the negative log-likelihood:

$$l(d) = \frac{d}{M} \log \sigma_1^2 + \frac{M-d}{M} \log \sigma_2^2$$
$$\text{with } \sigma_1^2 = \frac{1}{d}\sum_{i=1}^{d} \lambda_1^2 \;,\; \sigma_2^2 = \frac{1}{M-d} \sum_{i=1}^{d} \lambda_2^2 \qquad (3)$$

This value of d will be used to separate the eigenvalues.

4.3 Earlier signal analysis using SSA

To provide some background, we consider briefly our earlier works [4] using data from patient simulators and using SSA to analyse accelerometer signals from the Cube to detect abnormal conditions. Among them are:

i) selecting the dominant frequency from a particular RC from the accelerometer *y*-axis (vertical) signal.
ii) using SSA to prefilter signals and computing the average energy of the *xyz*-axis signals.
iii) prefiltering signals with SSA and then computing the area under the signals.
iv) determining the *most* dominant frequency among all three axes, using Multivariate SSA[8].

The way the patient simulators were asked to simulate movement disorders may have effect on the kinds of signal features produced. However, the kinds of analyses performed prepared us for what to expect for actual patient data.

5. RESULTS

We present some initial results in this section and show how the design of the sensors help us to interpret the obtained readings. This is followed by an analysis of the smoothness of movement.

5.1 Significant axis of movement

A typical plot of accelerometer signals between patients with a score of 3 and 2 are compared in Fig. 5. It can be seen that in general the movements of those with a score of 2 have more variation.

We subjected the data to the analyses performed before, but no discernible pattern of correlation was found. For example, where we used the significant frequency in the 2nd SSA reconstructed signal in the *y*-axis, it could not show any discernible pattern here.

But when we examined the entire set of patient data, the *z*-axis data which corresponds to the side-to-side movement of the Cube, showed results more correlated with the score. Specifically, it was noticed that the signal's coefficient of variation, which is the standard deviation divided by the mean value gave interesting results. By removing the mean, this value corresponds to the root mean squared (RMS) value of the AC component of our signals. The step to remove the mean should be done anyway to mitigate the effect of the constant pull of gravity on the accelerometer. This measure then gives an indication of the energy expended in the movement. We separate the RMS value of the signal into the component attributable to the signal, RMSS and that of noise, RMSN.

In our data set, each trial has an identification (ID) code formulated as *SCC_MM_T* where *S* is P for our subjects who were patients, *CC* the subject code, *MM* the movement type, which has a value of TS for our subjects and *T* being the trial number, 1 to 3.

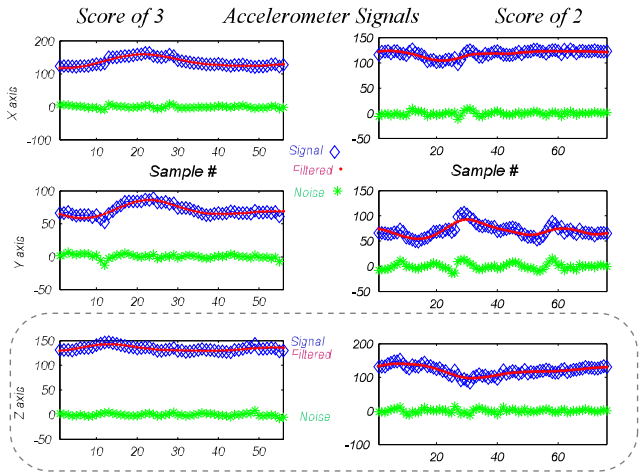

Fig. 5. Plots of signal, filtered and noise data : Left, score of 3 and right, score of 2 using SSA. RDE was used for automatic significant eigenvalue determination. Top to bottom are *xyz*-axis plots respectively.

The results are shown in Table 1. Some error is to be expected because of variation in execution of the moves and scoring. These are summarized in the confusion matrices found in Tables 2 and 3.

Table 1 RMS signal/noise value of accelerometer readings for patients with score 2 and 3. Each have 3 trials and the axis values of 1/2/3 represent *x/y/z* respectively. Shaded rows are *z*-axis.

Subject	Axis	RMS signal/noise		score
P20_TS_1	1	5.35	3.81	2
P20_TS_1	2	10.5	6.08	2
P20_TS_1	3	7.18	4.6	2
P20_TS_2	1	3.58	3.12	2
P20_TS_2	2	11.82	4.2	2
P20_TS_2	3	7.29	4.91	2
P20_TS_3	1	4.16	4.84	2
P20_TS_3	2	14.6	8.71	2
P20_TS_3	3	6.56	5.24	2
P21_TS_1	1	4.23	1.61	3
P21_TS_1	2	8.83	2.41	3
P21_TS_1	3	6.86	2.2	3
P21_TS_2	1	4.48	2.28	3
P21_TS_2	2	9.95	3.42	3
P21_TS_2	3	5.73	2.86	3
P21_TS_3	1	3.2	2.36	3
P21_TS_3	2	10.11	3.03	3
P21_TS_3	3	7.17	3.28	3

From examining the trial data, we found a threshold of 6.9 is sufficient to distinguish between the scores of 3 and 2. For a session, we take the average of the scores of the 3 trials and round up. For example, for *P20_TS_3*, the RMSS is 6.56 which would indicate a score of 3, but the other two trials *P20_TS_1* and *P20_TS_2* the RMSS are 7.18 and 7.29 the average is 7.01 - a score of 2. In Table 2, for scores of 1 and 2, a RMSS value of 12 separates between them.

Table 2 RMS signal/noise value of accelerometer readings for patients with score 1 and 2. Only the 3rd or *z*-axis shown.

Subject	Axis	RMS signal/noise		score
P27_TS_1	3	14.45	5.37	1
P27_TS_1	3	14.73	6.05	1
P27_TS_1	3	12.78	4.85	1
P28_TS_2	3	8.85	3.61	2
P28_TS_2	3	9.11	5.03	2
P28_TS_2	3	11.46	5.4	2

In summary, the steps to automatically score a test are:

i) for each subject and session, score each trial.
ii) normalize *z*-axis signals to zero mean.
iii) perform a SSA to obtain the λs.
iv) perform a Relevant Dimension Estimation of the signal.
v) note the threshold for significant λs.
vi) reconstruct the zero-mean signal-use significant λs.
vii) compute the RMSS.
viii)if RMSS < 6.9, score is 3, < 12, score is 2 else 1
ix) select the score that appears in most of the trials.
x) if there is a tie, average the RMS and use as score.

However, if there is improper execution of a move, a score of 1 is given, as per guidelines.

Next we present the results of our automated scoring. The accuracy is given on a per-trial basis and on a per-session basis, noting that the per-session score is the maximum result from all the per-trial results. It is unfortunate that only two score 1 sessions were recorded and one of them obtained this score because of improper handling of the Cube.

It should be noted that out of 75 trials, 31 trials were scored at 3, 38 scored at 2 and 6 scored at 1. Also, of the 26 valid sessions, 11 were scored at 3 and also 13 at 2 but 2 at a score of 1. The results are shown in the confusion matrices in Tables 3 and 4.

Table 3 Confusion matrix on per-trial scoring - bottom row shows the number of trials receiving the score.

Actual / Predict	3	2	1
3	19	6	0
2	9	28	0
1	3	4	6
# actual scored	31	38	6

From this, we can see that the accuracy for *trial* scoring is (total correct/total sessions) = (19+28+6)/75 = 71%

Table 4 Confusion matrix on per-session scoring - bottom row shows the number of subjects receiving the score.

Actual / Predict	3	2	1
3	7	3	0
2	3	9	0
1	1	1	2
# actual scored	11	13	2

From this, we can see that the accuracy for session scoring is (total correct/total sessions) = (7+9+2)/26 = 69%.

It bears recalling from Sec. 3.1 that we are attempting to objectify what is essentially a rather subjective rating, given by different scorers with no means of normalizing results. Informal, unpublished results with other team members on the grant team, working on other aspects of the ARAT have not been able to achieve this kind of recognition rate.

6. CONCLUSIONS

In summary, we attempted to automatically score the ARAT Test4. Using SSA with RDE we introduced a new time domain feature, namely the Root Mean Squared value of the SSA filtered *z*-axis accelerometer signal to get an accuracy of 69% amidst rather uncontrolled circumstances.

The *z*-axis data corresponds to the side by side movement of the Cube. It may be that this movement is visually more prominent to the assessor. This is a significant finding that validates our approach and paves the way for continuing research in this direction.

Future work will involve the analyses of other accelerometer signals, other types of eigenvalue analyses for a more robust determination and characterisation of movement disorders. There is also the need to secure adequate subjects with an even distribution of scores that are needed to be validated against. A more thorough briefing procedure for clinicians acting as scorers would also be needed.

ACKNOWLEDGEMENT

This work was funded by the Ministry of Education of Singapore under grant number 2010MOE-IF-005. We thank Eric Breitenberger his software available. We also thank the anonymous reviewers for their useful comments which have greatly helped to improve the quality of this paper

REFERENCES

1. K. Andrews and J. Steward, "Stroke Recovery: He can but does he?" *Rheumatology*, vol.16, no.1, pp.43-48, 1978.
2. R. C. Lyle, "A Performance Test for Assessment of Upper Limb Function in Physical Rehabilitation Treatment and Research," *Int. J. Of Rehab. Research*, vol.4, no.4, pp.483-492, 1981.
3. N. Yozbatiran, L. Der-Yeghiaian and S. C. Cramer, "A Standardized Approach to Performing the Action Research Arm Test," *Neurorehabilitation and Neural Repair*, vol.22, no.1, pp.78-90, 2008.
4. T. K. M. Lee, J. G. Lim, S. Sanei, and S. S. W. Gan,"Advances on Singular Spectrum Analysis of Rehabilitative Assessment Data", *J. Med. Imaging Health Inf.,* vol. 5, pp. 350-358, 2015.
5. R. Vautard, P. Yiou and M. Ghil, "Singular Spectrum Analysis: A toolkit for short, noisy chaotic signals", *Physica D,* vol. 58, pp. 95-126.
6. G. Raîche, T. A Walls, D. Magis, M.Riopel and J. Blais, "Non-graphical solutions for Cattell's scree test," *Methodology: European J. of Research Methods for the Behavioral and Social Sci.*, vol.9 , no.1, pp.23-29, 2013.
7. M. L. Braun, J. Buhmann and K. -R. Müller, "On Relevant Dimensions in Kernel Feature Spaces," *J Mach Learning Research*, vol.9 , pp.1875-1908, 2008.
8. Tracey K. M. Lee, Sharon S. W. Gan , J. G. Lim, Saeid Sanei, "A Multivariate Singular Spectrum Analysis Approach to Clinically-Motivated Movement Biometrics," *Proc. 22nd European Signal Processing Conf*, 2014.