Development and Evaluation of a Clinically Robust Measurement System for Visuo-Spatial Neglect: Assessment and Conclusions

Richard Guest
University of Kent

Introduction and research context

Visual-spatial neglect (VSN) has long been recognised as a major factor in delaying post-stroke rehabilitation. Accurate detection of the condition is therefore critical in selecting an appropriate rehabilitation methodology. A common method for the diagnosis and severity of VSN is to assess the outcome of a series of hand drawn tasks. Previous work [1] by the author has explored the diagnostic ability of analysing both the outcome and the constructional elements of task performance using a computer-based drawing capture system, concentrating mainly on cancellation (target location and marking) based tasks. Using this system, patient task responses were made using a cordless pen on a sheet of paper overlaid on a conventional graphics tablet. The tablet captured the position of the pen (along with other information such as pressure and tilt) which was stored in a time-stamped file for later processing. From this file it is possible to extract features detailing the outcome of the drawing (static features such as image size) alongside constructional properties (dynamic features such as completion time and pen velocity) [2].

The work conducted for this study had four aims: i) to build upon existing work by exploring the diagnostic significance of geometric drawing tasks through the development and implementation of novel feature extraction algorithms, ii) to develop a clinical capture tool and demonstrator, iii) to further explore the relationship and diagnostic capability of these new features alongside previously extracted test performance metrics within a number of clinical and demographic groupings and iv) conduct a clinical trial on a small test population to establish the performance of the developed robust capture tool and extracted features. These aims, which have been successfully met through the work undertaken in the study, will be reviewed in the following sections.

Aim 1 - Drawing Task Performance Feature Classification and Diagnosis

Alongside other pen-based tasks, our previous work had captured responses from six drawing tests – four figure copying (FC) shapes (a square, a five pointed star, a cube and a cross – the models shown in Figure 1) and two drawing from memory (DFM) tasks where the subject was asked to draw a shape (a square and a cube) on a blank piece of paper without prompting. In the previous trial [3], only a basic analysis had been performed on these responses (analysing features such as the number of sides or corners within a drawing, the position within the shape that the person started the drawing [16]). The work conducted within this study aimed to further the analysis of the diagnostic properties of these drawing tasks by answering two main questions in relation to drawing performance. Firstly, by defining and extracting a series of generic features, identify which shape response produces the best diagnostic indicator of VSN and secondly to question the necessity of administering all six drawing tasks, determining whether a comparable performance (or even improvement) can be obtained through using a smaller test battery, thereby reducing the fatigue in patient performance. Given the inherent difficulty in distinguishing VSN subjects from stroke patients without VSN [4] (VSN has many levels of severity which results in mild cases often exhibiting similar diagnostic results as non-VSN patients – this issue is specifically addressed by Aim 3) the investigation does not aim for a 0% error rate but merely to provide a consistent assessment tool with the highest achievable classification accuracy to be available in a clinician's diagnosis battery.

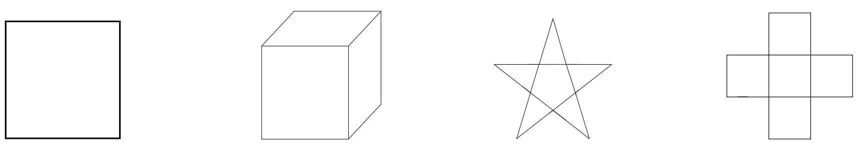

Figure 1: Shape copying models.

28 VSN subjects (identified using standard clinical assessments) were included in the experiementation. 55 stroke subjects without VSN were used as a stroke control group. A series of 12 generic performance features (4 static, 8 dynamic) were defined and extracted from each drawing response. These are outlined in Table 1. The experiment aimed to assess the ability of a series of trainable automatic classifiers to diagnose a subject based upon the feature results. The leave-one-out error-estimation technique [5] was

employed to estimate classifier error rates due to its suitability where only a relatively small sample size is available and because the method is an almost unbiased estimator of the true error rate of a particular classifier.

Feature Name	Feature Type	Comments
Starting Point	Dynamic	Drawing is divided into quadrants. The starting point signifies in which quadrant the drawing was started.
Number of Strokes	Dynamic	The number of individual line strokes drawn during drawing execution.
Task Execution Time	Dynamic	Time taken to complete the each drawing task.
Maximum and Mean Pressure	Dynamic	Pen pressure was recorded by the tablet at 255 discrete levels
Average Pen Velocity	Dynamic	Average pen speed
Average Pen Acceleration	Dynamic	Average change in pen speed
Pen Contact Ratio	Dynamic	Ratio of the time when pen is off the tablet to the time when pen is on the tablet,
Overall Distance Travelled	Static	Distance travelled by the pen in executing the drawing.
Image Height, Width and Area	Static	Measure of Image Height, Width and Area

Table 1: Drawing features

As an initial investigation of the classification ability of the extracted features, a single vector of 72 features across all drawing tasks (12 per task) was used individually to train a series of standard classifiers - Fisher linear discriminant, nearest mean (nmc), k-nearest neighbour (knnc), linear classifier using Karhunen-Loeve expansion on the joint data (kljlc) and a decision tree (treec). Using the leave-one-out estimation system the Fisher classifier provided the best classification rate (responses correctly classified as being drawn by a VSN or stroke control subject) of 60.0%, providing a benchmark against which improvement in other experiments could be compared.

A second experiment concentrated on analysis of individual drawing tasks, assessing the discriminatory strength of each shape. At this stage it was possible to analyse the most compact representation of the feature space for each task, eliminating the redundant features that fail to significantly contribute to the classification of the responses without degrading the discrimination power. Reducing the number of irrelevant/redundant features reduces the running time of a learning algorithm and, in many cases, improves learning. A forward feature selection (FSS) algorithm was applied to each individual drawing task feature set in a stepwise manner, from a target set of 1 feature up to all 12 features. At each step, the selected features were used to classify the two groups of subjects. Again, the one nearest neighbour leave-one-out classification performance measure was used as the evaluation methodology. The Star FC task has the highest classification rate obtained by any of the classifiers (74.1% using 5 features). The Cube DFM task has the second highest recognition rate of 72% (using 6 features). It can be noted that all these "best" recognition rates are higher than those obtained from using the entire test battery. Dynamic features, representing the drawing strategy, dominated these sub-sets with the number of strokes and the pen contact ratio being retained for all but one of the drawing tasks, indicating the important and novel contribution they make to the diagnostic task. It also indicates that some "traditional" static features are redundant with respect to diagnostic accuracy. Ranking the discriminatory power of individual tasks the Star FC, Cube DFM and Cross FC show the highest classification ability.

A third experiment addressed the problem of assessment efficiency and patient fatigue by investigating the optimum number of drawing tasks to be included in the test battery to give the best diagnosis performance. Based on a hierarchical combination of the selected features from each task, this experiment serves to determine if it is necessary to include all drawing tasks to obtain comparable (or improved) classification results. The selected features detailed in Table 1 for each shape were combined in a hierarchical manner, starting with the shape which exhibited the most discriminatory power. The features from the shape with the second most discriminatory power were then added, followed by the third most powerful and so on, until the selected features from all the shapes were added together. At each stage, classification was carried out using the five classifiers. It was found that the Fisher classifier obtained the best recognition rate of 78.3% (the best result across all 3 experiments) when using selected features from a combination of the best 3 tasks, the Star FC, the Cube DFM and the Cross FC; a total of 17 features.

The experiments have shown that it possible to reduce classification error rate by examining a subset of features and tasks rather than the entire test battery. We have introduced algorithmic objectivity into the assessment of the shapes and have demonstrated that the novel dynamic time-based and constructional features are important diagnostic indicators. The diagnostic information obtained from the investigation can be used alongside other forms of assessment in forming a final diagnosis. A more detailed account of this work can be found in [6].

Sequence Extraction

Our previous work assessing the constructional strategy of cancellation tasks [7] indicated the importance of this diagnostic indicator. One of the major investigations of this work package was to automatically analyse the constructional sequence of shapes by segmenting and ordering the drawing into components (sides of shapes) and attempting to classify the resulting sequence. The aim of the investigation was to demonstrate that the execution strategy - the sequence in which drawing strokes are made - can be effectively used as a diagnostic indictor of VSN in stroke patients. For classification purposes we implemented a recogniser using Hidden Markov Models (HMM) [8]. HMMs are statistical models of sequential data that have been used successfully in image processing and computer vision, speech recognition and modelling of biological sequences among other applications [9, 10]. The HMM is a natural way of recognising a pattern which can be represented as a sequence of discrete observation symbols. The sample sequences extracted from the tasks represent successive pen strokes, which should have a sequential relationship likely to be captured by left-to-right HMM model employed in this study.

Two experiments were conducted using data from the square and cross models. In the first experiment, sequences were automatically extracted using a devised algorithm which were then utilised in the training and testing of the HMMs. In the second experiment, drawing sequences were obtained manually using a playback routine which reproduced the drawings dynamically utilising the timing data captured during the drawing task. These manually obtained sequences were presented to the HMM. Comparing the results from the two experiments allows an observation as to the robustness of the generalisation ability of the HMM given that some of the sequences in the first experiment may not be correctly detected. It also enables the accuracy of the automated sequence extraction method to be measured.

Extracting the sequence from the drawing data was achieved by assessing the list of stored pen coordinates. Figure 2 shows a flowchart of the method used to automatically extract side sequences from shape drawing response files.

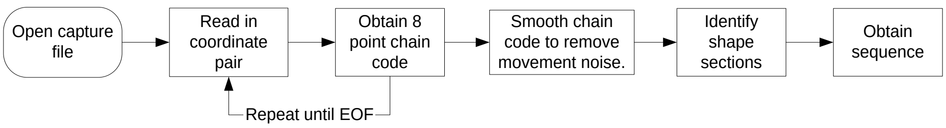

Figure 2: Sequence extraction

Table 2 shows the performance of the automatic sequence analysis routine in terms of correctly recongised sequences verified against manual analysis. It can be seen that the square task produced the most accurate analysis results. Incorrect sequence detection was due to a number of factors. In some cases all the straight lines would be successfully extracted, but the sequence would be distorted by the subject returning within the drawing sequence to repeat a line they had drawn earlier. Some cases would result in two parallel lines being drawn whilst the subject attempts to re-emphasize a particular line. In the case of the cross, besides these distortions, the subject might not have followed the anticipated model line sequences.

In the diagnosis of patients, an overall recognition rate of 69% was obtained using the sequences extracted from the cross drawing task with the square tasks at 62% and 60% for the DFM and FC task respectively. As would be expected, the recognition rates are higher for manually extracted ‘correct’ sequences, however the small differences (72%, 64% and 60% respectively) between the two experiment’s results shows both the minimal trade-off between accuracy and speed in using the automated stroke recognition system and the generalisation ability of the HMM classifier.

Work is continuing to apply the automatic extraction techniques to the star and cube drawing models. More information on this study is to appear in [12].

Multiple classification

As a final investigation into drawing feature analysis, a multi-classification architecture [11] was established to utilise the diverse nature of the information provided by the different static/dynamic features and sequential information in an attempt to further improve diagnostic ability. In order to increase the number of "decision makers" in the multiple classifier strategy, feature sets obtained from an individual shape were divided into two sets based on a performance-ranking. Features were ranked in order of performance using a Euclidean distance based nearest neighbour strategy. In order to generate balanced groups with no dominating feature, the two groups were formed by extracting alternate features in the ranking order, i.e. Group 1 comprised of features with ranking 1,3,5,7,11 and Group 2 features ranking 2,4,6,8,10,12. The HMM sequential analysis on the shape under consideration was included as a third classifier. Figure 3 shows this configuration where features extracted from a single drawing are divided into the three groupings.

Group	Drawing Task	Sequence Successful Extraction Rate (%)
VSN	Square (DFM)	82.4
	Square (FC)	82.4
	Cross (FC)	56.6
Stroke Control	Square (DFM)	92.7
	Square (FC)	83.7
	Cross (FC)	71.2

Table 2: Automatic sequence recognition rates.

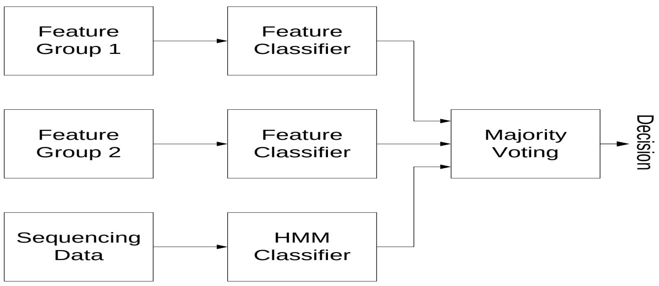

Figure 3: Multiple classification scheme

Initially each group was used separately to reach a VSN/Non-VSN classification decision. The decisions from the three classifiers were then used in a majority combination scheme to reach an overall final decision.

The results show that that we get improvements for the two FC tasks, with a recognition rate of 69.5% (using a Fisher classifier) for the square and 78.1% (using a kljl classifier) for the cross, thereby demonstrating the effectiveness of the multi-classifier system in producing the most accurate diagnostic results across experiments.

The development of automated analysis procedures has shown that important diagnostic information can be obtained from shape drawing tasks. Most importantly we have shown that novel dynamic features can also add to the diagnostic outcome of the task.

Aim 2 - Development of a clinical capture tool/demonstrator software

To facilitate the clinical trial, a windows based capture tool was developed (Figure 4). The software runs under Windows XP/2K and interfaces with an attached graphics tablet using the WinTab protocol. Figure 4 shows the capture of a cube drawing task. As the test subjects draws on a piece of paper placed on the tablet, the drawing is recreated on the screen, the thickness of the drawn line reflecting the applied pressure. Special consideration was given to user interface aspects of this design as it would be solely operated by hospital staff, abstracting the use away from the patient and therefore removing any anxieties over the use of technology. The capture routine features simple to use controls (accept or reject the current drawing) and a progress bar for test battery capture. Data is automatically stored in an individual directory for each test subject ID. Figure 4 shows also shows the developed shape feature extraction demonstrator. Again designed with clinical use in mind, the program enables the display and playback of captured files in real-time, extraction of features and classification.

Aim 3 - Study of previously collected data

Our previous study [1] had collected drawing, bisection (location of midpoint within a single line) and cancellation task data from 108 test subjects (including age matched controls). Our existing investigations had only considered the performance of features against an independent clinical measure, grouping test subjects as either VSN or non-VSN according to the output of the Rivermead BIT [13]. VSN has long been recognised as being a condition with varying levels of debilitation therefore a single performance feature extracted from our system should ideally be able to recognise a range of VSN severity levels [4]. This investigation also aimed to establish the relationship between feature outcome and other factors such as stroke locations, age, gender, stroke type. In assessing the results, any inter-group significance indicates an ability of the feature to distinguish between finer levels of VSN. The outcome of this study will ultimately be of interest to clinicians and neurophysiologists as it presents new information about constructional aspects of test performance

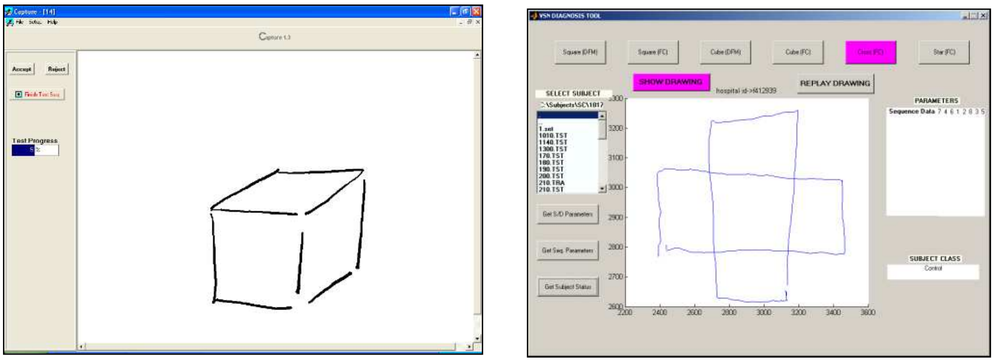

Figure 4: Capture software and drawing replay and feature analysis program

The results from the trial showed that there was no effect due to age, stroke location, gender, month of stroke and number of days post stroke the test was performed. Significant features emerged however when regrouping subjects by grades of VSN performance: Dividing the VSN population into two groups (severe and mild) according to their BIT score produced a range of significant differences within the cancellations tasks, in particular the number of cancellations to the left of the page and, using the dynamic data, intercancellation and movement times within this side of the overlay. Interesting, there are two main features extracted from drawing tasks that are sensitive to levels of VSN, these being the number of individual pen strokes made and the overall pen distance travelled. The square FC and cube DFM tasks seem the most sensitive to this. Dividing the VSN group into 4 subgroups (severe, semi-severe, semi-mild and mild VSN) based on the BIT score quartiles there are very few features that are able to detect between these groups due to the sample size. The features that are able to detect significance between groupings are the cancellation task – in particular the total number of cancellations made and the drawing time within the bottom right quadrant.

Utilising the stroke type and location record for each subject enabled us to investigate the effect of these data on test performance. As an initial investigation we divided the population according to their stroke type (Infarct, Haemorrhage, Atrophy or Ischemia). Our findings indicate that the bisection and cancellation seems the best to distinguish between Infarct and Haemorrhage – both using static and dynamic features. The drawing tasks show other significant group differences by stroke type and location again using both static and dynamic data.

We also explored the interrater reliability of clinician responses showing good reliability in some aspects, but significant subjectivity in other areas of assessment [15]

Aim 4 - Clinical Trial

A clinical trial was conducted in order to evaluate both the usability functionality of the developed clinical software and the accuracy in diagnosis using the devised methods described in Aim 1. The trial was carried out over a period of seven months collecting drawing data from 22 subjects. Of these subjects, 6 were age-matched controls (i.e. no medical history). The other subjects were stroke patients in the Stroke Unit at Kent and Canterbury Hospital. Full Local Ethics Committee approval was granted for this trial.

Stroke patients were first administered the Rivermead BIT to provide an independent assessment on their stroke condition. Out of the 16 stroke patients, 2 of them exhibited VSN symptoms – one severe and one mild VSN patient. The collected shape drawing data for each patient was classified using methods described above. The results showed that the two VSN patients from the study period were correctly classified. Out of the 14 stroke control cases, 3 were incorrectly classified as VSN patients. Analysing these misclassified cases it was found that these were all test subjects with BIT scores close to the VSN/non-VSN cut-off indicating the difficulty in automated diagnosis.

More information on this study is to appear in [14]. A larger scale clinical trial is planned to further test the diagnostic ability of the system. The devised software performed well under the six months of clinical use with only minor modifications to the user interface.

Research Impact and Benefits to Society

The research conducted in this project has been of international standing in the areas of hand-drawing and classification research – our publications to date resulting from this work verify this. We have applied novel multi-classifier and sequence analysis methodologies to the diagnosis of VSN and have devised a solid platform for future experimentation and research. Dissemination within the medical community will provide information on the diagnosis and symptoms of VSN.

References

[1] Computer recording of standard test of visual neglect in stroke patients, Jonathan Potter, Tony Deighton, Mahood Patel, Michael Fairhurst, Richard Guest, Nick Donnelly, Clinical Rehabilitation, 14, pp. 441-446, 2000.
[2] Analysing constructional aspects of figure completion for the diagnosis of visuospatial neglect, Richard Guest, Michael Fairhurst, Jonathan Potter. Nick Donnelly, In Proc. ICPR2000, Vol. 4, pp.316-319, 2000.
[3] Novel rule-based static and dynamic feature extraction from figure copying tasks for the detection of visuo-spatial neglect, Richard Guest, Michael Fairhurst, Jonathan Potter, Nick Donnelly, In Proc. IWFHR 2000, pp. 547-552, 2000.
[4] Visuo-spatial neglect - underlying factors and test sensitivity, Halligan et al., Lancet, Vol. II, 908-911, 1989.
[5] Evaluation of attributes obtained in statistical decision rules, Lunts et al., Engineering Cybernetics, 3, pp. 98-109, 1967.
[6] Optimisation Procedures for Diagnostic Processing Of Hand-Drawn Geometric Figures, Richard Guest, Samuel Chindaro, Michael Fairhurst and Jonathan Potter, Electronic Letters, Vol. 39, Iss. 2, 2003, pp. 205-206.
[7] Diagnosis of visuo-spatial neglect using dynamic sequence features from a cancellation task, Richard Guest, Michael Fairhurst, Jonathan, Pattern Analysis and Applications, 5, pp. 261-270, 2002.
[8] An Introduction to Hidden Markov Models, Rabiner et al., IEEE ASSP Magazine, pp. 4-16, 1986.
[9] A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Rabiner, Proc. of the IEEE, Vol. 7, No. 2, pp. 257-286, 1989.
[10] Applications of hidden Markov chains in image analysis, Aas et al., Pattern Recognition, 32(4), p. 703, 1999.
[11] An evaluation of multi-expert configurations for the recognition of handwritten numerals, Rahman et al., Pattern Recognition, 31, pp. 1255-1273, 1998.
[12] Automatic Classification of Hand Drawn Geometric Shapes using Constructional Sequence Analysis, Richard Guest, Samuel Chindaro, Michael Fairhurst and Jonathan Potter, Proc ICDAR2003, pp. 990-994.
[13] The Behavioural Inattention Test, Wilson et al., Thames Valley Test Company, Fareham, Hampshire, 1987.
[14] Assessing visuo-spatial neglect through feature selection from shape drawing performance and sequence analysis. Sam Chindaro, Richard Guest, Michael Fairhurst, M and Jonathan Potter. International Journal of Pattern Recognition and Artificial Intelligence, 18(07), 1253-1266. 2004.
[15] Assessing visual inattention: study of inter-rater reliability, S Hannaford, G Gower, Jonathan Potter, Richard Guest, Michael Fairhurst, International Journal of Therapy and Rehabilitation, 10:2, 72-75, 2003
[16] A clustering approach to corner point analysis in hand drawn images, Richard Guest, Michael Fairhurst, In Proc 16th International Conference on Pattern Recognition, 2002.,vol 3. pp 940-943.