Developing a generic approach to online automated analysis of writing and drawing tests in clinical patient profiling

M. C. FAIRHURST AND T. LINNELL
University of Kent, Canterbury, England

S. GLÉNAT
Université de Rouen, Rouen, France

R. M. GUEST
University of Kent, Canterbury, England

AND

L. HEUTTE AND T. PAQUET
Université de Rouen, Rouen, France

Writing and drawing tests are widely used in the clinical environment for the diagnosis of a variety of neuropsychological conditions. Conventional assessment of these tests involves the inspection by trained assessors of the completed patient response. This article describes the development of a computer-based framework for data capture, automated feature analysis, and result reporting for a range of drawing- and writing-based test batteries. In developing this framework, we have exploited the commonality between tasks while allowing for both flexibility in configuration across condition-specific testing requirements and extensibility for future test development. Using the two example clinical conditions of visuospatial neglect and dyspraxia, we illustrate the advantages of utilizing a computer-based analysis system, describe a structured approach to system implementation, and demonstrate the generality of this implementation for different conditions of interest, which extends to feature selection and design.

Investigating behavioral dysfunction can provide important information about impairments attributable to various neurological conditions. An understanding of the effects of conditions ranging from cerebrovascular accident (CVA; i.e., stroke), through forms of dementia, to developmental coordination disorders can benefit from detailed behavioral profiling of the individuals affected (Fairhurst, Hoque, & Razian, 2004; Guest, Donnelly, Fairhurst, & Potter, 2004). Thus, effective diagnosis and detailed assessment of a condition may ultimately depend on the acquisition of reliable information about behavioral characteristics and detailed patient profiling.

Although some approaches to establishing diagnosis and assessment rely on detailed, wide-ranging, and highly specialized investigative techniques, it is also true that many of these conditions are susceptible to investigation through relatively simple and more easily administered testing procedures. An example of such an approach is the use of "pencil-and-paper" tests that aim to capture and analyze behavioral characteristics in order to describe the capabilities of patients in handwriting and drawing. In evaluating

the perceptual capabilities of stroke patients, for example, cancellation tests or 3-D shape copying can reveal information about both 2-D spatial perceptual deficits and perception of depth and perspective in the visual field (Kirk & Kertesz, 1989). The assessment of a child's perceptual and motor capabilities is another example, since, through the analysis of hand-drawn copied figures or constructed visual patterns, therapists and clinicians can extract information about the child's visual coordination capabilities (Bender, 1946; Montheil, 1993; Rey, 1959).

More importantly, this type of evaluative testing, based on an analysis of writing and drawing tasks, is widespread across a range of conditions and can be found as part of standard assessment protocols in the evaluation of CVA, the investigation of Alzheimer's and Parkinson's patients, and the study of developmental coordination disorders, such as dyspraxia in young children (Lezak, Howieson, & Loring, 2004; Portwood, 2000). Such a degree of commonality across testing regimes has both advantages and disadvantages. In terms of the disadvantages, the use of similar hand-drawn testing procedures across a variety of

R. M. Guest, r.m.guest@kent.ac.uk

clinical scenarios can lead to inappropriate tests and measurements being applied to conditions for which they were not designed or for which the test commonality does not account for condition-specific measurements that can be extracted from the tasks and that could lead to enhanced sensitivity to test subject performance. On the other hand, the widespread adoption of a small number of approaches that can influence a wide range of different scenarios suggests that efficiencies of scale and commonality might be sought successfully. These are the ideas behind our attempt to design a system that can both implement commonality between tasks and allow for task-specific flexibility and independence in measurement, in order to enhance the resolution of information available to clinicians.

This article addresses some of these important issues and, in particular, investigates an approach to the exploitation of the possible advantages of a common framework for testing that firmly recognizes the potential for focusing on generic testing regimes while, at the same time, acknowledging the need for condition-specific options where appropriate. We describe a novel approach to testing that improves the efficiency and effectiveness of the execution of writing and drawing tests and enhances the richness and potency of the information that can be extracted from conventional assessment procedures. Importantly, this approach also lends itself to implementation as a powerful generic system that nonetheless retains enough flexibility to support widespread clinical adoption.

A RATIONALE FOR ONLINE DATA CAPTURE IN WRITING AND DRAWING TESTS

Traditionally, typical patient assessments in writing and drawing tests involve the capture of data using direct pencil-and-paper implementation, in which the patient generates a specified output through freehand expression. The resulting end-product of the test is then visually inspected by a human observer (generally a clinician or trained clinical support worker), and appropriate diagnostic or evaluative information is extracted. This information can take the form of a qualitative description and analysis of the patient response or, more often, of a quantitative analysis based on a specified set of analytical rules. This sort of assessment can often result in the assignment of a specific "score" to the patient output, on the basis of which some conclusion about patient status can be drawn or a direct contribution to a diagnostic evaluation can be made.

A typical example is the use of Albert's cancellation test to assess the spatial perceptual ability of a stroke patient (Albert, 1973). Counting the target line segments canceled (that is, located and marked by a pen or pencil) by a particular patient can be indicative of the severity of the effect of CVA on the patient's ability to detect the objects in the visual field, whereas an analysis of the spatial distribution of canceled and noncanceled elements can provide insight into the nature of the damage caused by the CVA. As part of a wider test battery, such information can also contribute to the diagnosis of visuospatial neglect in the patient (Wilson, Cockburn, & Halligan, 1987). Fol-

lowing the same approach, the Bender test (and its many versions and modifications) is widely used by clinicians to assess the constructional ability of a patient (Bender, 1946; nferNelson, 2003). Comprising five figures that are hand-copied, the production of each figure is manually scored according to defined rules (Zazzo, 1967) that relate to, for example, the size of the drawing or its orientation. The overall score gives reference indicators for the diagnosis; within a wider test battery, this score can contribute to the diagnosis of dyspraxia.

Though effective in principle, such (essentially *offline*) analysis of individual patient performance is both restricted in scope and prone to potentially serious drawbacks. Some illustrations of these problems follow below.

1. Although direct visual inspection can determine both qualitative and quantitative indicators of performance, these are almost always related to the analysis of the end-product of the patient activity (i.e., the finished test output, seen as a complete entity) rather than to anything that can provide an indicator of the underlying execution process involved. The only way in which information of the latter type can be obtained is by continuous observation of the test (with associated detailed observation and note-taking) while the test procedure is being executed. In many situations, it may be important to distinguish between different types of behavioral characteristics in describing an individual patient performance. For example, it may be important to distinguish between aspects of performance that relate specifically to, say, motor functioning, and other aspects that describe cognitive function or strategies adopted in execution. Such distinctions, which can make a substantial impact on the nature of the information available from a particular test, may not be discernible from a typical test procedure or may incur significant overheads, even if they can, in principle, be identified.

2. Many of the tests commonly adopted can be complex in design and can require very detailed "rules" by which their analysis should be carried out. This can lead to inter-rater variability in scoring, which can undermine effective objective assessment. More seriously, for some tests the evaluation guidelines specified can require a considerable degree of subjectivity in assessment that can exacerbate the difficulty of ensuring objectivity and repeatability in testing.

3. Although in many implementations hand-drawing tasks provide a quick and efficient method for evaluation, in others the testing process can be time consuming both for the patient and for the staff administering the test. A test battery comprising many individual overlays can cause fatigue in a patient, thereby affecting test performance. This can raise issues of both cost effectiveness and clinical suitability, especially if the information generated is subject to uncertainty or potential inaccuracy.

Moving to a system of *online* data capture (whereby sequential/constructional information about the drawing process is captured and stored) and automated analysis can therefore bring significant rewards in relation to all of the issues above. As will be shown, online data capture makes available a much richer source of potentially valuable diagnostic data (in the example given above, for

instance, online capture can readily track spatial trajectories of pen movements in the line-canceling process); the algorithmic application of analytical procedures ensures objectivity and repeatability of interpretation; and automatic processing of the data significantly reduces the resource implications for staff. Similarly, automated testing offers the potential for more streamlined and less intrusive procedures from a patient perspective.

For these reasons, an effective system for online capture and analysis of typical writing and drawing tasks has been a priority for development.

AN ONLINE CAPTURE AND ASSESSMENT INFRASTRUCTURE

The key to the performance assessment methods to be considered in this article is the capture of *online* information about the execution of the patient's writing or drawing activity, so that subsequent (automated) data analysis can extract appropriate diagnostic or profiling characteristics. A typical infrastructure for capturing these data is a standard computer-linked graphics tablet. In this practical scenario, the conventional paper-based test sheets can be placed on top of the tablet surface and positioned in front of the patient on a table. The use of a cordless inking pen of conventional design allows the patient to execute the requirements of the testing procedure directly onto paper and to gain the familiar visual feedback from the drawing process. There is evidence that, because the patient is using a test apparatus identical to that used when taking the conventional "gold-standard" test (i.e., a normal-sized inking pen and a standard piece of paper), test equivalence is maintained (Potter et al., 2000). At the same time, positional information about the pen is stored sequentially for later analysis (it is also worth observing here that many commercial graphics-capture systems also allow the collection of other pen-dependent data—e.g., data relating to the pressure applied in the writing process and, in many cases, to the angle and tilt of the pen as it moves across the writing surface). Thus, the experimental setup is made to parallel almost exactly the data collection conditions prevailing when entirely manual testing is undertaken. Hence, the interface for online data collection remains standard but allows the capture and analysis of relevant data automatically.

Figure 1 shows a schematic of the capture process that also indicates a range of possible further processing stages, involving the extraction and possible selection of performance-related features that might contribute to diagnostic or therapeutic protocols postcapture.

Clearly, a variety of fundamentally different types of performance data can be acquired within this framework. For example, performance features can be readily extracted from the captured drawing that relate directly to the features used in the nonautomated form of testing and assessment. Such features are essentially assessable indicators of the quality and form of the finished drawing (generally denoted *static* features), examples of which might include any of the following: particular quantitative measurements of geometric properties, completeness of the drawing, relative positioning of drawing components, indicators of writing style or functional deficit, and similar indicators of the ultimate quality of the finished drawing as executed by the patient.

More importantly, this novel online data capture approach also offers the potential to extract a much richer set of features that could not readily be obtained through conventional testing, features relating specifically to the characteristics of the drawing execution pattern. Such *dynamic* features would typically include measurements related to timing information, pen velocity or acceleration profiles, number of strokes, and pen-up/pen-down times, as well as a variety of similar measurements that can provide insights into the fundamental physical mechanisms of drawing execution. These dynamic features are an especially important aspect of our approach. In addition, we might identify a third fundamental feature type (though strictly speaking it is a subset of the dynamic feature set already described). These could be characterized as *strategy* features, encompassing aspects of performance such as specific stroke sequencing, start/end points of drawing-element construction, pause durations, and so on. Obviously, these additional features are only worth extracting if they provide an aid to diagnosis. Our previous work (Donnelly et al., 1999; Fairhurst et al., 2004; Glénat, Heutte, Paquet, & Mellier, 2005) has shown the diagnostic capability of such constructional data in terms of both diagnostic classification and as additional individual performance metrics that researchers have hitherto been unable to extract.

Figure 1. A schematic of the online data capture process.

Figure 2. System structure.

Thus, we define a complete online assessment procedure that can not only reproduce conventional (offline) testing but can also extend, in a more efficient and reliable manner, to the capture and analysis of additional behavioral profiling data not accessible via conventional means.

FUNDAMENTAL SYSTEM REQUIREMENTS

The scenario outlined above clearly suggests that many advantages could result from engineering a *generic* clinical evaluation tool to encompass assessment of a range of clinical conditions. The principal requirements of such a system may be summarized as follows:

- A capability for generalized test-independent data capture;
- A high degree of modularity in implementation, for ease of adding or modifying specific features to be extracted;
- The incorporation of a core set of “universal” features that underpin all specific testing scenarios;
- The availability of test-specific supplementary feature sets;
- Adoption of flexible options for analysis of the captured data on the basis of the features selected;
- Ease of feedback to clinical staff of clinically relevant information embedded in the analysis.

Figure 2 shows a schematic of a system structure and design ideals appropriate for such a generic tool. The flow of data is shown between system subcomponents.

Hence, three key issues must be satisfactorily addressed in order to define an effective framework for the specification of the generic tool envisaged. These are (1) a robust data acquisition strategy, (2) a strategy for feature specification and efficient categorization and structuring of features, and (3) a strategy for analysis and feedback of clinical assessment/diagnostic indicators. Each of these issues will be discussed in the following sections. In order to illustrate

the underlying arguments, we will consider two different clinical conditions of interest, both of which commonly use writing/drawing tests as part of an assessment test battery.

The first scenario of interest is the identification of visuospatial neglect in CVA. *Visuospatial neglect* relates to a dysfunction caused by brain damage, the main effect of which is to cause subjects to fail to respond to stimuli in the visual field on the side opposite the location of the lesion (Halligan, Marshall, & Wade, 1989). Traditional testing (Stone et al., 1991) has exploited this effect by assessing the identification of objects within the visual field. Diagnosis of this and other neuropsychological conditions is critical for the selection of an appropriate rehabilitation process to compensate for the effects of neglect. Inadequate detection of neglect at an early stage of therapy will result in a performance deficit from the patient during rehabilitation and a failure to respond to treatment, prolonging the time scale required for recovery and, hence, increasing both the distress of the patient and the costs associated with care and treatment (Edmans & Lincoln, 1990).

Neglect is conventionally assessed by observations and formal testing procedures undertaken by a range of healthcare staff, but a key component of testing is a series of standard pencil-and-paper-based geometric shape drawing tests that can be used to *quantify* performance. The responses to these tasks are traditionally evaluated by therapists or trained assessors using a series of carefully specified assessment criteria (typically relating to the presence or absence of components within the drawing). Figure 3 shows examples of geometric shapes drawn by neglect and asymptomatic patients. The omission of key components of the square and the cross is evident in the neglect responses.

In the work reported here, we will focus on the three types of tests commonly adopted to evaluate this condition:

- Line bisection—location and marking of a mid-point on a predrawn line
- Cancellation—location and marking of a series of targets printed on a page

Figure 3. Figure-drawing models (A), as well as responses from neglect (B and D) and asymptomatic (C and E) patients.

- Drawing—drawing (either freehand or copying) of a series of geometric shapes

Discussion of the analysis of patient output in these tests, as well as examples of features that characterize patient performance, can be found in Azouvi et al. (2002); Chindaro, Guest, Fairhurst, and Potter (2004); and Lezak et al. (2004).

The second scenario concerns the diagnosis and assessment of dyspraxia in children. *Dyspraxia* is an impairment of the organization of movement (Landgren, Pettersson, Kjellman, & Gillberg, 1996). It can be defined as an immaturity in the way the brain processes information that results in messages not being properly or fully transmitted. Dyspraxia affects the planning of what to do and how to do it and may affect some children of primary school age.

Diagnosis is usually made using gesture tests or spatial exercises, drawing tests, and, for children who can write, handwriting tests. Each of these tasks assesses different levels in the organization of movements and is thus complementary to the others.

Among the pen-and-paper tests commonly used to diagnose dyspraxia, the Bender test (Bender, 1946; Zazzo, 1967), the Rey complex figure (Montheil, 1993; Rey, 1959), and the BHK test (a.k.a. the Concise Evaluation Scale for Children's Handwriting; Charles, Soppelsa, & Albaret, 2003) are widely used by clinicians; the first two are drawing tests, and the third is a test based on handwriting analysis.

With the help of clinicians, we have designed a set of tests for a population corresponding to children in primary school (6–11 years old). Since our goal is to allow a dyspraxia diagnosis that is as accurate as possible, we have selected a set of complementary and separate tasks that are of two different types: drawing and handwriting. In the drawing test, the subject must copy the five classical figures shown in Figure 4A. The handwriting test consists of

several different types of simple writing tasks that enable an analysis of the handwriting process, which is, by definition, very complex (Glénat et al., 2005; Rémi, Frelicot, & Courtellemont, 2002). Note that, since handwriting disorders are often associated with dyspraxia, handwriting tests are proposed here as a complementary analysis in order to refine the evaluation of graphical disorders, thus providing qualitative clues complementary to the dyspraxia diagnosis process. To avoid the multitude of factors that can influence handwriting (such as deficits in motor planning, attention, and language, to cite a few), handwriting tasks are assessed without spelling constraints (i.e., the child is asked to write his or her first name, "papa," "maman," or other well-known words). Handwriting examination is one part of a wider outcome that assesses spatial abilities and constructional praxes by means of drawing tests.

Developmental coordination disorders may produce low graphomotor production speed or persistent deformations of strokes and loops. The interest here is in characterizing handwriting gestures by providing information on processes in terms of sequences and planning, such as the number of pen pauses. Through this analysis, we aim to examine the effect of dyspraxia on handwriting and determine how handwriting disabilities decrease under rehabilitation. We also want to have a means for finding that handwriting aspects are affected according to dyspraxic level, even if handwriting is not directly part of classical tests for evaluating dyspraxia.

The examples of production shown in Figures 4B–4E for both drawing and handwriting tasks illustrate the great diversity of drawings, which makes high-level feature extraction difficult.

Details of the two "gold standards" that have been used as baseline data in the development of the system are described in the Feature Specification and Structuring section below.

Figure 4. Models for the drawing test (A), as well as examples of drawing (B and C) and handwriting (D and E).

Using these two conditions as developmental examples, the design principles described above have been used as the basis for developing a generic patient evaluation system for computer-based neuropsychological hand-drawing testing, which is discussed in the remainder of this article.

STRUCTURAL CONSIDERATIONS FOR A GENERIC CLINICAL TOOL

A detailed description of the configuration of the proposed system is shown in Figure 5. Designed according to the structure and criteria discussed in the previous two sections, the system comprises three main elements. Data are captured using a standard graphics tablet, write-on screen, or tablet PC (as alternative capture technologies). Data from these devices are stored in separate files for each drawing response as a series of time-stamped pen coordinates and normalized pressure samples.

This data capture module is highly configurable, in that the number of individual tasks, the use of any “back projection” (overlay) images with the tablet PC, and the number of such overlays can be specified within an external text

file for ease of modification and creation. The data capture module also accesses a subject database to enable the creation and modification of anonymized subject data.

The pen capture files are read by the analysis module. Again highly configurable, this module consists of a separate plug-in executable file with a well-defined input/output interface, to allow for future extensions (making changes or additions to existing tests or creating entirely new tests) to be written with little or no modification to the rest of the system. This module can analyze specific overlay data individually or, alternatively, the test (i.e., the complete set of overlays) as a whole by looking at the relevant pen capture files created at the data capture stage. This module contains a classification/marketing system that has the ability to read configuration parameters from an external file. Once the analysis is made, the results can be written to results files. These can include test results (i.e., pass/fail or a predicted continuous marked score) as well as individual feature values, such as pen speed profile, acceleration, and so forth.

The final module, the display/playback module, contains several functions. As well as producing performance

Figure 5. Detailed system structure.

reports, this module also enables individual pen capture files to be accessed directly, allowing real-time replay of patient responses (such as observing sections in which the pen was on or off the tablet surface) and graphical feature profiling (such as pen velocity against time, pressure values against time, etc.). Importantly, this module allows a clinician not only to analyze the classification/score assigned to a test battery from the analysis module, but also to drill down into individual feature values to explore a particular aspect of the test performance.

Each module is now described in detail.

A Robust Data Acquisition Strategy

The capture of handwritten data is an essential part of this system, and in order for the system to be as general purpose as possible, the data capture element must also be general purpose and extensible. A discussion on the design ethos and implementation strategies of the capture element is presented here.

Capture devices. To allow for maximum generality, support has been implemented for both a Wintab-compliant graphics tablet and a tablet PC or write-on monitor. When using a tablet PC rather than conventional paper to capture patient input, drawings are made directly onto a laptop screen that can be folded flat onto the table surface. As the subject draws on the screen, the drawing is displayed in virtual "ink," thereby providing visual feedback on the drawing process. Pressure is simulated by varying the width of the virtual ink. Although this approach deviates from the conventional test infrastructure, it has advantages in terms of speed (new test overlays can be instantly displayed on

the screen rather than having to replace a sheet of paper), cost of paper-based overheads, and portability.

Orientation. Some tasks require that the overlay or screen be physically oriented differently (e.g., handwriting tasks require portrait instead of the more common landscape orientation used for drawings). In order to support this functionality, the capture software must adjust the saved x - and y -coordinates for the tests that are captured in portrait in order to enable the correct extraction of data. To implement this, the tablet PC screen or graphics tablet is rotated 90° counterclockwise, and the saved data are adjusted accordingly.

Extensibility. We have indicated above a variety of different clinical conditions in which pen-and-paper-type tests are used as a recognized method of diagnosis. The data capture module allows for ease in implementation of new pencil-and-paper tests developed by the medical community, as well as for the modification of existing test batteries. To facilitate this, we have defined a configuration file format that allows for the specification of all elements of the test. These are as follows:

- **Overlay title**
- **Overlay images (bitmap filename)**
- **Is the test timed? (if so, time limit per overlay)**
- **Portrait or landscape orientation**

This configuration file format results in a straightforward procedure to create and modify tests as and when needed, either to complement existing testing procedures or to allow for the creation of tests for new medical conditions.

Integration. Most importantly for a generalized clinical tool for the capture and automatic scoring of subjects' test performance, the capture element must be fully integrated within the framework of the system. Examples of this include the abilities to edit subject data and to analyze and play back any captured tests.

Feature Specification and Structuring

The neuropsychological test batteries implemented within the assessment system currently proposed are specifically aimed at the evaluation of the two conditions of visuospatial neglect and dyspraxia in children. Both conditions are assessed by analyzing drawings captured dynamically.

To enable the development of the automatic classification system for visuospatial neglect, test data were collected from a series of CVA patients at four clinical test centers within the county of Kent in the U.K. Test subjects were first assessed using the Behavioural Inattention Test (BIT; Wilson et al., 1987) in order to obtain a baseline visuospatial neglect "severity" score. To ensure consistency within the administration and assessment of the BIT, data collection and assessment were conducted by a single therapist. On the basis of the individual BIT scores, test subjects were categorized into two groupings: neglect (subjects scoring below the predefined BIT threshold of 130) and stroke controls (subjects scoring 130 and above). Groupings were confirmed using CT scan information and other clinical evidence. Alongside the CVA patients, a group of 13 healthy age-matched control (AMC) subjects were also assessed. The AMC subjects were free from cerebrovascular diseases with normal visual-motor coordination. No significant relationships were identified between the BIT score and either (1) subject gender or age or (2) the length of the time between hemorrhage and diagnostic examination.

After the collection of BIT test data, test subjects were required to complete a series of 14 commonly used pencil-and-paper tasks (mirroring those found within the BIT)

on a piece of paper overlaid on a standard graphics tablet. They completed the following tasks: the OX cancellation task (with one overlay); the Albert (1973) cancellation task (one overlay); the figure completion task (diamond, man, and house; six overlays); the figure copying task (cross, cube, square, and star; four overlays); and the drawing-from-memory task (square and cube; two overlays). See Liang, Guest, Fairhurst, and Potter (2007) for full implementation details of these tasks.

For the investigation of dyspraxia, data were collected in two separate centers in Rouen, France: first in a series of schools, to obtain a reference population, and second within a clinical environment, to obtain a sample of dyspraxic children. In both cases, children performed a series of handwriting tests alongside the Bender (1946) and BHK (Charles et al., 2003) tests on a graphics tablet. The Bender and BHK tests were assessed manually by clinicians with reference to the standard scoring guide. A total of 8 dyspraxic children were collected alongside 178 "reference" subjects from schools. Full details on data collection and feature extraction for the assessment of dyspraxics can be found in Glénat, Heutte, Paquet, Guest, et al. (2007).

In the design of the automated assessment of these conditions, it was important to consider elements of commonality in measurement between the two conditions. This commonality may be shared across all tasks, a specific subset of tasks, or a specific subset of shapes. In order to efficiently accommodate this commonality, a set of common global features has been experimentally identified (Glénat & Linnell, 2006) that could be used to assess both conditions. This consists primarily of a set of dynamic features that are common to the majority of the tasks that provide a universal set applicable to any condition that may be evaluated through pen-based tests.

Alongside these global features are a number of task-specific measurements (i.e., the data extracted are only applicable to a particular test or series of tasks). A representation of feature groupings and interactions for the two

Figure 6. Feature interaction and groupings for the dyspraxia and neglect conditions.

conditions of immediate interest is provided in Figure 6. It is clear that the shapes, tasks, and conditions that must be extracted from the system overlap in several areas.

The *common* feature set includes features such as total time for the execution of the drawing, number of segments in the drawing (defined by monitoring pen-up events), pen-up/pen-down time ratio, total distance traversed by the pen, average pen pressure and pen average speed across the drawing, and so on. A *group-specific* set contains features, also experimentally identified (Glénat & Linnell, 2006), that may be common to a particular group of similar tests (e.g., drawing or handwriting tests) and are common to the analysis of specific subtasks within both conditions. Furthermore, *test-specific* features are common features that may be extracted from all tasks contained in a certain test, whereas *task-specific* features would be those specific to a particular task, such as the number of correctly canceled lines in Albert's (1973) cancellation task. Finally, we may define as *shape-specific* features the high-level features that accurately describe a particular shape, which may be found among different tests of a condition or between different conditions.

Some important issues should be considered if we are to ensure that a system is able to optimally manage the high diversity of tests that need to be analyzed. The two specific conditions considered for illustration here are very different, and the type of tests used to evaluate them are very diverse, whereas a wide variety of tests is used to evaluate each condition alone. This diversity requires the

implementation of robust algorithms able to accommodate the diversity and individual requirements of separate tasks within a more general framework. For example, it is common for the features extracted to characterize neglect to divide the patient response into left and right sides of the visual field, with the features for each side extracted separately, whereas a dyspraxia diagnosis generally stresses the use of geometrical features.

The list of common global features extracted for both conditions is described in Table 1. Some of the listed features may be group-specific for drawing tasks but are treated within the application framework as common global features, for easy access and robust implementation. In the feature selection section, the common global features required in each test are included within the group-specific features for that test in order to ease the selection process. Any future extensions to the analysis module to support additional medical conditions will require only a few modifications; new features can be implemented with ease alongside the existing global features available for use.

Because of the modular nature of the system and the defined interface between the modules, new analysis modules can be written to assess different conditions and tasks (using generic and new task-specific features) without the need to completely reconstruct the analysis application. This also enables the application to be used for the assessment of other hand-drawn or writing data, such as signatures and forensic information.

Table 1
Global Features

Feature	Explanation
Total time	The total time taken to complete the overlay
Number of segments	The number of distinct pen strokes used
Pen-up/pen-down ratio	The ratio of the time spent with the pen on the paper and off the paper
Total pen-up duration/number of pen-ups	The total time with the pen off the paper divided by the number of times the pen is taken off the paper
Pen-up proximity duration/total pen-up duration	Total time when pen is in the proximity (within 2 cm of the writing surface) divided by the total time the pen is not in contact with the writing surface
Duration of pen-down pauses/# pauses	The total time the pen pauses when writing divided by the number of times it pauses
Number of pen-down pauses	The number of times the pen stops moving while in contact with the paper
Average pressure*	The mean of the pressure values obtained for all pen-down packets
Std pressure*	Standard deviation of pressure
Total length	The total distance the pen moves during the test when the pen is down
Average direction	The average direction the pen has traveled in
Width/height ratio	The ratio between the width and height of the drawn shape
Average speed**	Mean speed of the pen
Max speed	Maximum value of the speed of the pen
Average acceleration	Mean acceleration of the pen
Max acceleration	Maximum acceleration of the pen
Average jerk	Mean jerk (change in acceleration over time) of the pen
Max jerk	Maximum jerk (change in acceleration over time) of the pen
# ux zero-crossings	Number of velocity-in-the-x-axis zero crossings
# uy zero-crossings	Number of velocity-in-the-y-axis zero crossings
# ax zero-crossings	Number of acceleration-in-the-x-axis zero crossings
# ay zero-crossings	Number of acceleration-in-the-y-axis zero crossings
# jx zero-crossings	Number of jerk-in-the-x-axis zero crossings
# jy zero-crossings	Number of jerk-in-the-y-axis zero crossings
Average ux > 0	Average positive x-axis velocity
Average ux < 0	Average negative x-axis velocity
Average uy > 0	Average positive y-axis velocity
Average uy < 0	Average negative y-axis velocity

*These features appear in a normalized form. **These features appear in an absolute form.

Analysis and Feedback of Clinical Indicators

The features that are extracted are applied to a classification scheme in order to obtain a score. For example, for the neglect condition, we wish to approximate the score that the patient would obtain by undertaking a standard BIT (Wilson et al., 1987). The BIT is a battery of pencil-and-paper-based tasks conducted by clinicians in order to assess the patient on a set of standard criteria (based on the outcomes of the drawing tasks), resulting in a BIT score value between 0 and 146. The classification method used for the neglect assessment scheme comprises a multistage neural network topology (see Stasiak, Guest, Linnell, & Fairhurst, 2006, for further details). For the assessment of dyspraxia, the system aims to approximate the score that would be given by the clinician to the five figure copies (as shown in Figure 4), according to the standard scoring guide (Zazzo, 1967). The classification method used in this case is a combination of feature extraction and multi-linear regression (Glénat et al., 2005); information about score approximation confidence can also be given to the clinician.

The results of the analyses are stored along with a subset of the extracted features that may be of interest to clinicians—such as average pressure, time to complete the test, and so on—so that the clinician or user of the software can view these results later in a presentable form, in the *playback module*. This module allows the user to choose a test that has already been taken and analyzed and view the patient actions during the test by playing them back, as if the test were being executed in real time. It is possible to speed up or slow down the replay of the test execution pat-

tern, zoom in and out of the graphical representation of the pen movements in order to show sections in greater detail, and show pen-up as well as pen-down movements—as long as the pen is within 1 cm of the capturing tablet, its movement is recorded. On this same screen, the researcher can then view the saved feature values calculated during the analysis stage and also view some graphical feature data, such as a speed profile graph showing how a subject varied in speed of response throughout the test. This ability to view analysis data during playback is illustrated in Figures 9 and 10 below, where the pressure placed on the drawing surface is plotted against time and a range of common features are displayed.

IMPLEMENTATION

Figure 7 shows the graphical user interface for the main part of the application. The emphasis here is on usability, hence the choice of large illustrated buttons and textual instructions. This is the main hub of the software, which allows the user to launch the capture, analysis, and playback modules.

Of particular interest to clinicians is the ability to replay the test responses in real time and produce a graphical representation of velocity and acceleration profiles within a drawing response (these functions are implemented in the playback and analysis modules, respectively). Figure 8 is an example of a test being played back in the playback module. In this example the pen-up movements can be seen as well as the pen-down movements. Figure 9 shows the pen pressure profile of a test response within the anal-

Figure 7. The graphical user interface of the application.

Figure 8. Real-time playback of test response.

ysis program. Time-based graphs can be displayed for pen pressure, velocity, and x- and y-coordinates, in order to visualize the change of a feature value over the course of the test.

An important aspect of the use of the application in a clinical setting is the presentation of the analysis and the results. As well as examining a “top-line” pass/fail or percentage result from the system, the clinician also has the

opportunity to explore individual feature measurements that may provide additional clinical information and performance comparison data, alongside the overall classification mark. Figure 10 shows this reporting process. For some task-specific high-level features, displaying the segmentation of an image by color enables the highlighting of specific identified features within the drawing, as can be seen in Figure 11.

Figure 9. Graphical pressure profiling.

Results	
Features Pen X Coordinate Profile Pen Y Coordinate Profile Pen Velocity Profile Pen Pressure	
Feature	Value
Global Timing and Pen-down/Pen-up Features:	
Drawing Duration (s)	12.834000
Number of Drawn Segments	5
Total Pen Contact Time (s)	8.410000
Total Non-Contact Time within Drawing Process (s)	4.424000
Ratio of Pen Drawing Time to Non-Contact Time	0.526040
Total Non-Contact Time With Pen in the Tablet Proximity Area (s)	4.140000
Number of Pauses with Pen in Contact	4
Average Pause Time with Pen in Contact (s)	0.150000
Average Normalised Pressure	0.735684
Global Spatial Features:	
Drawing Offset Distance from Left Border (mm)	33.010000
Drawing Offset Distance from Top Border (mm)	28.330000
Total Width of Drawing (mm)	34.690000
Total Height of Drawing (mm)	24.960000
Ratio of drawing width to drawing height	1.389824
Total Pen Travel Distance (mm)	116.240000
Global Dynamic Features:	
Average Pen Drawing Speed (mm/s)	13.821641
Maximum Pen Drawing Speed (mm/s)	79.000001
Maximum Pen Drawing Acceleration (mm/s ²)	0.790000
Detected Circle Features:	
Number of Drawn Segments Belonging to the Detected Circle	1
Average Radius (mm)	8.490000
Minimum Radius (mm)	7.200000
Maximum Radius (mm)	9.690000

Figure 10. Result reporting.

MEDDRAW : Playback

File Help

Test Info

Subject:

Age: 9 years 3 months

Gender: F

Test Group: Patient

Writing Hand: Right

Device Type: Graphics Tablet

View Specific Fields

Overlay to playback:

Bondor1.tst

Show Score Results

Colour Legend

- Detected circle
- Detected square
- Unlabelled pen-downs
- Pen-ups
- LMS approximated circle

Features

Feature	Value
Detected Circle Features:	
Number of Drawn Segments Belonging to the Detected Circle	1
Average Radius (mm)	17.54
Minimum Radius (mm)	12.36
Maximum Radius (mm)	20.50
Ratio of Radius Variation to Maximum Radius	0.40
Radius of the LMS Approximated Circle (mm)	17.68
Detected Square Features:	
Number of Drawn Segments Belonging to the Detected Square	5
Estimated Number of Sides in the Detected Square	9
Average Side Length (mm)	11.21
Ratio of Side Length Variation to Maximum Side Length	0.80
Average Deviation of the Side Direction from the 45° and 135° Axes	0.00
Average Pen Drawing Speed (mm/s)	7.05
Maximum Pen Drawing Speed (mm/s)	112.22
Maximum Pen Drawing Acceleration (mm/s ²)	1.25

Figure 11. High-level feature display.

Initial results from the system demonstrate that the devised analysis framework is able to accurately detect patient groupings within both of the implemented condition analysis modules and, furthermore, to produce a significantly correlated quantifiable score against the "gold standards" for the respective conditions. Full details of the results are contained in Stasiak et al. (2006) and Glénat and Linnell (2006). Through these results, we have demonstrated both the content and construct validity of the devised system.

DISCUSSION AND CONCLUSIONS

Seeking commonalities in diagnostic or assessment procedures can be a powerful catalyst for developing effective and efficient clinical procedures, especially where an assessment system embodies a sufficient degree of flexibility in its structure to accommodate local variability in patterns of use and individual differences in testing protocols.

It is apparent that a cluster of patient populations exists for which a significant and exploitable degree of commonality can be readily identified in relation to the requirements for (1) the implementation of pencil-and-paper tests such as figure copying, figure completion, and so on, and (2) the analysis of the outputs generated by patients in varying test environments. Although significant differences occur across such generally adopted tests, they share many characteristics that arise from the need to analyze different elements of the handwriting or drawing processes in the patients undergoing testing.

We have developed a technique that has been applied individually to a range of clinical conditions; in previously reported work, we have shown how this approach can enhance many aspects of traditional testing procedures (see, e.g., Glénat, Heutte, Paquet, & Mellier, 2005, 2007). Moreover, this work has clearly demonstrated that, by adopting online data capture and more sophisticated analysis of the behavioral characteristics thereby observable, information of clinical value yet hitherto unobtainable can now be readily accessed from traditional tests.

In this article, however, we have specifically introduced a new perspective on this type of investigation, by examining the principal issues that underlie a strategy for integrating the research strands relating to different clinical populations, as a means of developing a powerful generic clinical tool that can better facilitate established routine clinical testing with respect to individual conditions of clinical importance and can also be readily extended to other conditions possibly using similar testing procedures.

In particular, we have shown that a substantial degree of commonality exists in the behavioral characteristics that are of clinical significance across different patient populations, but also that test-dependent differences need to be preserved if the benefits of different assessment procedures are to be maximized. We have examined these issues in relation to two separate clinical populations and two typical test protocols associated with these groupings in order to illustrate these commonalities and differences. Most importantly, we have discussed in detail how such an

investigation can guide and inform the specification and implementation of a powerful generic assessment system that can both improve the effectiveness of current patient assessment procedures and extend and enhance the potential capabilities of these procedures.

Our approach benefits substantially from its degrees of flexibility and generality, but one further advantage can potentially be gained from its use because, although the structure we have described here supports the effective execution of standard established tests with respect to a particular population of interest, the availability of related modes of feature extraction and analysis within the same framework can stimulate experimentation and new insights into the assessment and diagnostic processes undertaken.

Given the generic nature of our framework, a number of possible directions for future work are available using the system we have developed, such as the following:

1. Researchers may explore additional neuropsychological conditions, utilizing pen-and-pencil-based tests as diagnostic indicators. In particular, the analysis of novel dynamic features that have typically not been assessed may lead to an enhancement in diagnostic resolution. Furthermore, because of the objective nature of the framework in terms of algorithmic assessment, in future work we aim to explore the replicability over time of the scoring mechanism results within the conditions explored in this prototype.

2. We will continue to work with clinicians on expanding the data replay and analysis functionality, the usefulness of which has been confirmed. By this means, we will provide further means for in-depth analysis and investigation, particularly with respect to dynamic features hitherto inaccessible.

3. The proposed framework enables the integration of novel features and analysis methodologies within the data capture system, thereby allowing applications to contexts such as biometric technologies (e.g., signature verification and authentication) and document forensics, through the use of dynamic (and inferred dynamic) constructional analyses of handwriting samples.

The approach described here is therefore both a means of improving existing clinical practice and a potentially highly innovative research tool with a range of applications.

AUTHOR NOTE

This study is part of the MEDDRAW Project (www.meddraw.org), sponsored by the Interreg IIIA E.U. Programme, the Government Office for the South East, Kent County Council, U.K., and the "Conseil Régional de Haute Normandie," France. The authors thank the sponsors and our medical partners: Jonathan Potter, Kent and Canterbury Hospital; Daniel Mellier, Psy.Co, Rouen; and Dr. Charollais, Mrs. Lemarchand, and Mrs. Monier, Centre Hospitalier Universitaire de Rouen. We also thank Elina Kaplani for her work on this project. Correspondence relating to this article may be sent to R. M. Guest, Department of Electronics, University of Kent, Canterbury, England (e-mail: r.m.guest@kent.ac.uk).

REFERENCES

- ALBERT, M. L. (1973). A simple test of visual neglect. *Neurology*, 23, 658-664.
- AZOUVI, P., SAMUEL, C., LOUIS-DREYFUS, A., BERNATI, T., BARTOLO-

- MEO, P., BEIS, J.-M., ET AL. (2002). Sensitivity of clinical and behavioural tests of spatial neglect after right hemisphere stroke. *Journal of Neurology, Neurosurgery, & Psychiatry*, 73, 160-166.
- BENDER, L. (1946). *Test moteur de structuration visuelle*. Paris: Editions du Centre de Psychologie Appliquée.
- CHARLES, M., SOPPELSA, R., & ALBARET, J.-M. (2003). *BHK-C: Echelle d'évaluation rapide de l'écriture chez l'enfant*. Paris: Editions & Applications Psychologiques.
- CHINDARO, S., GUEST, R. M., FAIRHURST, M. C., & POTTER, J. M. (2004). Assessing visuo-spatial neglect through feature selection and combination from geometric shape drawing performance and sequence analysis. *International Journal of Pattern Recognition & Artificial Intelligence*, 18, 1253-1266.
- DONNELLY, N., GUEST, R. M., FAIRHURST, M. C., POTTER, J. M., DEIGHTON, A., & PATEL, M. (1999). Developing algorithms to enhance the sensitivity of cancellation tests of visuospatial neglect. *Behavior Research Methods, Instruments, & Computers*, 31, 668-673.
- EDMANS, J. A., & LINCOLN, N. N. (1990). The relation between perceptual deficits after stroke and independence in activities of daily living. *British Journal of Occupational Therapy*, 53, 139-142.
- FAIRHURST, M. C., HOQUE, S., & RAZIAN, M. (2004). Improved screening of developmental dyspraxia using on-line image analysis. In *Proceedings of the 8th World Conference on Systems, Cybernetics & Informatics* (Vol. 3, pp. 160-165). Orlando, FL.
- GLÉNAT, S., HEUTTE, L., PAQUET, T., GUEST, R. M., FAIRHURST, M. C., & LINNELL, T. (2007). *The development of a computer-assisted tool for the assessment of neuropsychological drawing tasks*. Manuscript submitted for publication.
- GLÉNAT, S., HEUTTE, L., PAQUET, T., & MELLIER, D. (2005). Computer-based diagnosis of dyspraxia: The MEDDRAW Project. In *Proceedings of the 12th Biennial Conference of the International Graphonomics Society, Salerno, Italy* (pp. 49-53). Civitella, Italy: Editrice Zona.
- GLÉNAT, S., HEUTTE, L., PAQUET, T., & MELLIER, D. (2007). *Automatic drawing analysis in dyspraxia diagnosis*. Manuscript in preparation.
- GLÉNAT, S., & LINNELL, T. (2006). *MEDDRAW internal report*. Available at www.meddraw.org.
- GUEST, R. M., DONNELLY, N., FAIRHURST, M. C., & POTTER, J. M. (2004). Using image analysis techniques to analyze figure-copying performance of patients with visuospatial neglect and control groups. *Behavior Research Methods, Instruments, & Computers*, 36, 347-354.
- HALLIGAN, P. W., MARSHALL, J. C., & WADE, D. T. (1989). Visuospatial neglect: Underlying factors and test sensitivity. *Lancet*, 2, 908-911.
- KIRK, A., & KERTESZ, A. (1989). Hemispheric contributions to drawing. *Neuropsychologia*, 27, 881-886.
- LANDGREN, M., PETTERSSON, R., KJELLMAN, B., & GILLBERG, C. (1996). ADHD, DAMP and other neurodevelopmental/psychiatric disorders in 6-year-old children: Epidemiology and co-morbidity. *Developmental Medicine & Child Neurology*, 38, 891-906.
- LEZAK, M. D., HOWIESON, D. B., & LORING, D. W. (2004). *Neuropsychological assessment* (4th ed.). Oxford: Oxford University Press.
- LIANG, Y., GUEST, R. M., FAIRHURST, M. C., & POTTER, J. M. (2007). Feature-based assessment of visuo-spatial neglect patients using hand-drawing tasks. *Pattern Analysis & Applications*, 10, 361-374.
- MONTHEIL, M. C. (1993). *Manuel de la feuille de dépouillement de la figure de Rey*. Paris: Editions du Centre de Psychologie Appliquée.
- NFERNELSON (2003). *Bender visual-motor gestalt test* (2nd ed.). London: nferNelson.
- PORTWOOD, M. (2000). *Understanding developmental dyspraxia: A textbook for students and professionals*. London: David Fulton.
- POTTER, J., DEIGHTON, A., PATEL, M., FAIRHURST, M. C., GUEST, R. M., & DONNELLY, N. (2000). Computer recording of standard test of visual neglect in stroke patients. *Clinical Rehabilitation*, 14, 441-446.
- RÉMI, C., FRELICOT, C., & COURTELLEMONT, P. (2002). Automatic analysis of the structuring of children's drawings and writing. *Pattern Recognition*, 35, 1059-1069.
- REY, A. (1959). *Test de copie et de reproduction de mémoire de figures géométriques complexes*. Paris: Editions du Centre de Psychologie Appliquée.
- STASIAK, A., GUEST, R. M., LINNELL, T., & FAIRHURST, M. C. (2006). *Automatic estimation system of the neuropsychological state of patient after cerebral hemorrhage*. Manuscript in preparation.
- STONE, S. P., WILSON, B., WROOT, A., HALLIGAN, P. W., LANGE, L. S., MARSHALL, J. C., & GREENWOOD, R. J. (1991). The assessment of visuo-spatial neglect after acute stroke. *Journal of Neurology, Neurosurgery, & Psychiatry*, 54, 345-350.
- WILSON, B., COCKBURN, J., & HALLIGAN, P. (1987). *Behavioural inattention test*. Titchfield, U.K.: Thames Valley Test.
- ZAZZO, R. (1967). *Manuel pour l'examen psychologique de l'enfant* (6th ed.). Neuchâtel, Switzerland: Delachaux & Niestlé.

(Manuscript received October 19, 2006;
revision accepted for publication July 19, 2007.)