# A method for selecting an efficient diagnostic protocol for classification of perceptive and cognitive impairments in neurological patients

**Kunjan D. Rana**,
Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA, USA

**Benvy Caldwell**, and
Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA, USA. She is now with the Boston University, School of Public Health

**Lucia M. Vaina**
Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA and Harvard Medical School, Massachusetts General Hospital, Department of Neurology, USA (phone: 617-353-2455; fax: 617-353-6766)

Kunjan D. Rana: kdrana@bu.edu; Benvy Caldwell: benvy7@gmail.com; Lucia M. Vaina: vaina@bu.edu

## Abstract

An important and unresolved problem in the assessment of perceptual and cognitive deficits in neurological patients is how to choose from the many existing behavioral tests, a subset that is sufficient for an appropriate diagnosis. This problem has to be dealt with in clinical trials, as well as in rehabilitation settings and often even at bedside in acute care hospitals. The need for efficient, cost effective and accurate diagnostic-evaluations, in the context of clinician time constraints and concerns for patients' fatigue in long testing sessions, make it imperative to select a set of tests that will provide the best classification of the patient's deficits. However, the small sample size of the patient population complicates the selection methodology and the potential accuracy of the classifier. We propose a method that allows for ordering tests based on having progressive increases in classification using cross-validation to assess the classification power of the chosen test set. This method applies forward linear regression to find an ordering of the tests with leave-one-out cross-validation to quantify, without biasing to the training set, the classification power of the chosen tests.

## I. Introduction

This paper introduces a general method for selecting and ordering a set of behavioral perceptual-cognitive tasks that can be administered in sequence to a patient for diagnosis with each subsequent test providing optimal additional classification power amongst the remaining tests. It is necessary to diagnose perceptual-cognitive deficits in neurological patients based on the results of a well-selected battery of quantitative tasks that is easy to administer, and validated in a large population. However, the amount of testing a patient is able to bear is variable and often times unknown apriori. Thus, if a patient were unable to continue testing mid-battery, the set of tests completed should still allow for a low error rate in diagnosis with some optimality. Several test batteries exist for assessing a wide range of neurological impairments[1–3], but they suffer from a lack of ordering that would allow for appropriate classification when only a subset of the battery is complete. We propose a quantitative method that establishes a sequence of tests that provides optimal additional classification power per additional test administered.

Our method employs forward selection to iteratively find tests that improve classification rate through linear discriminants. Due to the large variances in patient performance as well as the difficulty of obtaining a sufficiently large number of patients, leave-one-out cross validation (LOOCV)[4] is used to assess the performance of the data. The resulting accuracy of classification from LOOCV is utilized to measure the gains in classification with the addition of each new test dimension.

In this paper, we use as an example dataset the patient results from a series of tests called the low-level visual motion screening battery[5]. The selected stroke patients from this study have completed the entire battery and are categorized by the anatomical locations of their lesion. For simplicity we only included patients with unilateral lesions.

## II. Methods

### A. Test Selection and Ordering

A typical method for selection of tests based on classification accuracy is stepwise regression. In the case of requiring a set of tests with progressive improvements in accuracy, one way apply the forward linear regression procedure. Forward linear regression uses the estimated group covariances across a subset of tests to generate linear boundaries. Initially, all tests are assessed separately and the one producing the greatest separation between classification groups is labeled as the most significant test. Then, the separation is recomputed amongst the remaining tests with the addition of the most significant test to choose the second-most significant test. This procedure is repeated, while recomputing separation amongst each remaining test along with the selected significant tests, until no tests are remaining.

The problem with this method is that it is susceptible to "Type III errors," that is the classification accuracy is biased by its training data. In other words, we may build a classifier that may fit all training data with 100% accuracy, but would not necessarily have such accuracy when tested with new data. Ideally, we could test with new data. However, in our class of problems, we have only a small set of labeled data due to the difficulty of obtaining patients for study. Thus, we need to rely on cross-validation methods to test data. Note that since we are simply obtaining an ordering of tests, we simply need to know how "new" data (in this case the validation set) is classified based on the rules of generating our classifier given a subset of the tests.

In our algorithm (see Alg 1), the classifier uses leave-one-out cross validation (LOOCV) nested inside the forward selection method. Each iteration of LOOCV (step 7i) classifies a single subject while using the rest of the labeled dataset to compute the optimal classifier. With our dataset, we use linear discriminant analysis for our linear classifier. The classification error across all iterations is averaged together (step 7ii) to produce the error rate for the current set of tests. The test with the best accuracy is added to the ordered set of tests (steps 7iii and 8), which is the output of the algorithm after all tests have been tested.

### B. Training Dataset

The training data set consists of the result of forty-nine patients with a first ever unilateral, cortical infarct. The low-level motion screening battery is used to assess deficits in motion processing ability. Six tests, described in the following section and taken from our previously published studies [5–19], are part of the low level visual motion battery, and each is administered in two conditions, ipsilesional and contralesional, resulting in a total of twelve tasks used in the analysis. The ipsilesional and contralesional conditions refer to the visual field in which the stimulus is presented. For example, ipsilesional means that the test stimuli are presented in the visual field on the same side as the lesion (i.e. left hemisphere

lesion, stimuli presented in the left hemifield). Patients are divided into four classification groups based on their lesion location: Occipital-temporal, Occipital-parietal, Dorsal-parietal, and Frontal.

## C. Low-level Visual Motion Screening Battery

**1) Direction discrimination—**All the dots in the stimulus moved upwards and at a variable angle to the left or right of true vertical (Figure 1a), which was indicated by a short clearly visible line placed 0.5° above the display aperture. In a two alternative forced choice (2AFC) procedure, subjects reported whether the dot-field moved to the right or to the left of the vertical line. Threshold was the angle at which performance was 79% correct.

**2) Speed discrimination—**This task measured the perception of relative speed of two random dot kinematograms (RDK) which shown schematically in Figure 1c. The RDK's were displayed sequentially, with a 500 ms inter-stimulus interval. In each interval, every dot's trajectory changed randomly from frame to frame, but the speed was the same for all the dots. The variable was the ratio of speed difference between the two intervals. The standard speed, presented first or second at random, was 3°/s and the speed in the other interval varied from trial to trial, starting from a maximum of 6°/s (ratio = 2). In a two temporal alternative forced choice procedure, subjects reported in which interval (the first or the second) the dots moved faster. Threshold was the speed ratio at which performance was 79% correct.

**3) Motion coherence—**This stimulus display, adapted from [20], was designed to isolate motion-sensitive mechanisms by using a controlled motion signal whose strength did not alter the average spatial and temporal structure of the stimulus (as adapted by [14] from [20]). The display (schematized in Figure 1e) consisted of stochastic RDKs in which a specifiable percentage of the dots had a constant velocity and correlated motion signal while the remainder moved in random directions at random speeds, providing masking motion noise. The strength of the motion signal was varied by changing the percentage of dots moving coherently between 0 (just noise) and 100 (all dots are signal and move in the same direction). In each frame, the position of the noise dots was random, and at 0% coherence the display appeared as a fluctuating pattern of spatiotemporal noise. The motion content of the display (direction) could be extracted only by integrating brief local motion signals over time and space[21, 22]. In a four alternative forced choice task, subjects reported whether the overall direction of the RDK was up, down, left, or right. Threshold was the percentage of signal dots at which direction discrimination (DDT) was 79% correct.

**4) Motion discontinuity—**The display was an RDK with identical statistical properties to that described in Exp 3 except that in half of the trials (discontinuous) an illusory line divided the display into two equal fields of dynamic random dots (Figure 1g) and the other half the trials (homogeneous) contained no such division. The signal dots moved upwards or downwards. The illusory line arose from the opposite direction of motion of the 'signal' dots within the two halves of the stimulus aperture. To prevent any use of spatial local cues, the illusory line had four possible orientations and the centre of the line was slightly (less than 0.5°) and randomly offset from the centre of the stimulus aperture. In a 2AFC task, subjects reported whether the display was discontinuous or homogeneous. Threshold was the percentage of signal dots at which subjects could discriminate between the homogeneous and discontinuous displays at 79% correct.

**5) Two-dimensional form-from-motion—**As in Exps 3 and 4, the stimulus was an RDK of variable proportion of signal dots embedded in masking motion noise. A two-dimensional form, defined solely by the relative motion of two oppositely moving fields of

signal dots and resulting in an illusory line outlining a two-dimensional form (either a 'plus' or a 'minus', of equal areas (schematized in Figure 1i) appeared in the centre of the stimulus aperture. In a 2AFC task, subjects reported whether the two-dimensional form was a 'plus' or a 'minus'. Task difficulty was titrated by varying the proportion of signal dots and threshold was the percentage of coherently moving dots where performance was 79% correct.

**6) Motion coherence – radial—**This task is similar to that of Exp 3 except that the signal dots move radially in the frontal plane from centre to periphery (expansion) or the reverse (contraction), illustrated in Figure 1k. To ensure that subjects perceived planar motion, all dots had an equal displacement at all distances from the centre, preventing the depth illusion that radial motion stimuli can produce. The proportion of dots moving coherently and radially was titrated as above and the subject reported whether the pattern was expanding or contracting. Threshold was the percentage of signal dots at which performance was 79% correct.

## III. Results

This method was applied to the dataset from [1] to reduce the number of tests required to screen a patient for visual motion deficits. The results of the validation test using the data set of the four patient groups are shown in Figure 2, where we have plotted the error rate as a function of the tests included in each step of the analysis. Each point represents the error rate of the classifier (1 – accuracy) at predicting the lesion group of each patient. The tests along the x-axis are ordered based on their classification result. The first test selected, Motion Discontinuity contralesional (MDTc), produced the best classification. The second iteration selected the Direction Discrimination contralesional (DDTc) to be the test that, in combination with MDTc, produced the best separation between groups. This process continued to order the remaining tests based on their contribution towards classification. Even in the first iteration with a single test included, the classifier performed better than chance with 50% accuracy (out of four lesion groups) at predicting lesion location.

The error rate drops below 30% by 8 tests. The error rate increases as the number of tests is increased past 9 tests. The reason for this effect is most likely due to the curse of dimensionality. Adding more dimensions that do not provide additional information will increase the noise in the system, thus leading to a higher error rate. Adding additional dimensions in the forward selection procedure likely increases the error. Thus, we can conclude that these additional tests do not provide any additional, useful information for diagnosis and we can exclude them from the diagnostic battery.

## IV. Conclusion

We have proposed a method for ordering a set of psychophysical tests by their significance in classification in terms of the classification error rate. Although the dataset spanned low-level visual motion tasks and included only forty-nine patients, this method can be expanded to include any arbitrary family of behavioral tasks with any number of subjects. The forward selection procedure grows quadratically in the number of tests rather than combinatorially as would an exhaustive search over the test space. Also, the number of LOOCV iterations is proportional to the number of data points. If we had a larger test set, we could utilize other cross-validation schemes to reduce computation time. Since we are validation over a small population, LOOCV does not require too many iterations while reducing the loss of information in the classifier, thus better reflecting classification accuracy given all points.

In the proposed method, we did not make assumptions on how much time each patient is allotted for performing the behavioral tasks. If, however, we knew apriori that each patient can take a set number of tests (or be involved in testing for a prespecified amount of time), we could replace the forward linear regression step with a more complex stepwise regression step that attempts to find the optimal combination of tests for a given set size. In this case, all tests must be administered to the subject. However, this procedure is incapable of ensuring optimality if a patient fails to complete the battery. In the method discussed in this paper, discontinuing in the middle of the testing battery would still include the tests that are progressively optimal in improving classification.

## Acknowledgments

## References

1. Adams RJ, MacNeil DE, Dove C, Courage ML. Optics and psychophysics in a clinical setting: Success of a screening battery for assessing visual functioning in human infants. Journal of Vision. May 10.2008 8:414.

2. Griffiths TD, Dean JL, Woods W, Rees A, Green GGR. The Newcastle Auditory Battery (NAB). A temporal and spatial test battery for use on adult naive subjects. Hear Res. Apr.2001 154:165–9. [PubMed: 11423227]

3. Robbins TW, James M, Owen AM, Sahakian BJ, McInnes L, Rabbitt P. Cambridge Neuropsychological Test Automated Battery (CANTAB): A Factor Analytic Study of a Large Sample of Normal Elderly Volunteers. Dementia and Geriatric Cognitive Disorders. 1994; 5:266–281.

4. Vapnik, VN. Estimation of dependences based on empirical data. New York: Springer-Verlag; 1982.

5. Vaina LM, Sikoglu EM, Soloviev S, Lemay M, Squatrito S, Pandiani G, Cowey A. Functional and anatomical profile of visual motion impairments in stroke patients correlate with fMRI in normal subjects. J Neuropsychol. Oct 8.2009

6. Vaina LM. Selective impairment of visual motion interpretation following lesions of the right occipito-parietal area in humans. Biological Cybernetics. 1989; 61:347–359. [PubMed: 2790066]

7. Vaina LM, Grzywacz NM, LeMay M. Structure from motion with impaired local-speed and global motion-field computations. Neural Comp. 1990; 2:420–435.

8. Vaina LM, LeMay M, Bienfang DC, Choi AY, Nakayama K. Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: a case study. Visual Neuroscience. Oct.1990 5:353–369. [PubMed: 2265150]

9. Vaina L. Functional Segregation of Color and Motion Processing in the Human Visual Cortex: Clinical Evidence. Cerebral Cortex. 1994; 5:555–572. [PubMed: 7833656]

10. Vaina LM, Grzywacz NM, Kikinis R. Segregation of computation underlying perception of motion discontinuity and coherence. Neuroreport. Nov 21.1994 5:2289–2294. [PubMed: 7881048]

11. Vaina LM. Akinetopsia, achromatopsia and blindsight: recent studies on perception without awareness. Synthese. 1996; 105:1–19.

12. Cowey A, Vaina LM. Blindness to form from motion despite intact static form perception and motion detection. Neuropsychologia. 2000; 38:566–578. [PubMed: 10689034]

13. Vaina LM, Soloviev S, Bienfang DC, Cowey A. A lesion of cortical area V2 selectively impairs the perception of the direction of first-order visual motion. NeuroReport. Apr 7.2000 11:1039–1044. [PubMed: 10790879]

14. Vaina LM, Cowey A, Eskew RT, LeMay M, Kemper T. Regional cerebral correlates of global motion perception: Evidence from unilateral cerebral brain damage. Brain. Feb.2001 124:310–321. [PubMed: 11157558]

15. Vaina LM, Cowey A, LeMay M, Bienfang DC, Kikinis R. Visual deficits in a patient with 'kaleidoscopic disintegration of the visual world'. European Journal of Neurology. Sep.2002 9:463–477. [PubMed: 12220377]

16. Vaina LM, Gryzwacz NM, Saiviroonporn P, LeMay M, Bienfang DC, Cowey A. Can spatial and temporal motion integration compensate for deficits in local motion mechanisms? Neuropsychologia. 2003; 41:1817–1836. [PubMed: 14527545]

17. Royden CS, Vaina LM. Is precise discrimination of low level motion needed for heading discrimination? Neuroreport. Apr 29.2004 15:1013–7. [PubMed: 15076725]

18. Vaina LM, Gross CG. Perceptual deficits in patients with impaired recognition of biological motion after temporal lobe lesions. Proceedings of the National Academy of Sciences of the United States of America. Nov 30.2004 101:16947–16951. [PubMed: 15556997]

19. Vaina LM, Cowey A, Jakab M, Kikinis R. Deficits of motion integration and segregation in patients with unilateral extrastriate lesions. Brain. Sep.2005 128:2134–45. [PubMed: 15975945]

20. Newsome WT, Paré EB. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). Journal of Neuroscience. 1988; 8:2201–2211. [PubMed: 3385495]

21. Downing CJ, Movshon JA. Spatial and temporal summation in the detection of motion in stochastic random dot displays. Invest Ophthal Visual Sci (Suppl). 1989; 30:72.

22. Newsome WT, Britten KH, Salzman CD, Movshon JA. Neuronal Mechanisms of Motion Perception. Cold Spring Harb Symp Quant Biol. 1990; 55:697–705. [PubMed: 2132847]
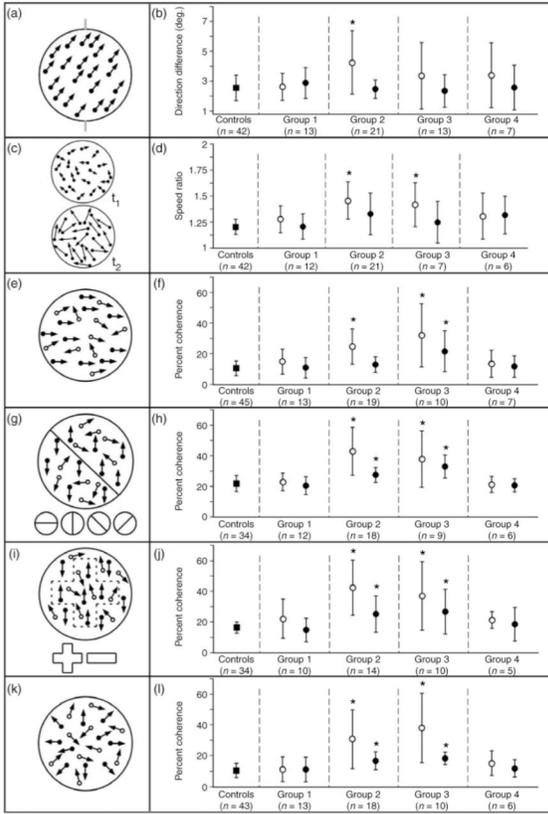
**Fig. 1.**
The visual motion tests and results from the stroke patients and healthy controls. The left column of panels represents schematic views of the visual motion displays. (a) DDT; (c) local SDT; (e) MCT (translation); (g) MDT; (i) 2D-FFM; (k) MCT-radial (RM). In all the tests, each dot is represented as a vector indicating the magnitude and direction of motion. The filled circles represent signal dots (moving in the same direction) and the open circles represent noise dots. The second column of panels (b, d, f, h, j, l) represents the behavioural results for each test for the control subjects and each group of patients. Each data point represents the group mean ± SD of the thresholds obtained for the particular test. The open circles indicate the contralesional visual field and the filled circles indicate the ipsilesional visual field. The '*' symbols illustrate instances where the patients performed significantly worse than control subjects. As there was no statistically significant difference between performance in the right and left visual fields of the controls, the results were combined. This figure and additional information can be found in [1].
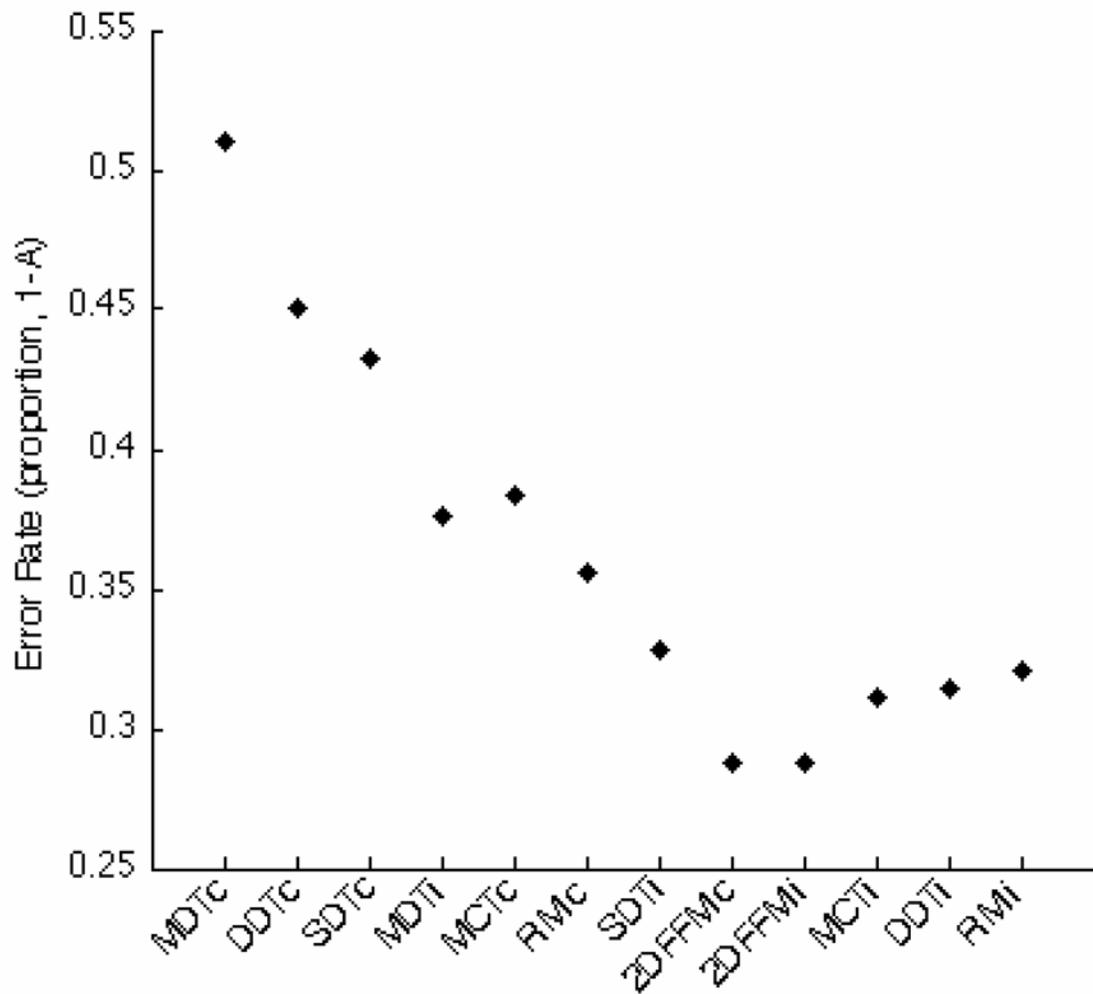
**Fig. 2.**
Results of the classification showing the error rate as a function of the tests included in each step of the analysis. Each point represents the accuracy of the classifier at predicting the lesion group of each patient. The first test is that which produced the best classification alone. The following iteration selected the test that in combination with the first produced the best separation between groups resulting in better accuracy.patients. Each data point represents the classification error (which is 1 – accuracy) obtained for the particular test averaged over all LOOCV steps.

**Algorithm I**

Test Selection and Ordering

| | |
|---|---|
| **1** | U ← {Set of Behavioral Tasks} |
| **2** | V ← ∅ |
| **3** | X ← Subject Behavioral Task Performance |
| **4** | Y ← Subject Classification Labels |
| **5** | T ← ∅ |
| **6** | B ← zero [Best classification accuracy] |
| **7** | For each test index i in U |

      **i.** For each subject j in X

            • X{−j} ← X \ X{j} [All X excluding subject j]

            • Find Classifier F given tests V ∪ U{i} and data X{−j}

            • Classify X{j} using classifier F and compare to Y{j}

      **ii.** Compute accuracy A = # times correctly classified / size of X

      **iii.** If A > B:

            • T ← U{i}

            • B ← A

| | |
|---|---|
| **8** | V{end+1} ← T [test T is appended to the end of V] |
| **9** | U ← U \ T [test T is removed from U] |
| **10** | Go back to 5 until U is empty |

Algorithm 1. An algorithm for finding the optimal progressive ordering of behavioral tasks. The classifier to use is generalized in 7i, so that any classification method can be used. The resulting ordering is found in V with the first element corresponding to the first task to be administered to a patient, which is the test that provides the best classification accuracy.