Classification and Clustering of Brain Injuries from Motion Data of Patients in a Virtual Reality Environment

Natan Silnitsky

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE MASTER DEGREE

University of Haifa
Faculty of Social Sciences
Department of Computer Sciences

September 2011

Classification and Clustering of Brain Injuries from Motion Data of Patients in a Virtual Reality Environment

By: Natan Silnitsky

Supervised by: Prof. Larry Manevitz (Computer Science dept, University of Haifa),
Dr. Uri Feintuch (School of Occupational Therapy, Hadassah-Hebrew University Medical Center)

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE MASTER DEGREE

University of Haifa
Faculty of Social Sciences
Department of Computer Sciences

September 2011

Approved by: ________________ Date: ________
(supervisor)

Approved by: U. Feint Date: 27/9/2011
(supervisor)

Approved by: ________________ Date: ________
(Chairperson of the MSc Committee)

Table of Contents

List of figures

List of Tables

Classification and Clustering of Brain Injuries from Motion Data of Patients in a Virtual Reality Environment

Natan Silnitsky

Abstract

Virtual Reality (VR) has been found to be an effective rehabilitation tool for brain injury patients. We show that motion data from these VR sessions can be effectively used to both cluster and classify patients according to types of injury. Neural Network and other tools were used to differentially classify patients with traumatic brain injury, cerebral vascular accident (stroke) with and without spatial neglect and healthy individuals solely from the motion data. Clustering techniques also successfully duplicated the classification division. These results have potential implications for scientific research, automated diagnosis and integrated individually adaptive therapies in the virtual reality technology.

Portions of this thesis have been presented in the following conferences:

- ICNC, International Conference on Neural Computation, 2010
- BISFAI, Bar-Ilan Symposium on Foundations of Artificial Intelligence, 2011

Portions of this thesis have been publicized in the following conference proceedings:

- Feintuch U., Manevitz L., Silnitsky N. Classification and clustering of brain injuries from motion data of patients in a virtual reality environment. Proceedings of the International Conference on Fuzzy Computation and 2nd International Conference on Neural Computation, 2010

1 Introduction

1.1 Background

1.1.1 Virtual Reality in Rehabilitation

Recent advances in computer science and engineering have allowed scientists and clinicians to introduce virtual reality (VR) technology to various medical fields in general, and to rehabilitation in particular. Virtual reality applications let patients function in simulated environments where they are safe on one side, but practice real-world functions on the other side (see review at Weiss et al., 2006). For example, a stroke patient may practice virtual street crossing in the clinic before trying to cross a street in the physical world (Kats et al., 2005).

Beyond the ecological validity offered by virtual environments, they are also carefully controlled so they can be standardized, and the behavior of the patients is monitored and recorded. The collected data can be analyzed and used for clinical diagnosis or progress evaluation as well as general scientific research.

In the current study we focused, as a proof of concept, at the rehabilitation of brain injuries, and in particular at the population of CerebroVascular Accident (Stroke) patients.

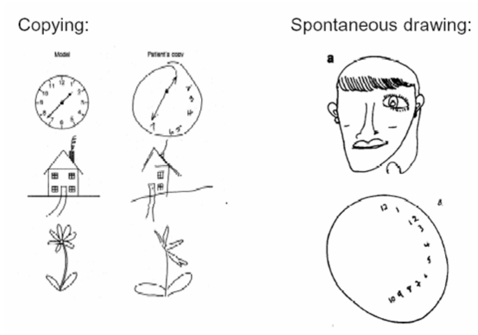


Figure 1. Various Pen and Paper Neglect tests

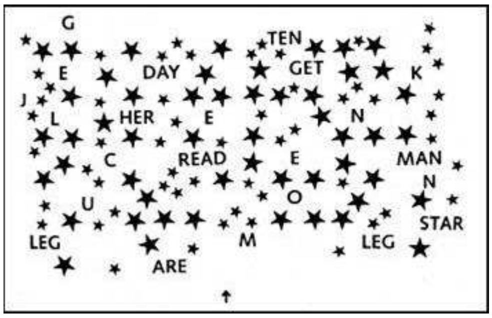


Figure 2. Pen and Paper Neglect test of star cancellation.

A stroke is a lesion of the brain resulting from a disturbance in the blood supply to the brain, due to obstruction or rupture of a blood vessel. Stroke causes a neurological deficit which may lead to various types of disabilities such as cognitive, emotional and motor impairments. In some cases stroke leads to spatial neglect. Patients with neglect are impaired in directing attention to selective part of space, usually the half of space that is opposite the injured hemisphere, and are unaware of their deficit (Robertson and Halligan, 1999).

Neglect is commonly assessed using paper-and-pencil tests (See Figure 1 and Figure 2). However, these tests have several substantial drawbacks that often lead to a misdiagnosis of less severe cases. For example, a stroke patient who had passed the traditional tests and even got back his driver license yet experienced multiple car accidents which occurred due to lack of attention and awareness to the neglected visual hemifield (Deouell, Sacher and Soroker, 2005). Other studies have also shown the weakness of conventional tests in neglect, and the potential of using virtual reality technology for accurate assessment of this neurological condition (Dvorkin et al., 2008, 2011).

Figure 3. The VRROOM 3D platform.

Several types of VR methods are used for the investigation and treatment of stroke. The main one we used for this study implements a 3D environment, where the patient has to reach and "touch" a virtual ball appearing at various spatial locations (see Figure 3). Each reaching trial produces a data vector which includes the x,y,z coordinates and orientations (6 degrees of freedom) of the moving hand at 60 Hz sampling rate.

Beyond the 3D experiment we also used machine learning tools in order to analyze data gathered by a 2D VR system, where subjects perform virtual shopping. In this VR application the data is only two-dimensional and also is very noisy. Finally, in this experiment we included traumatic brain injury (TBI) participants, which constitute another common patient population.

1.1.2 Machine Learning Approach

As virtual reality platforms produce very large amounts of data, many researchers end up reducing the analysis to simple outcome measures such as reaction time, accuracy level etc. Thus meaningful information may be ignored due to the difficulties involved in identifying the relevant clinical features hiding in the data.

We propose that such patient data are prime candidates for analysis using machine learning tools. This study aims to explore how various approaches may be used for analysis of patient data under constraints posed by the clinical conditions.

We focus on analysis based solely on motion data, since this is common in almost all VR platforms.

In analyzing the data produced by these VR platforms, we had to overcome several hurdles. First and foremost, the sample size in these studies is quite small for technical and clinical reasons. Moreover, as the participants get tired with time, the length of each experimental session is rather limited.

Thus we had to find ways to process noisy and scarce data. These issues will be addressed later on. As even simple human motor performance is quite challenging for meaningful analysis, we approached this challenge using three levels of classifications as each one may yield a solution for a different clinical or scientific challenge.

Two-class classification: This approach may be quite valuable when it comes to differential diagnosis (DD). Several types of brain lesions may produce very similar performance, not always easily clear even to a professional eye. Thus it would be of clinical benefit to detect which of the suspected conditions the patient suffers from. After training on some clear cut cases, it may be possible to generalize and classify the more questionable cases. Such classification may assist for example in differentiating between Neglect and Hemianopsia (cortical blindness), as in both conditions patient behave in very similar ways.

Zero-class classification (clustering): Clustering techniques lend themselves for analysis of heterogeneous populations, like stroke patients. Since no two stroke patients are identical, clustering them into subclasses leads to better classification than the coarse ones used today, i.e., mild, severe, and so forth.

One-class classification: One-class filters are those that are trained and produced using only data from one-class, yet it produces a classification on new data that says the data point is in the class or not. This is inherently a much more difficult task than 2-class classification, for the simple reasons that there are no negative samples to learn from. Although harder to implement, this approach has the advantage that it is in principle scalable in the sense that it may lead to the creation of arrays of condition-specific classifiers.

After sufficient training of a set of one-class filters, one can bring a novel data vector and test it on these filters to see whether this patient tested positive for the "mild neglect" or "severe neglect" or "mild upper-left hemi-field but otherwise OK" etc. This may focus the clinicians in treating only the impaired faculties of the patients.

Since data is so expensive and scarce, we were also interested in the possibility of combining data from different VR platforms. Thus we can expand the data pool for each of the relevant pathological conditions (e.g., Neglect, TBI) and base our clinical diagnosis on a larger sample. For example, one can train a neural network (NN) (Fausett, 1994) on data from two platforms in two separate hospitals where neglect patients perform various tasks. Once trained, the NN would be able to assist clinicians in both hospitals and even from another clinic using yet another platform.

Of course, this is a complex problem because different systems have different ways of dealing with the data, even of different dimensions, and the users perform different tasks. Nonetheless, if they are all motion data, we thought it worthwhile to see how they can be merged.

All of the previously described classification categories will benefit from a large, combined, multi-platform VR data set.

1.2 Contribution of study

As virtual reality technology gains popularity in the field of rehabilitation, clinicians and scientists need an appropriate way to interpret the data acquired in VR therapy sessions. We believe that finding efficient ways to harness machine learning for analysis of human behavior has a significant potential to better understanding of brain injuries. These injuries manifest themselves in such a wide spectrum, so patients may suffer from inaccurate evaluation of their condition. Also, better analysis of movement patterns may greatly assist neuroscientists in their pursuit of better understanding of brain mechanisms such as perception, attention, motor planning and control. Specifically we can imagine using these tools in the following fields:

1. Differential diagnosis - detect which of the suspected conditions the patient suffers from.

2. Prognosis - build a virtual model of the individual which we would want to extract automatically from data based on his performance in the VR sessions. Then an individual rehabilitative protocol can be obtained by simulating the behavior of an avatar in the VR; and simply testing how the avatar improves under a large variety of protocols.
3. Individualized treatment – once a good one is established, it can be immediately applied to the patient who is being treated in the same VR environment.

1.2.1 Goals of study

In the present study we aim to demonstrate the abilities of machine learning tools to classify patients according to their pathology using captured motor data.

To do this our primary tool is 2-class classification. However because of scaling concerns (over the number of possible diagnoses) we utilized the 1-class approach.

We also want to see how the scale and availability of patient data can be increased in order to make the learning system as general as possible and not VR platform specific. This would also allow for cross platform classification.

Another goal is to see if one can classify patients of a certain disease according to the severity of their condition, using clustering techniques working on the same captured motor data.

In the following sections we shall demonstrate the feasibility of these approaches suggesting the relevance of machine learning tools.

2 Methods

We gathered data from 2 VR platforms. Data analysis was done using NN (in various arrangement, as described below), Support Vector Machines (SVM) (Cristianini and Shawe-Taylor, 2000) and k-means (Loyd, 1982), with the main tool being NN.

2.1 Data from 3D Experiment - VRROOM

Population: 29 volunteers participated in a study performed by Dvorkin et al., 2008. Ten of them were diagnosed as suffering from stroke without clinical signs for neglect, nine suffered stroke and showed signs for neglect. The other ten were healthy adults in similar ages. The patients were diagnosed as having different levels of severity of their medical condition, ranging from mild to severe.

Procedure: Participants were positioned in front of the VRROOM (Virtual Reality and Robotics Optical Operations Machine) system (Patton et al., 2006), shown in Figure 3. On each trial a virtual target appeared randomly in space in one of 49 possible positions. Participants were instructed to reach toward the target as soon as they detected a target appearing within the scene, using their unimpaired arm. Each subject was presented with 343 target stimuli altogether.

Analysis: The data vectors were first preprocessed in order to eliminate premature movement initiations or omissions (i.e., when the subject did not respond within three seconds). Also, any hand movement prior to the stimulus appearance was ignored as it is not part of the experiment.

The input vectors were of **several types**:

- Long vectors - including the data from the onset of the target stimulus in the virtual environment till the end of the hand movement.
- Movement vectors – consisting of data from the response of the subject, i.e., only from the moment the subject started a physical response.
- Initial/final vectors – These vectors included the initial/final 130 data points of the movement. As oppose to the former types, these vectors were of fixed lengths.

2.2 Data from 2D experiment - GestureXtreme

Population: 99 volunteers participated in a study performed by Rand et al., 2004. 54 were healthy adults, 11 adults who suffered from CVA (without neglect), 9 children suffering from TBI and 25 healthy children.

Procedure: A virtual supermarket was presented to the participants using the GestureXtreme platform (www.GestureTek.com). This system is based on video motion capture technology where user is captured by video camera and sees his image in immersive 2D VR environment on the screen (Figure 4). Motion tracking algorithm produces two-dimensional coordinates of the user's movements. The participants were instructed to touch certain virtual products according to a shopping list (Rand et al., 2004). The system did not have markers on the user's body, except for a hand glove so the motion tracking was very limited in its abilities. Thus the collected motion data was very noisy and fragmented.

Figure 4. A sample view of a subject within a GestureXtreme virtual Environment.

Analysis: In this experiment we implemented learning tools in a challenging virtual environment.

The data vectors were first preprocessed in order to find least noisy segments where the movements of the hand are consistent over a period of several seconds. At a rate of 15 frames per seconds, a typical segment consisted of several coherent chunks of 7-10 second durations. Thus each participant produced eventually about 750 measurements (x,y,t) of his hand.

The noisy and fragmented nature of the data prevented us from creating input vector of whole movements or even long snippets. A snippet is a consecutive sequence of differentials (dx,dy,dt). For the 2D experiment we used snippets of length five.

2.3 Combined Platforms

Population: the only identical categories in both experiments were healthy and non-neglect stroke. The combined population included 64 healthy adults, 20 stroke patients and 25 healthy children.

Analysis: The characteristics of each platform were studied by plotting the trajectories of select subjects' trials. An example can be seen in Figure 5 and Figure 6.

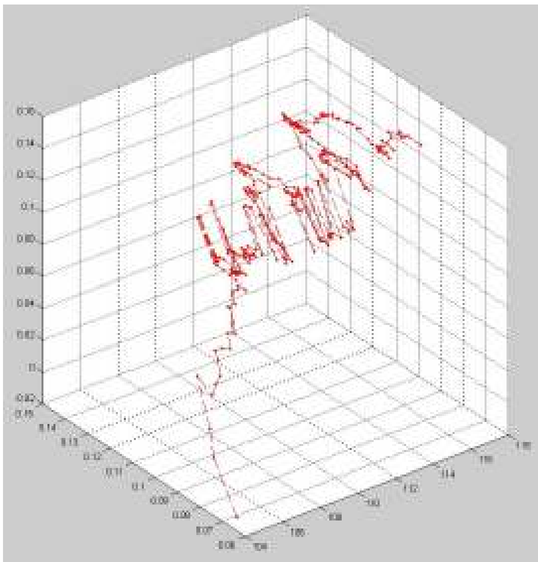


Figure 5. A sample trajectory of a healthy subject trial (x,y,t) on the GestureXtreme platform

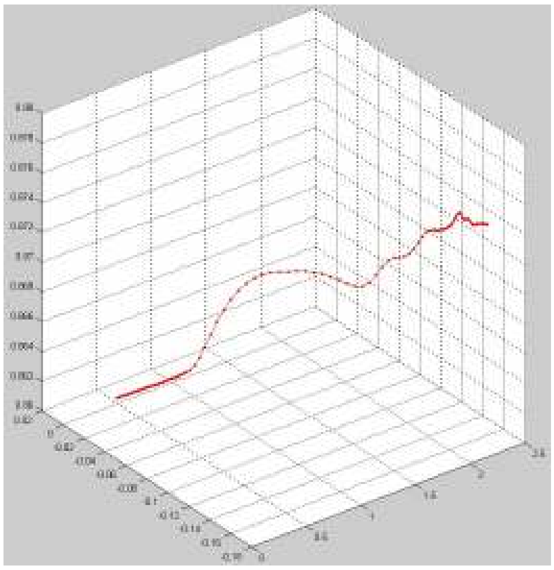


Figure 6. A sample 2D projected trajectory of a healthy subject trial (x,y,t) on the VRROOM platform

In order to experiment on data from both experiments, a geometric transformation was needed. The 3D data was transformed to 2D by removing the Z axis: (x,y,z) -> (x,y,0). This projection was chosen because the nature of the subject movement in the 3D experiment was such that there was relatively little movement in the z dimension. Beyond the obvious difference of dimensionality between the two platforms, the merging of the data raised another challenge for this task. As can be seen in Figure 5 and Figure 6 the GestureXtreme data are very noisy and unstable.

We used varying lengths of "snippets" from both data sets, seeking the optimal length that might remove the platform specific movement characterization and hoped that the intrinsic information will remain.

There was considerable more data in the 3D dataset, so we selected a representative subset of the 3D data, which took into account all of the different target locations.

2.4 Architecture and Training

2.4.1 Two-Class

GestureXtreme (2D): For this experiment we used a feed forward network architecture with one hidden layer, which received as input a 15 element vector – 5 consecutive hand movements vectors (dx,dy,dt). The hidden layer had 5 elements. All together an architecture of 15-5-1 (Figure 7). For the more difficult case (TBI v. CVA) a network of the structure 15-20-10-1 (2 hidden layers) was applied. Previous 2-class classification work on the GestureXtreme has been published (Feintuch, Manevitz, Mednikov et al., 2006). The results shown here confirm their results.

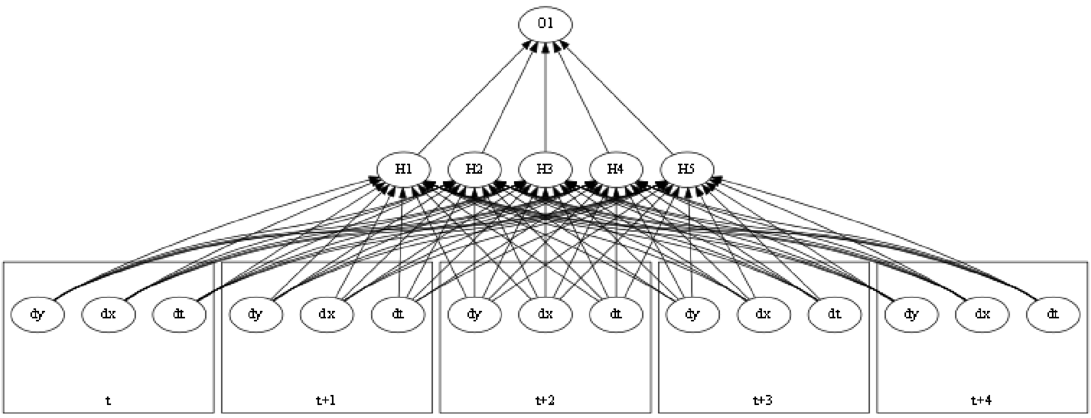


Figure 7. NN for the GestureXtreme experiment.

VRROOM (3D): Here we used the same feed forward network architecture with a different input layer, consisting of various lengths, consecutive hand movement vectors (x,y,z). 1400 elements for a long vector (1400-5-1), 1000 elements for a movement vector (1000-5-1), 130 elements for initial/final vectors (130-5-1).

Combined Platforms: After an extensive empirical search for the optimal "snippet" length, we settled on length of 90 data points (dx,dy,dt). We found that 2 hidden layers were more effective, and thus the final network structure was 270-300-10-1.

At first, the training method was Levenberg-Marquardt (Marquardt, 1963). Later on we discovered that the resilient back-propagation algorithm (Riedmiller and Braun, 1993), which modifies the update-values for each weight according to the behavior of the sequence of signs of the partial derivatives in each dimension of the weight-space, obtains the same stable results only with a much faster processing time.

Empirically, we found that it was sufficient to run training for 50 epochs for the simpler classification task of healthy v. CVA. It was necessary to increase the number of epochs from 50 to 300 for the more difficult task of CVA v. TBI classification.

Cross-validation: One subject was removed during the training session. It was used for testing of the generalization. This was repeated as many times as there were subjects and percent of successful classifications was calculated.

2.4.2 One-Class

This was run only on 3D data from VRROOM. For these experiments we used a back propagation neural network, structured such that the number of neurons in the hidden layer is smaller than the input layer and output layer, which have the same number of neurons. The "compression" rate (ratio between number of neurons in the input and hidden layers) was chosen to be 5 as is exemplified in Figure 8. The input layer was of various lengths, 1000 elements for a movement vector (1000-200-1000) and 130 elements for initial/final vectors (130-26-130).

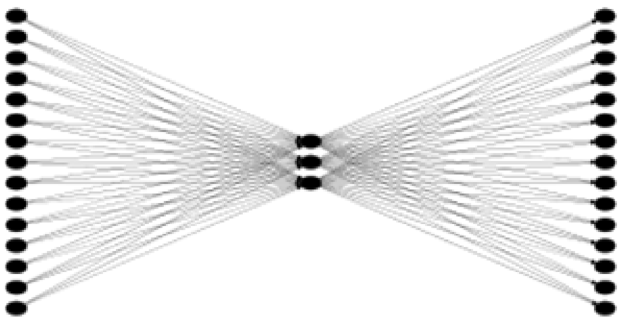


Figure 8. An example of a NN with a compression rate of 5

We used the resilient back propagation training method, with the number of epochs being 300. A simple Euclidean distance metric was used to measure how close the output was compared to the input. Subjects at a distance below a manually chosen threshold were inside the class, while subjects at a distance above the threshold were outside the class. For a discussion on automatic threshold selection methods see Manevitz and Yousef, 2007. An example for severe neglect is shown in Figure 9

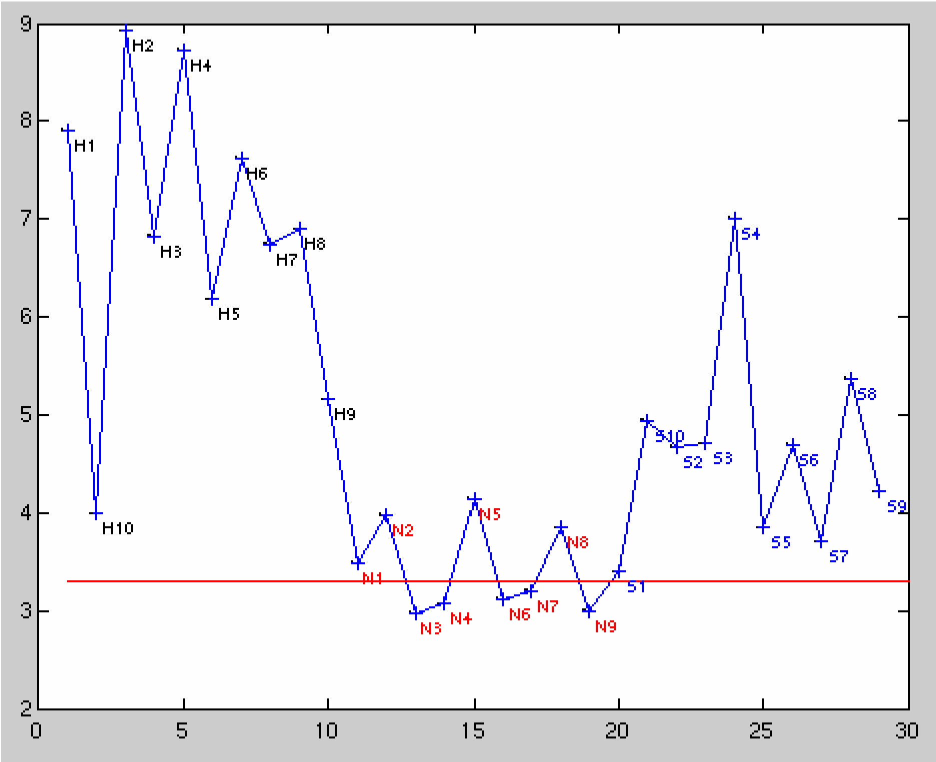


Figure 9. Severe neglect classifier experiment. The threshold was chosen to be 3.3

2.4.3 *Zero-Class (Clustering)*

For both of these experiments we used a Kohonen Self Organizational Map (SOM) network (Fausett, 1994).

VRROOM (3D): The input layer was of various lengths, 1000 elements for a movement vector and 130 elements for initial/final vectors.

We experimented on different lengths of line topologies (from 2 clusters up to 300).

Each data point was placed in a cluster numbered between 1 and the total number of clusters. We emphasize that the clustering is done on data points *without* consideration of which individual patient, the data point belongs to. In Figure 10 you can see, for example, the histogram; color separated for subject, for Kohonen Clustering to 7 neurons of all of the 1278 data points of Neglect patients (only some of the patients are shown for clarity). N9 appears to be on one extreme while N5 on the other. N4 seems closer to N9, than N1 is.

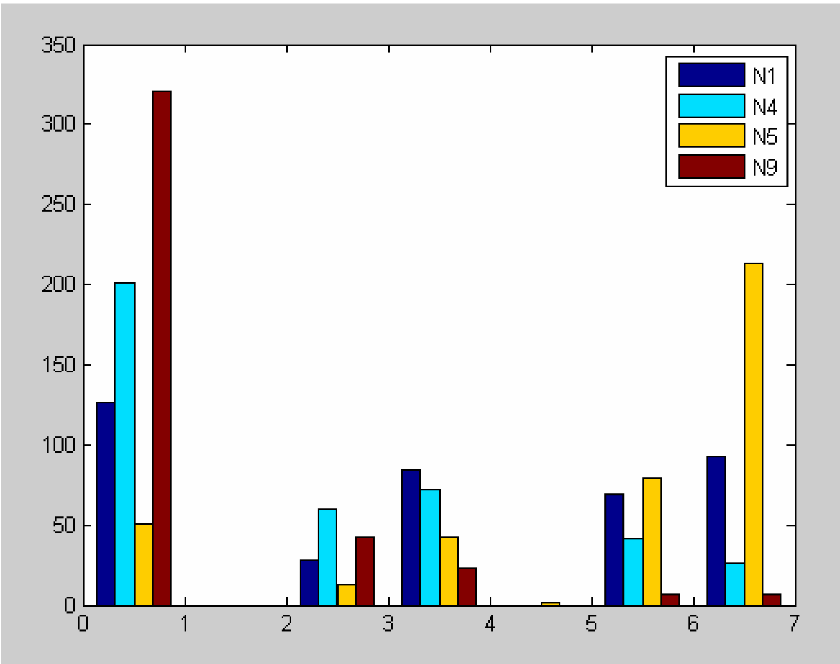
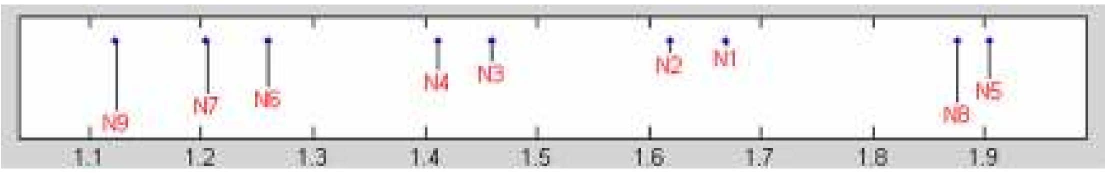


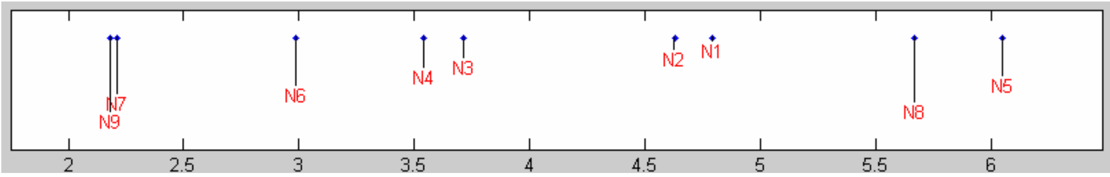
Figure 10. Histogram for Kohonen Clustering to 7 clusters of data points of 4 Neglect patients.

In order to have a simple way of showing the relationship between the different subjects, we place a subject on a linear scale in the correct place relative to the other subjects. This is possible by calculating the average of its data points' cluster numbers.

For example in Figure 11, the linear scale for Neglect subjects is presented.

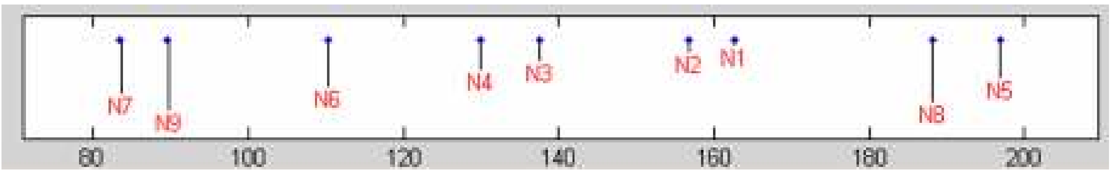


a – 2 neurons



b – 7 neurons

c – 50 neurons



d – 300 neurons

Figure 11. Clustering results for Neglect subjects. Kohonen line topology of various amount of neurons: a - 2 neurons, b – 7 neurons, c - 50, d - 300

Demonstrated in this example, when using the kohonen line topology, the order of the patients was strikingly preserved, no matter the amount of clusters. Except in the case of patients N7 and N9, all the rest of the group ordering and spacing were preserved. The N7-N9 switch was local. They kept their topological distance from N6 and the others.

The topology we have chosen as a good representation was that of a line with 7 clusters. For a movement vector the architecture was 1000-7 and for initial/final vectors it was 130-7. Training was 50 epochs.

GestureXtreme (2D): In this experiment we used the same linear scale technique and network topology as in the 3D experiment.

For length 5 snippets the architecture was 15-7 (see Figure 12). Training was 50 epochs as well.

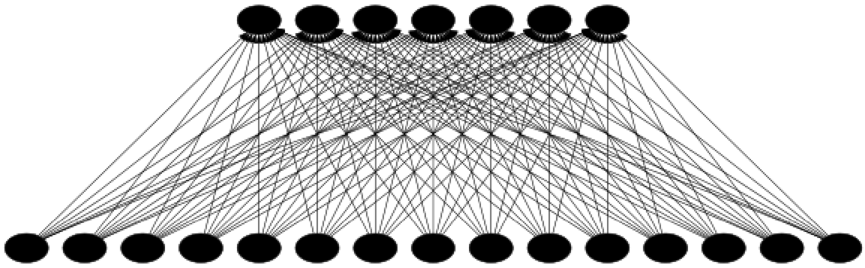


Figure 12. Kohonen SOM clustering algorithm NN architecture length 5 snippet vectors (15 neurons) for the chosen 7 clusters line topology

2.5 Data point result analysis

All experiments' training and testing were conducted on each of the subjects' data points (whether long vectors or short vectors).

For substantive clinically beneficial results, we transformed the results to be subject specific instead of subject's data point specific. Towards this end the data points result were averaged by each subject and then put to the specific metric of the classification kind (threshold for 1-class and 2-class, and putting on a liner scale 0-class). In some of the 2-class results we also show data point rates for comparison.

These rates are calculated using a simple plurality rule on all subjects' data point results combined together without consideration of any specific subject.

1 Results

1.1 Terminology

When describing the results in text and tables there are four main populations whose subjects are referred to by a combination of letter and number:

- Healthy participants are denoted as H. In the 2D experiment HA represent healthy adults and HC represents healthy children.
- Stroke (a.k.a. CVA) who were not diagnosed as suffering from neglect are denoted as S.
- Stroke patients who are also suffering from neglect are denoted as N.
- People with traumatic brain injury (TBI) are denoted as T.

1.2 Two-Class

1.2.1 3D experiment

As seen in Table 1 the generalization success rates in classification of long vectors were 72-89%. It is not surprising to see that the best rate was achieved for the Healthy/Neglect classification, for neglect is a condition which tends to be explicitly manifested. From a clinical point of view the distinction between neglect and CVA is (72%) is certainly more meaningful, since traditional assessments often lead to a misdiagnosis of less severe cases of neglect.

As explained earlier, long vectors include all data from the onset of the target stimulus till the end of the hand movement. This includes the target detection as well as both movement planning and execution. Thus the distinction between different populations may be the result of a cognitive perceptual component, (i.e., reflecting the target detection latency of response phases), or a motor component. Such evidence has of course a scientific merit but it does not require a neural network to measure response time.

While there is ample evidence for a perceptual deficit associated with neglect, motor control studies have produced a large amount of contradictory data.
Hence we also attempted to perform a 2-class classification using movement vectors. In this case the input included only data from movement initiation to its completion.

Furthermore, as neglect, almost by definition, manifests itself in one half of the visual field, we chose to use only the relevant hemi-field in the input data.

The classification results resemble very much the ones produced with the long vectors, ranging from 72% to 89%. This implies that the distinction between the populations manifests itself in more complicated ways than reaction time.
In order to further investigate the differences between these populations, we used another length of input. This was done by preparing a vector consisting of either the initial or the final movement segment (length of 130 data points). This approach may assist in focusing the research to the critical point of the hand trajectory, where the difference may lie.

Table 1. Success rates of 2-class classification in 3D data.

Vector size	Populations	BP NN Success Rate By Patient	BP NN Success Rate By Data Point
Long	Healthy/CVA	78%	62% (4010/6458)
Long	Healthy/Neglect	89%	79% (4557/5791)
Long	Neglect/CVA	72%	63% (3584/5703)
Movement	Healthy/CVA	78%	68% (4378/6458)
Movement	Healthy/Neglect	89%	78% (4515/5791)
Movement	Neglect/CVA	72%	66% (3737/5703)
Initial segment	Healthy/CVA	39%	43% (2791/6458)
Initial segment	Healthy/Neglect	83%	70% (4058/5791)
Initial segment	Neglect/CVA	78%	61% (3470/5703)
Final segment	Healthy/CVA	67%	53% (3435/6458)
Final segment	Healthy/Neglect	89%	61% (3547/5791)
Final segment	Neglect/CVA	61%	52% (2962/5703)

While significant, the classification results are not as decisive, ranging from 39% to 83%. However, when comparing the success level of the classifications, it seems that it was easier for the NN to classify healthy from CVA or from neglect in the final segment, compared to the initial segment (67% vs. 39% and 89% vs. 83% respectively). On the other hand, the more challenging classification, the one between CVA and neglect patients seems to be more distinct in the initial segment (78%) rather than the final segment (61%). Such results imply that the deficit caused by neglect is more likely to be expressed at earlier stages of motion execution. The Healthy vs. CVA by-patient result for initial segment (39%) is lower than the data points result (43%), which is closer to 50%. It's very likely that the patient percentage will be lower than data point percentage when the success rate is around 50% (similarly, it is likely that patient percentage will be higher than data point percentage

when success rate are much higher). In addition, further experimentation showed that longer and longer inputs of initial segment increased the success rate linearly up to the maximal result of complete movement (78%).
It should be noted that the key findings of this analysis were also reproduced using SVM with a linear kernel and the results were comparable.

3.2.2 2D experiment

Compared to the previous VR system, this platform produced very noisy data, thus the preprocessing reduced the inputs to short vectors snippets, each one covering 5 consecutive position measurements.

The results, appearing in Table 2, indicate that the NN had high success (72%-95%) rates in comparing the three populations who participated in this experiment, namely healthy, CVA and traumatic brain injury. It is interesting to mention that in the CVA group there was a patient who was consistently misclassified as healthy even when his data were the training phase. Reviewing closely his medical files revealed that this particular patient indeed suffered from CVA but he had only cognitive impairments but no physical disability. This anecdote, beyond demonstrating the clinical potential of the system, suggests that this particular NN classified according to movement features of the subjects' motor behavior rather than by cognitive features. It is not surprising as in this experiment the input consisted of rather short movement segments, which apparently are not sufficient for containing meaningful cognitive information.

Table 2. Success rates of 2-class classification in 2D data.

Input Vector Length	Populations	BP NN Success Rate By Patient	BP NN Success Rate By Data Point
Length 5 snippets	Healthy/CVA	85%	90% (11500/12670)
Length 5 snippets	Healthy/TBI	95%	95% (11167/11806)
Length 5 snippets	TBI/CVA	72%	59% (1414/2393)
Length 5 snippets	Healthy Adults/ Healthy Children	50%	80% (10641/13351)

Since the TBI patients were all much younger than the CVA patients, it was possible that it is the age difference that accounts for the classification between these populations, rather than the clinical condition. To control for this variance we tried to classify the healthy children from the healthy adults. As seen in Table 2, the classification failed (50%), so it appears that age did not play a role in the CVA/TBI classification.

1.2.2 Combined Platforms

The only subject categories suitable here are healthy and non-neglect stroke, since these conditions were the only ones mutual to both experiments.

The results, appearing in Table 3, indicate that for combined platform training, the NN had high success rates only for relatively long "snippets" – 90 data points (270 neurons). Attempts to have a shorter snippet more resembling the one used for the GestureXtreme-only experiments (5 data points) gave unsatisfactory results. In addition network architectures using 1 hidden layer (regardless of its size) also was not able to successfully classify the different populations. It might be possible that the network is actually learning two separate networks within itself for the different platforms, but we think the chance for that is relatively small due to the tightness of the architecture. Even though the nature of the movement is quite different between the platforms, as can be seen from Figure 5 and Figure 6, the results with the longer vectors are quite encouraging.

Table 3 Success rates of 2-class classification in combined platform

Vector size	Training Set Origin	BP NN Average Success	Notes
30 data points	VRROOM and GestureXtreme	75%	Failed with VRROOM healthy
90 data points	VRROOM and GestureXtreme	90%	
90 data points	VRROOM only	50%	Failed with all GestureXtreme stroke
90 data points	GestureXtreme only	50%	Failed with all VRROOM stroke

On the other hand we were unable to succeed at Cross Platform (training on one platform and testing on the other). When training on VRROOM data only or on VIVID data only, the NN could classify the other platform population only at chance levels. It is possible that this might be rectifiable by a more careful matching of the training and testing environment, but we have no results in this direction.

One-Class

1.2.3 3D experiment

The success rates in 1-class classification are not as good as in 2-class classification. This is to be expected as an inherent characteristic of 1-class compared with 2-class networks. Table 4 states the most interesting results. It is not surprising to see that the best classifier is achieved for the healthy population. We hypothesize that this is because the variance between healthy individuals is relatively small. The same could not be said for the CVA and Neglect populations, for which the individual behavior seems to vary more substantially from the milder to the more severe cases.

In order to further investigate this we chose to also train on input data only from trials where the target for reaching was in the neglected hemi-field. This caused the results to be significantly improved, especially for CVA and Neglect, which may point to the more stark difference in movement of Neglect patients and CVA patients when the former have difficulty perceiving the left hemi-field targets.

Table 4 Success rates of 1-class classification in 3D data

Vector size	Targets Included	Training Set Population	Compression NN Average Success	Notes
Movement	All	Healthy	93%	
Movement	All	CVA	69%	Failed on most Healthy
Movement	All	Neglect	50%	
Movement	All	Non-Mild Neglect	76%	
Movement	All	Severe Neglect	100%	
Movement	Left Only	Healthy	93%	
Movement	Left Only	CVA	83%	
Movement	Left Only	Neglect	62%	Failed on most CVA
Movement	Left Only	Non-Mild Neglect	83%	
Movement	Left Only	Severe Neglect	100%	
Movement	Right Only	Neglect	50%	
Movement	Right Only	Non-Mild Neglect	50%	
Movement	Right Only	Severe Neglect	93%	
Initial segment	All	Healthy	83%	
Initial segment	All	CVA	62%	Failed on most Healthy
Initial segment	All	Neglect	50%	
Initial segment	All	Severe Neglect	97%	
Final segment	All	Healthy	62%	Failed on most

				CVA
Final segment	All	CVA	62%	Failed on most Healthy
Final segment	All	Neglect	59%	Failed on most CVA
Final segment	All	Severe Neglect	66%	Failed on most Healthy

By experimenting on smaller and smaller sub-sets of Neglect (Full, Non-Mild Severe Only) a pattern emerged. The more homogenous the group of Neglect subjects and the more severe the group was, the better the separation from the other groups.

It is worthy to note that by removing just two extremely mild neglect subject "outliers" ("Non-Mild" Neglect) the performance improved considerably (from 50% to 76% for all targets and from 62% to 83% for left target only). Figure 13 and Figure 14 show the difference between using neglect and non-mild neglect for 1-class classification running on left target trials only. The margin between the neglect group and the CVA group increases substantially in the latter case.

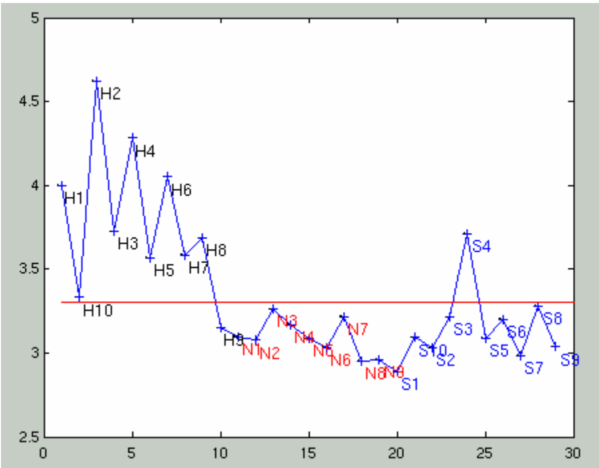


Figure 13. Neglect classifier for "Left targets trials only" experiment. The threshold was chosen to be 3.3

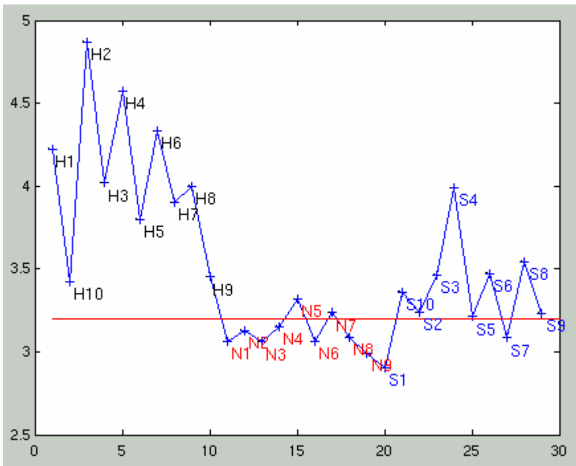


Figure 14. Non-Mild Neglect classifier for "Left targets trials only" experiment. The threshold was chosen to be 3.2

Interestingly when using shorter input vectors to represent the initial and final segments of movement, the Healthy classifier results were less successful especially in the latter case. Possibly the initial movement is more robust and uniform in this population. But the minute corrections needed to reach and stay at the target may be more varied between different healthy subjects.

Following these results, we decided not to pursue 1-class experiments on the GestureXtreme platform, due to the nature of the data. The 2D data were in short snippets, which, as we conclude from the 3D results, do not produce meaningful outcomes. This limitation, of course, prevented us from conducting combined-platform experiments as well.

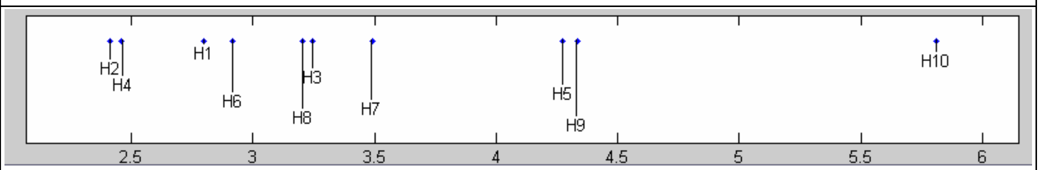
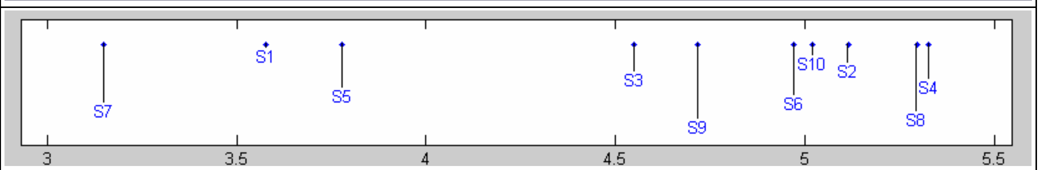
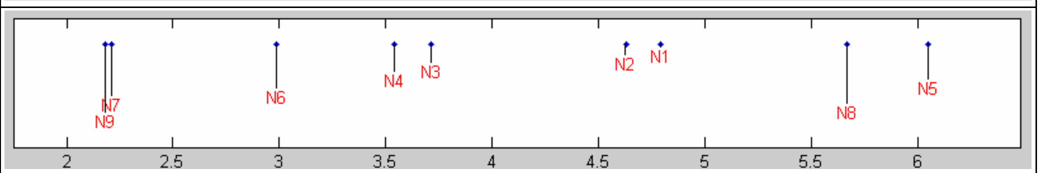
1.3 Zero-class (Clustering)

3.3.1 3D experiment

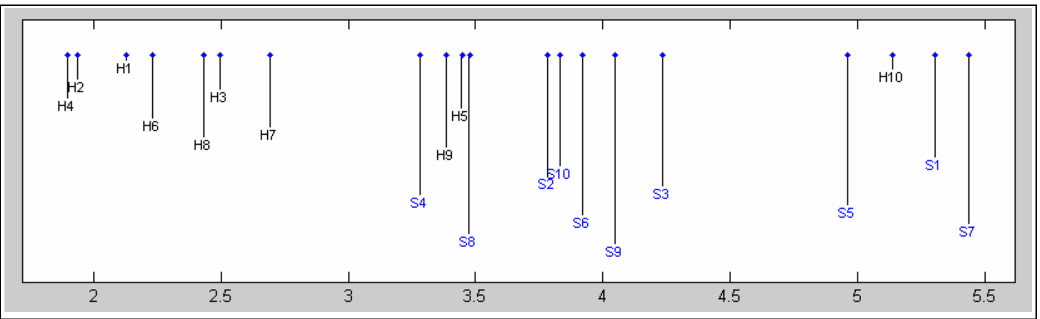
Stroke causes a wide array of damages leading to many types of medical conditions. Some of these sub-categories have received a distinctive title, such as neglect. Yet the definitions are rather broad, and the cut-off points are not completely well defined. In this phase of the study we picked various subsets of the patients and put them on a linear scale using clustering tools as described in the methods section. Following this, the patients' medical records were examined in order to test the clinical validity of their placements, and whether they point to meaningful directions. We chose various population types, and employed the Kohonen algorithm to cluster them in a line of 7 clusters. The main results appear in Figure 15, followed by the central findings.

First we were interested in finding out how homogenous our healthy control group is by itself (Figure 15, Panel A). Most of the healthy subjects were placed close together. One subject however, referred to as H10, was placed all by his own, for unclear reasons.

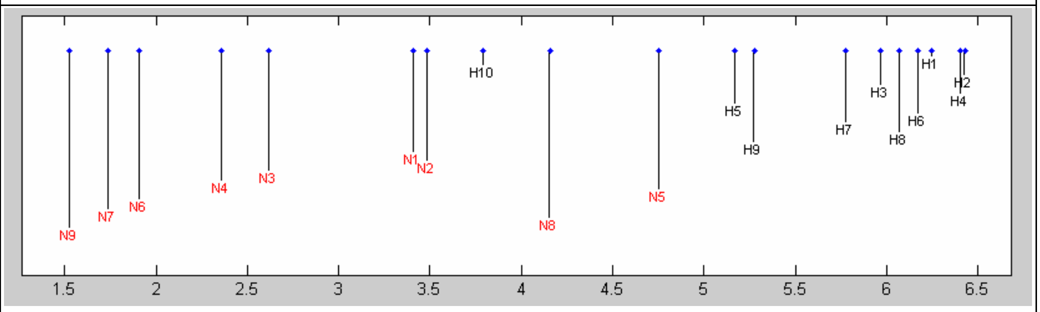
Figure 15. Clusters produced for 3D data.

Vector size, Populations, (Num. of clusters allowed)	Kohonen Scale
A) Movement, Healthy, (7)	
B) Movement, CVA, (7)	
C) Movement, Neglect, (7)	

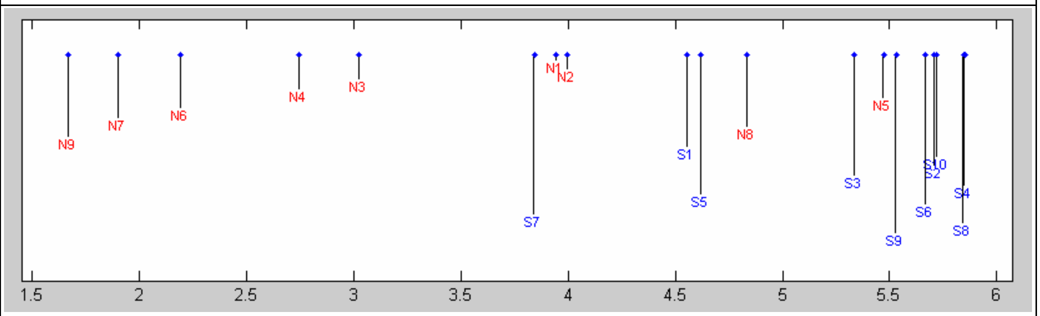
D) Movement, Healthy/CVA, (7)



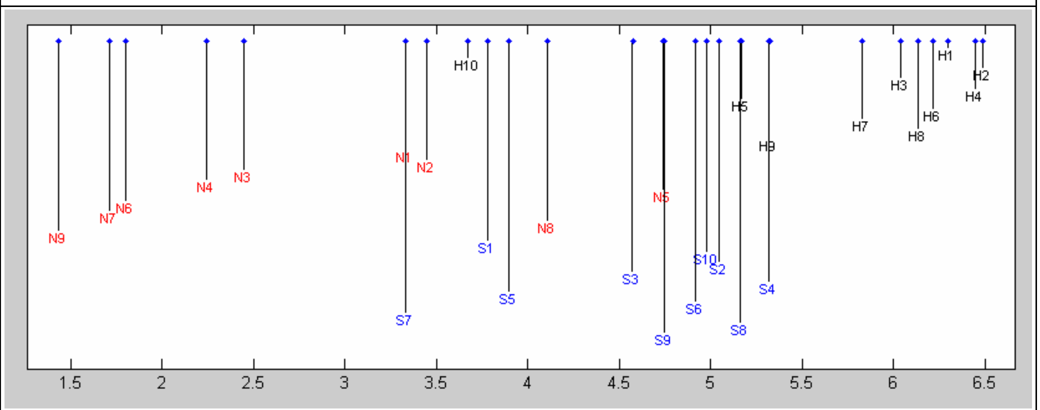
E) Movement, Healthy/ Neglect, (7)



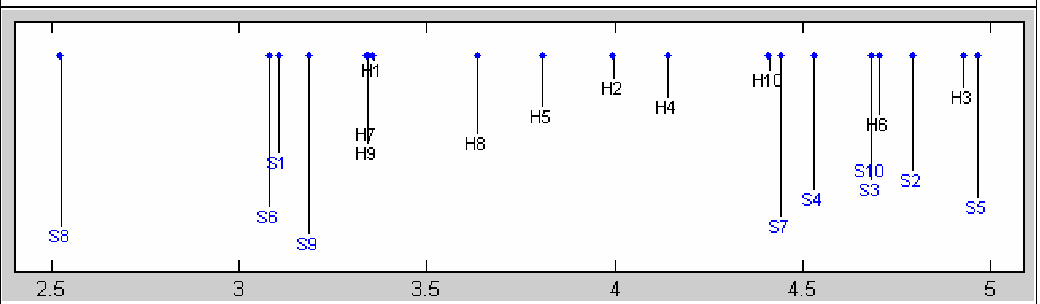
F) Movement, Neglect/CVA, (7)



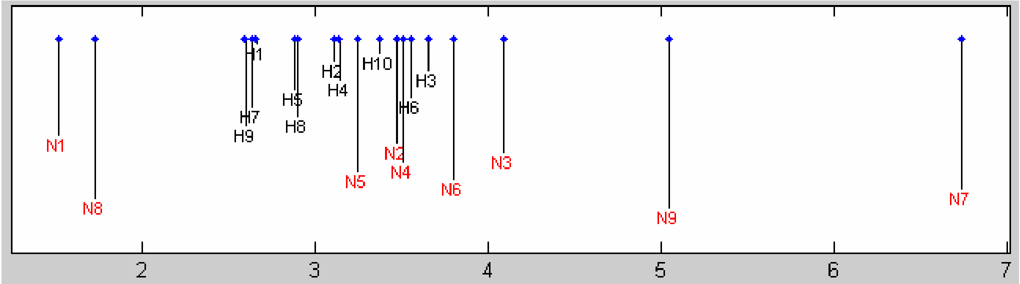
G) Movement, All, (7)



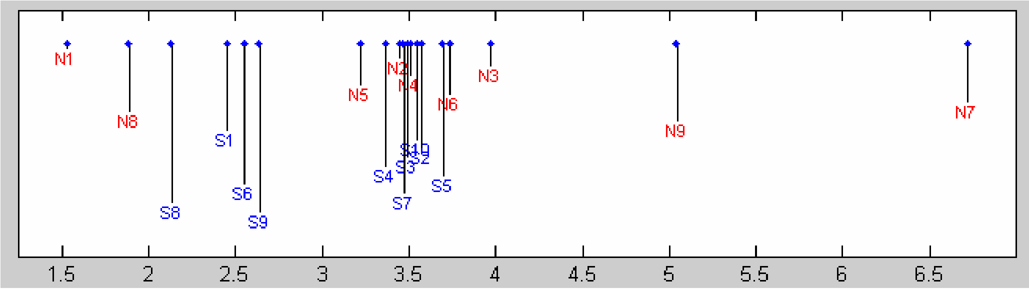
H) Initial segment, Healthy/CVA, (7)



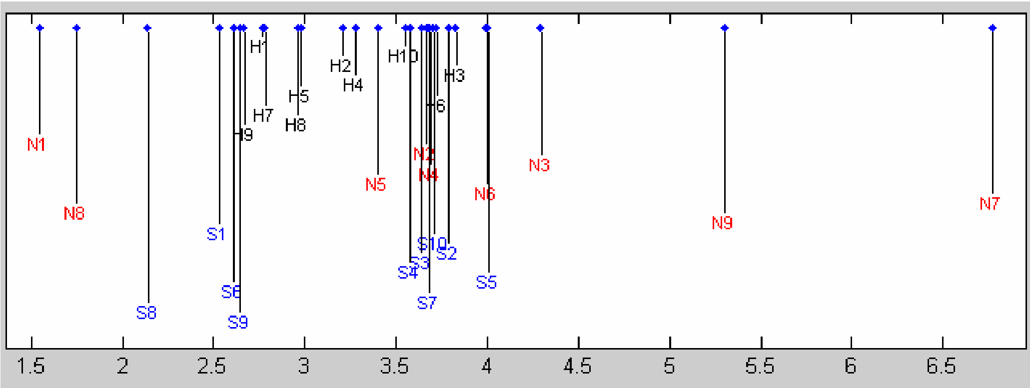
I) Initial segment, Healthy/ Neglect, (7)



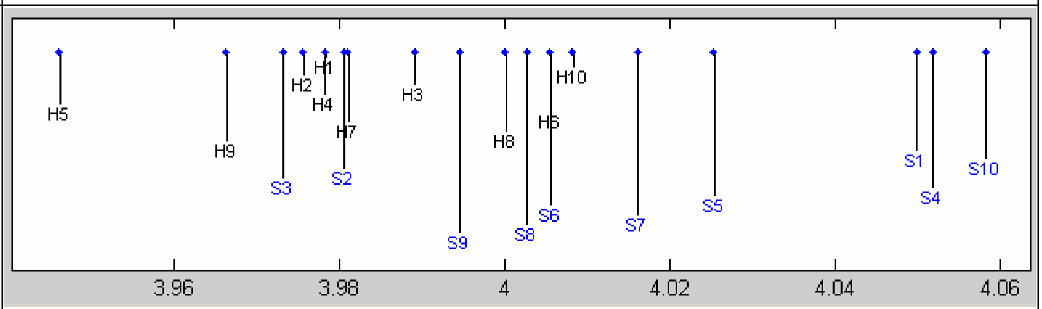
J) Initial segment, Neglect/CVA, (7)



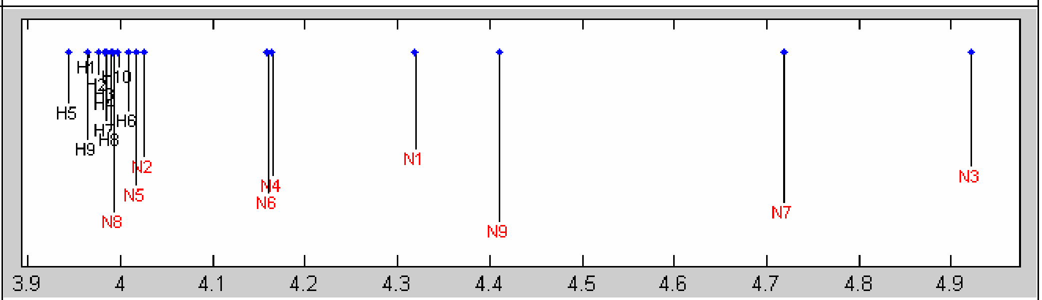
K) Initial segment, All, (2)

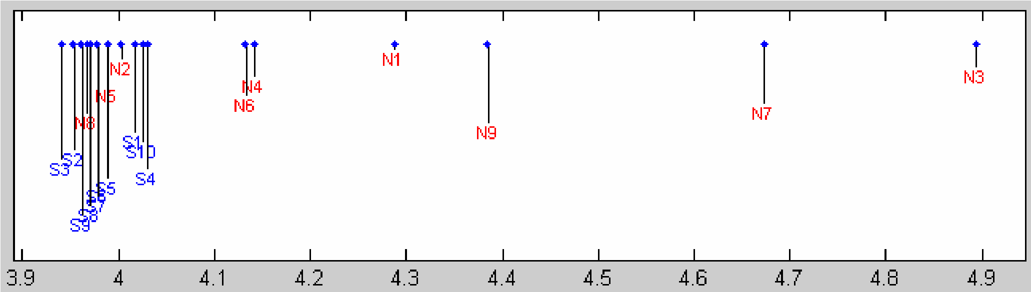
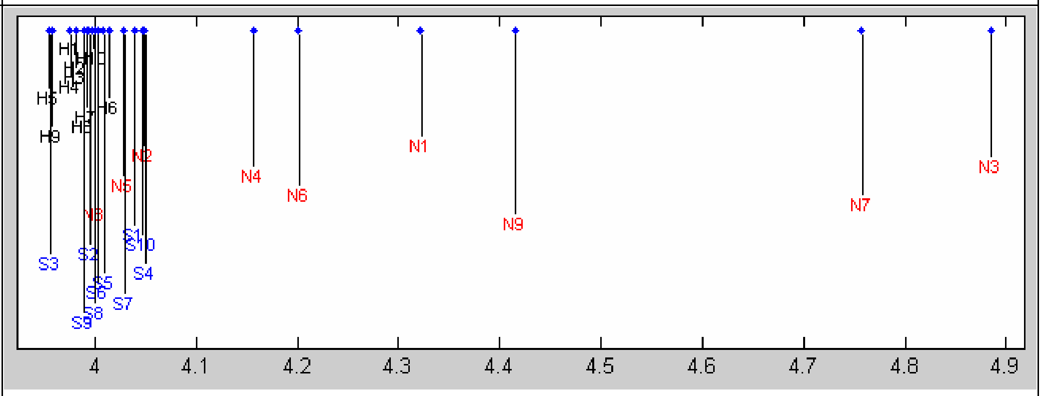


L) Final segment, Healthy/CVA, (2)



M) Final segment, Healthy/ Neglect, (2)



N) Final segment, Neglect/CVA, (2)	
O) Final segment, All ,(2)	

The CVA population by itself appeared to be divided into two distinct groups (B). The neglect patients, however, were placed into several groupings (C). Reviewing their medical records revealed that all the patients closer to cluster #7, towards the right end of the scale were diagnosed with only mild neglect, while the left end patients were more severe cases. Thus the neglect patients were successfully placed on a severity scale.

When comparing pairs of populations, some interesting clusters have emerged. When healthy and CVA subjects were pooled together (D), all the CVA subjects were close to each other with three healthy patients who were closer to them than to the other healthy patients. This suggests that the border between healthy and stroke is not always clear cut. The healthy and neglect populations (E), were placed apart, where the healthy were separate (except for H10), and the neglect were divided again into two groups, severe and mild.

When pooling together the two patient populations, CVA and neglect (F), all of the mild neglect patients, N1, N2, N5 and N8, performed well enough to be "upgraded" to the CVA side. This suggests that the distance between neglect and non-neglect stroke patients is positively correlated to the severity of the neglect symptoms. A similar trend was observed when we clustered all the subjects (G), as the severe neglect patients were in one side, some of the healthy were in the other, and a middle cluster included all the CVA, the mild neglect and even three healthy subjects.

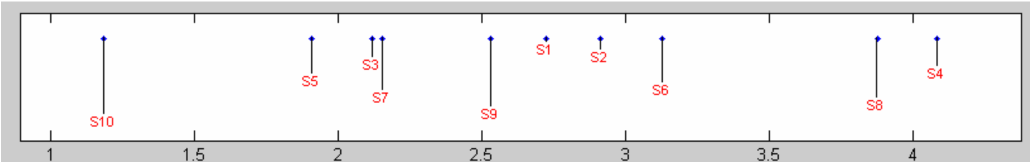
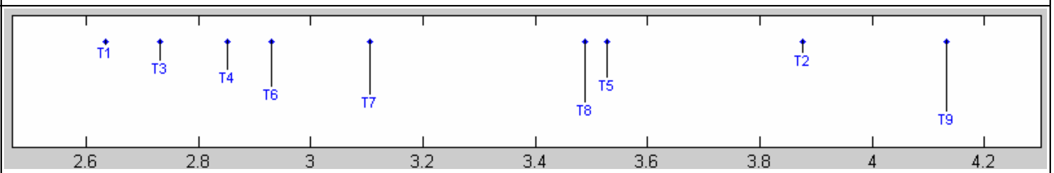
As before, we also focused at the initial and final segments of the motion. The initial segment (H-K) produced a similar pattern, although much less distinct. For example, when clustering the CVA and neglect populations, some of the severe neglect subjects (N4 and N6) were closer to stroke subjects than to other severe neglect subjects. The clustering of the final segment (L-O) produced essentially one group for all healthy and stroke subjects and put severe neglect farther away. It seems that the more the data are fragmented, the harder it is for the system to differentiate between the different severities.

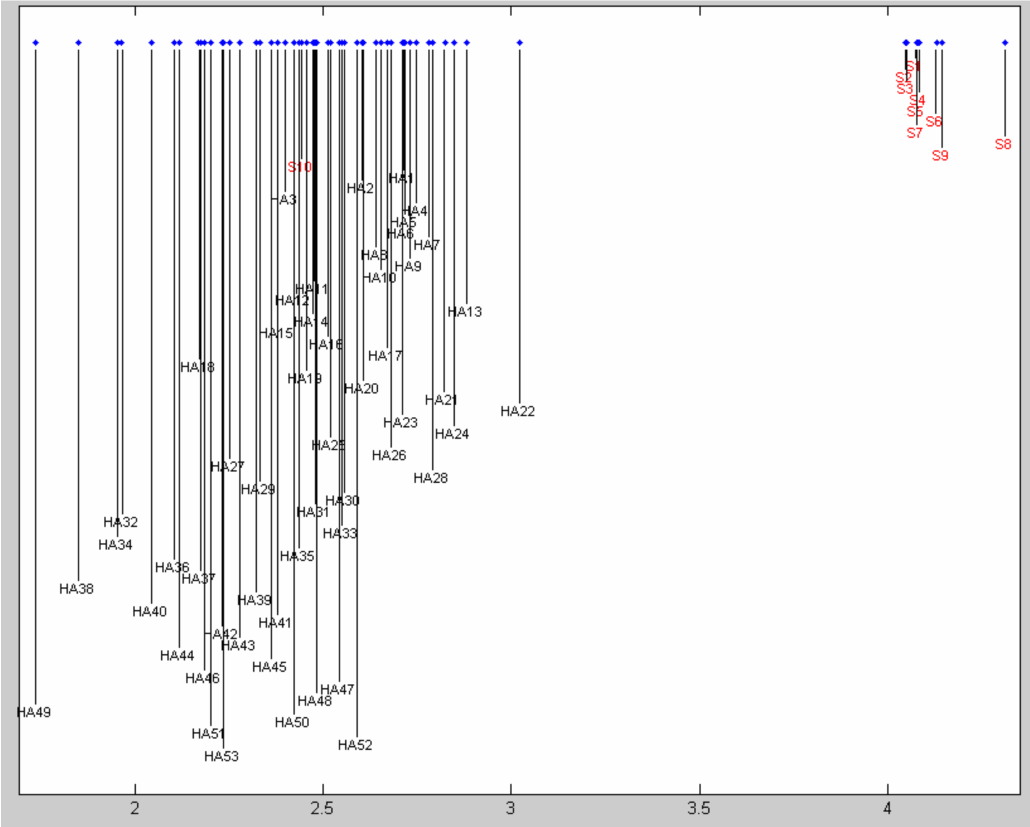
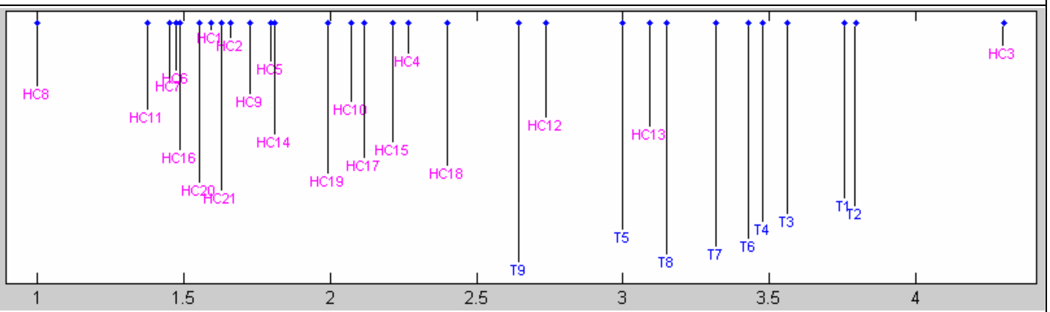
The key findings were reproduced also when employing k-means. The clusters k-means generated showed similar trends to kohonen. K-means does not take topology (the neighboring clusters) into consideration, so inherently this method is more limited in its ability to help put the various population on a single scale.

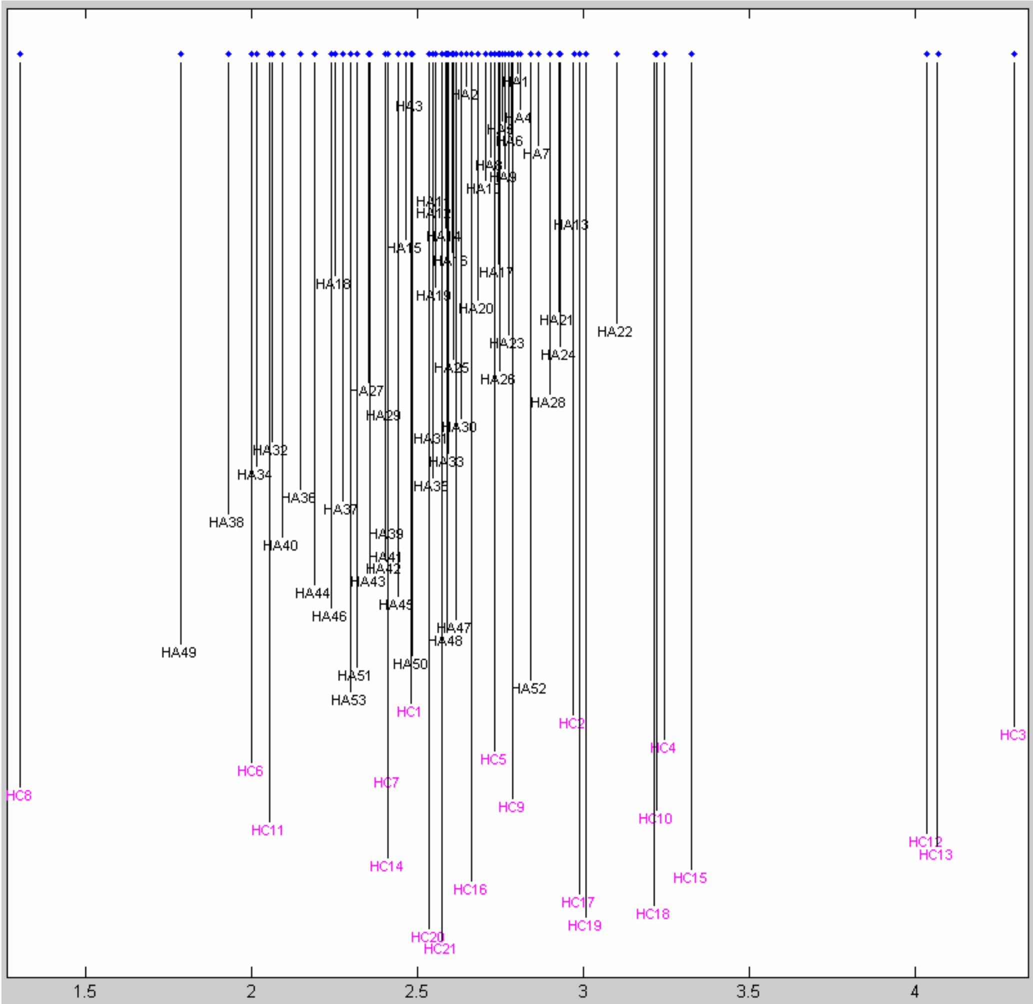
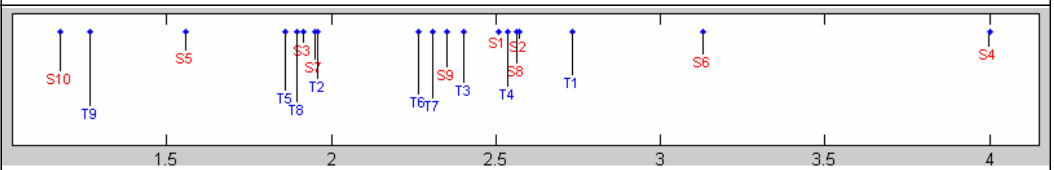
1.3.1 2D experiment

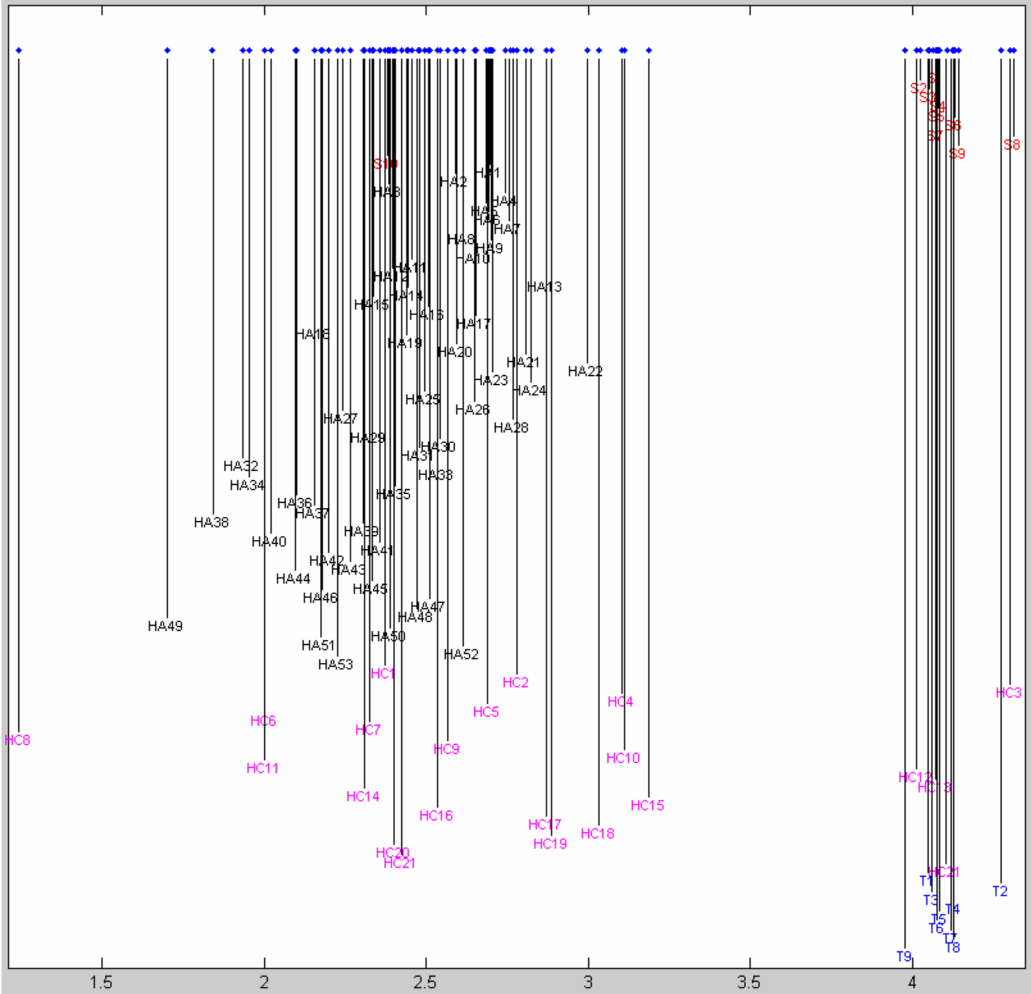
Looking at Kohonen clustering for the 2D data (SeeFigure 16), the clustering of Healthy and CVA adhered to the medical condition (Figure 16, Panel C). The two populations were far apart on the linear scale, aside from S10, who, as mentioned earlier, suffered no motor disability. When clustering Healthy children and TBI children (D), the two groups were mostly separate from each other. Healthy children were placed together with healthy adults (E). When clustering the two patient populations the TBI population seems to be more homogenous and clustered close to each other (F). However due to the heterogeneous patients' nature of the CVA population one cannot detect two separate clusters for stroke and TBI. When pooling all of the classes together (G), we essentially replicate the previous findings.

Figure 16. Clusters produced for 2D data.

Input vector length, Populations, (Num. of clusters allowed)	Kohonen Clusters
A) Length 5 snippet, CVA, (7)	
B) Length 5 snippet, TBI, (7)	

C) Length 5 snippet, Healthy Adults/ CVA (7)	
D) Length 5 snippet, Healthy Children/ TBI children, (7)	

E) Length 5 snippet, Healthy Children/ Healthy Adults, (7)	
F) Length 5 snippet, TBI/CVA, (7)	

G) Length 5 snippet, All, (7)	

2 Discussion and future directions

2.1 Discussion

In this study we demonstrated how machine learning tools may assist the clinician or scientist in analyzing data collected by VR platforms. This can be done even though these data are based on very small samples and even when the data are extremely noisy and partial. We proposed several approaches for achieving meaningful results.

First, two-class classification may assist in differential diagnosis. This was demonstrated as in both experiments, as several patient populations were differentiated one from another notably above chance level: CVA vs. neglect (3D Experiment) and CVA vs. TBI (2D Experiment). In this study, we picked medical conditions where we could assess the patients also in conventional methods. We believe that our approach may aid also in more hard to distinguish conditions such as spatial neglect and hemianopsia, which are related to different brain mechanisms, but lead to a similar behavior.

Furthermore, from the scientific aspect, running such classifications can be done while using different segments of the data as input. The results may direct the researcher to the key components in motion or behavior, which are sensitive to the classification. For example, the data in the 3D experiment suggest that perhaps the difference in reaching behavior between neglect patients and non neglect CVA patients lies at the very beginning of the motion, where the classification between them is quite high (78%). Such pointers may aid researchers in hypothesizing models of brain functions and in designing the experiments to validate them.

The zero-class approach in the 3D study, suggested how the rigid distinction between various conditions may be misleading. By having the mild neglect subjects N1,N2,N5,N8 close to non-neglect CVA patients on the linear scale, it was shown how some neglect patients behave in a similar way to non-neglect CVA patients or how certain CVA patients have mild neglect-like behavior. This approach can point the rehabilitation professionals to better understanding and organization of heterogeneous or wide spectrum disorders. This is true for many other broadly defined pathologies. For example it would be of significant value if zero-class clustering may aid in separating the wide spectrum of attention deficit disorders (ADD) into meaningful sub-categories.

Under the 2D environment the results were not as conclusive. We cannot discern if it is due to the intrinsic characteristics of these diseases (CVA and TBI) or due to the high level of noise the GestureXtreme platform incurs.

Another way to better define and profile complicated medical conditions may be achieved by implementing one-class classifications. Our results indicate that one can create a successful one-class classifier for homogenous groups even when training on small-size samples. For example, when severe Neglect and CVA patients are presented with stimuli on the left side, statistically significant 1-class filters are produced for either class. Also a healthy classifier has a very high success rate (93%) of differentiation.

The more homogenous the group of Neglect subjects and the more severe the group was, the better the separation from the other groups. (Compare for example Neglect sub-class results when presented with stimuli on the left side: all Neglect – 62%, Non-Mild Negelct – 83%, Severe Neglect – 100%).

After sufficient training of a set of one-class filters, one can bring a novel data vector and test it on these filters to see whether this patient tested positive for the "mild neglect" or "severe neglect" or "mild upper-left hemi-field but otherwise OK" etc. This may focus the clinicians in treating only the impaired faculties of the patients.

For 2-class combined platforms (Healthy vs. Stroke), even though the nature of the movement is quite different between the GestureXtreme and Vivid platforms, the results are quite encouraging. This may lead to expanded data pools for each of the relevant pathological conditions.

On the other hand we were unable to succeed at 2-class cross platform (training on one platform and testing on the other). We conjecture that the two platforms may be too different even for short snippets. It is possible that having two more closely related platforms (in terms of noiseless signals and kind of motor activity) might result in improved results.

2.2 Future directions

The work reported on here, seems to show that brain damage can be both classified, and clustered using various machine learning techniques. However, we have inconclusive results on the use of one-class techniques; that is, it appears to be reliable in some cases, but not in others. Preliminary work studying this was done, but future work should try to make it clear in which cases the method can be relied on.

We also point out that the "Kohonen-style" topological clustering has the advantage in that one can try to find topologies that help visualize the structure of the disease. We showed a simple example of this in a linear structure on the neglect scale. More complex topologies are probably appropriate for other conditions.

The work on merging data across platforms has great potential. This work needs to be extended over many data sets from a variety of platforms. Success in this area would help researchers and applications as then data can be leveraged from many sources.

Finally, in our vision for the long range, we see the possibility of "closing the loop" and using the classification and clustering methodology as keys for making rehabilitation protocols both adaptive and individualized. This is especially tempting in the context of rehabilitation in the virtual reality environment. What is needed is the development of a virtual model of the individual which we would want to extract automatically from data based on his performance in the VR sessions. Then an individual rehabilitative protocol can be obtained by simulating the behavior of an avatar in the VR; and simply testing how the avatar improves under a large variety of protocols. Once a good one is established, it can be immediately applied to the patient who is being treated in the same VR environment.

References

Cristianini, N., Shawe-Taylor, J. An introduction to support vector machines : and other kernel-based learning methods. Cambridge University Press (2000).

Deouell LY, Sacher Y, Soroker N. Assessment of spatial attention after brain damage with a dynamic reaction time test. *J Intern Neuropsychol Society 2005*; 11: 697-707.

Dvorkin AY, Bogey RA, Harvey RL & Patton JL (2011) Mapping the neglected space: Gradients of detection revealed by virtual reality. *Neurorehabilitation and Neural Repair. Published online* 11 July 2011. DOI: 10.1177/1545968311410068

Dvorkin AY, Rymer WZ, Harvey RL, Bogey RA, Patton JL (2008) Assessment and monitoring of recovery of spatial neglect within a virtual environment. *In: IEEE Virtual Rehabilitation.* p. 88-92, Vancouver, Canada.

Fausett, L.(1994) Fundamentals of neural networks : architectures, algorithms, and applications. Englewood Cliffs, N.J.: Prentice-Hall.

Feintuch, U., Manevitz, L., Mednikov, E., Rand, D, Dvorkin, A., Kizony R., Shahar, M. & Weiss, P. L. (2006). Integrating Artificial Intelligence and Virtual Reality in the diagnostic process – Feasibility study. In: B. K. Wiederhold, S. Bouchard, & G. Riva (Eds). *Annual Review of CyberTherapy and Telemedicine*. San Diego, CA. pp 207-208.

Katz, N., Ring H., Naveh, Y., Kizony, R., Feintuch, U. and Weiss, P.L. (2005). Interactive virtual environment training for safe street crossing of right hemisphere stroke patients with Unilateral Spatial Neglect. *Disability and Rehabilitation*, 29(2), 177-181.

Lloyd, S. P. “Least squares quantization in PCM,” *IEEE Transactions on Information Theory*, vol. IT-28, pp. 129–1373, Mar. 1982.

Manevitz, L. and Yousef, M., One-Class Document Classification via Neural Networks. *Neurocomputing,* vol. 70, pp. 1466-1481, 2007.

Marquardt Donald (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. SIAM Journal on Applied Mathematics 11 (2): 431–441.

Patton, J. L., Dawe, G., Scharver, C., Mussa-Ivaldi, F.A., and Kenyon, R. (2006) Robotics and virtual reality: A perfect marriage for motor control research and rehabilitation. *Assistive Technology* 18: 181-195.

Rand, D., Katz, N., Shahar, M., Kizony, R., and Weiss, P. L.: The virtual mall: development of a functional virtual environment for stroke rehabilitation**.** *Abstracts of the 55th Annual Conference of the Israeli Association of Physical and Rehabilitation Medicine.* Tel Aviv 2004.

Riedmiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. Proceedings of the International Conference on Neural Networks, pages 586 – 591

Robertson, I.H. and Halligan, P. W. Spatial neglect: a clinical handbook for diagnosis and treatment. Hove, UK: Psychology Press, 1999.

Weiss, P.L., Kizony, R., Feintuch, U., & Katz, N. Virtual reality in neurorehabilitation. In M.E. Selzer, S. Clarke, L.G. Cohen, P. Duncan, & F. Gage (Eds.). *Textbook of Neural Repair and Rehabilitation - Medical Rehabilitation*. pp 182-197. Cambridge: Cambridge University Press. Cambridge. 2006.

סיווג וקיבוץ של פציעות מוח מנתוני תנועה של חולים
בסביבת מציאות מדומה

נתן סילניצקי

<u>תקציר</u>

בשנים האחרונות התבססה טכנולוגיית מציאות מדומה ככלי יעיל לשיקום חולים עם פגיעות מוח הודות לכך שהיא מאפשרת לחולים להתנהג באופן טבעי בסביבה מדומה. אנחנו מראים שניתן להשתמש בצורה יעילה בנתוני תנועה מתוכניות מציאות מדומה כדי לסווג ולקבץ חולים על פי סוג הפתולוגיה שלהם. במחקר נעשה שימוש ברשתות נוירונים וכלים אחרים כדי לסווג בין חולים עם פציעות טראומה, שבץ עם או בלי הזנחת צד ויחידים בריאים רק על פי נתוני התנועה. שיטות קיבוץ שיכפלו בהצלחה את החלוקה של הסיווג. לתוצאות אלו ישנה משמעות ישומית עבור מחקר מדעי וקליני, ע"י הוספת יכולות ממוכנות לתהליך האבחון וההערכה.

סיווג וקיבוץ של פציעות מוח מנתוני תנועה של חולים
בסביבת מציאות מדומה

מאת: נתן סילניצקי

בהנחיית: פרופ' לארי מנביץ (אוניברסיטת חיפה)
דר' אורי פיינטוך (האוניברסיטה העברית)

עבודת גמר מחקרית (תיזה) המוגשת כמילוי חלק מהדרישות
לקבלת התואר "מוסמך האוניברסיטה"

אוניברסיטת חיפה
הפקולטה למדעי החברה
החוג למדעי המחשב

ספטמבר, 2011

סיווג וקיבוץ של פציעות מוח מנתוני תנועה של חולים
בסביבת מציאות מדומה

נתן סילניצקי

עבודת גמר מחקרית (תיזה) המוגשת כמילוי חלק מהדרישות
לקבלת התואר "מוסמך האוניברסיטה"

אוניברסיטת חיפה
הפקולטה למדעי החברה
החוג למדעי המחשב

ספטמבר, 2011