Cogn Comput manuscript No.
(will be inserted by the editor)

The Importance of Pen Motion Pattern Groups for Semi-Automatic Classification of Handwriting into Mental Workload Classes

Murad Badarna · **Ilan Shimshoni** · **Gil Luria** · **Sara Rosenblum**

Received: date / Accepted: date

Abstract – Background: In this paper we introduce the pen motion pattern groups (PMPGs) and their contribution to the classification of handwriting into cognitive mental workload classes. We demonstrate the importance of PMPGs by providing an efficient semi-automatic machine learning based classification framework that distinguishes between handwritten texts written by the same person under different mental workloads. Our evaluation framework is non-language-dependent since we used stroke features, which are not language-specific, and it takes into account the variability in behavioral biometrics between different writers.

– Methods: The handwritten text data was collected using the Computerized Penmanship Evaluation Tool. This digitizer provided accurate temporal measures throughout the writing. As a first stage, the participants were asked to write a given text in the Hebrew language. Then, as a second stage, the participants' cognitive workload was manipulated by asking them to hold a number in their memory during the entire writing task.

– Results: In our experiments we show that incorporating the PMPGs into the classification process yielded an average cognitive load discrimination accuracy of 92.16%, which decreased to 72.90% when the PMPGs were not considered.

– Conclusions: The separation of handwritten strokes into PMPGs allows us to account for the fact that the strokes are affected differently under different cognitive mental workloads. This novel distinction between PMPGs is important since the handwriting process in each PMPG is different from a perceptual motor and brain-hand control point of view.

Moreover, most of the features that are influenced by cognitive workload are those that cannot be discerned by an expert when looking at a handwritten text on paper, such as azimuth, tilt, velocity, acceleration and pressure.

Keywords Handwriting · Classification · Computerized measures · Mental workload · Digitizer

Murad Badarna
Department of Information Systems, Faculty of Social Sciences, University of Haifa, Haifa, Israel
E-mail: mbadarna@gmail.com

Ilan Shimshoni
Department of Information Systems, Faculty of Social Sciences, University of Haifa, Haifa, Israel

Gil Luria
Department of Human Services, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel

Sara Rosenblum
Laboratory of Complex Human Activity and Participation (CHAP), Department of Occupational Therapy, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel

Introduction

The human cognitive system is of interest to researchers developing computational applications for various purposes in multidisciplinary fields [47, 58, 53, 3, 14, 55, 27, 38]. For more than three decades researchers have attempted to understand human cognitive abilities. However, despite initial promise, there has been little advance towards profound comprehension [27]. More specifically, research is still needed with regard to relationships between the cognitive system and actual activity performance characteristics. Currently, data are obtained mainly through sensors [58]. The present study is concerned with obtaining information about handwriting activity by capturing visual-motor parameters and producing from them a spatio-temporal sequence [48, 26] via an electronic tablet (digitizer). The digitizer is a noninvasive, unobtrusive tool that detects behavioral biometrics of perceptual-motor task performance [8]. Such insights may be valuable for developing computational medical evaluations as well as for a variety of human–computer interactions involving cognitive abilities [46].

Multiple studies have demonstrated the advantages of computerized, objective, spatial, temporal, and pressure measures supplied by the digitizer, for capturing the handwriting process (e.g., [15, 21, 10]). These studies fall into two categories. The first category of studies provided good indicators of clinical and applied conditions (e.g., mild cognitive impairment or Alzheimer's disease [56], and depression [42]). The second category of studies provided evidence that the computerized measures are sensitive to the mental cognitive workload of healthy individuals. For instance, Luria and Rosenblum [25, 24, 40, 38] have shown that the computerized measures of writing under certain workload conditions significantly differ from those obtained when writing under other mental workloads. Moreover, studies (e.g., [33]) have shown that the non-language-dependent measures reflect individual handwriting uniqueness and low/high cognitive load [23].

The overall body of research on computerized handwriting analysis offers evidence that physical, cognitive, clinical, psychological, and situational characteristics of individuals can be captured through these computerized outcome measures. Spatial, temporal, and pressure measures that reflect attributes of writing are sometimes visibly evident on the written page (such as legibility, letter size). The interesting point is that these measures also capture behavior that occurs above the page (in the air, between strokes, tilt velocity, azimuth velocity, etc.) and has no visible outcome on the written page [1, 43, 25].

In order to analyze handwriting automatically, many researchers used stroke features to extract valuable information about individual handwriting uniqueness and to provide indicators of clinical and applied conditions (e.g., analyze behavior organization of children with developmental coordination disorders [35] and analyze handwriting of people with developmental coordination disorders [34, 37, 40]). Some previous works offered evidence of statistical differences between handwritten texts written under different mental workloads of the same writer ([24, 25, 23]). However, all these authors analyzed all strokes similarly without the ability to differentiate between different strokes. Some of these authors even mentioned the problem as a technological limitation of their research [24, 25]. Our study aims to fill this gap using a semi-automatic classification algorithm. We used stroke features in order to build a classification framework for cognitive mental workloads.

There is also considerable variability in the behavioral biometrics between different writers in relation to handwriting. The handwriting strokes have unique characteristics for each writer; therefore, using a single classifier over all writers would not be useful to classify strokes into their related mental workload class.

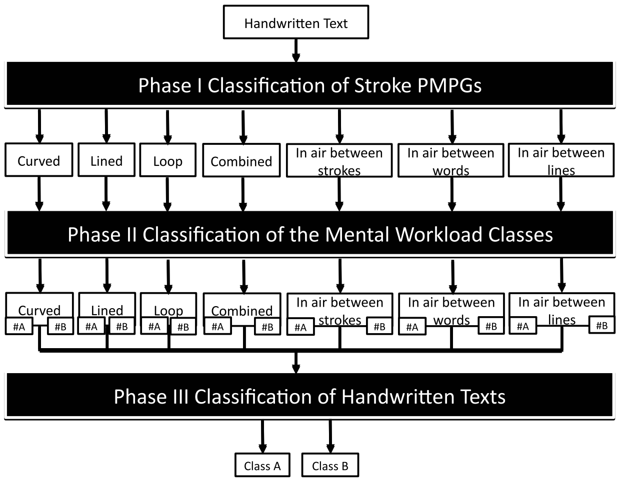


Fig. 1 The handwriting analysis framework

In our study, we separate the analysis of handwriting into **p**en **m**otion **p**attern **g**roups (PMPGs). We solve the problem of variability in handwriting by using a separate classifier for each writer, which is trained on his or her own handwriting strokes in order to classify the PMPGs (see Fig. 1). Next, as a second phase, given the output of each writer's classifier, we trained another classifier that uses these PMPGs to classify each writer's stroke into its related mental workload class. This classification task is quite challenging. We can, however, exploit the fact that the entire handwritten text was performed under a single mental workload class. Therefore, we use the output of the latter classifier as input to a single classifier that attempts to classify the entire handwritten text segment into a single mental workload class.

The proposed handwriting analysis framework is general in the sense that it can be used to classify handwritten text into any two different mental classes A and B. In this paper we show how our framework can be used to distinguish between a given text written under a low mental workload (class A), where the writer is simply asked to write the text, and under a high mental workload (class B), where the writer is asked to write the text while trying to remember a number.

There are several applications of our semi-automatic classification framework in the clinical field [24, 42]. The detection of cognitive load in handwriting has been proven a valid measure of illness (such as Parkinson's disease [5], Alzheimer's disease [56] and [36]) and even as a valid measure for the detection of deception regarding health [23]. Our study takes these existing tools a step further to yield a better and more valid detector of these conditions.

Our proposed classification framework is non-language-dependent since we used stroke features, which are not language specific, as multiple studies have shown [33, 25, 24, 40]. Each PMPG is different from a perceptual motor, brain-hand control point of view. Although the proportion of each PMPG may vary between languages, our classification framework is not affected since this information is not considered by our analysis.

Our study distinguishes itself from previous studies in the literature in the following manner:

- To the best of our knowledge, none of the previous studies provided a technique for semi-automatic classification of handwriting under different cognitive mental workloads. Most previous works presented evidence only of statistical differences [25, 23, 24, 40]. In contrast, in this study, we propose an efficient technique for semi-automatic handwriting analysis using a machine learning based classification framework [18].
- Several studies have proposed machine learning based techniques that successfully extracted valuable information from handwritten text (e.g., gender classification [44], diagnosis of Parkinson's disease [17, 4, 5], signs of aging [7, 6], on-line signature verification [11, 9], automatic diagnosis of Alzheimer's disease [22], analysis of neurological disorders [45], analysis of child development on the basis of handwritten performance [52], recognition of extraversion [12], and classification of emotional states [19]). However, all these methods distinguish between texts written by two groups of writers (users); none has tried to distinguish between two states of the same writer (user). The classification between handwritten strokes of the same writer in two mental workload states is difficult because the differences in the same writer's handwritten strokes are small. Therefore, we need additional knowledge such as the PMPGs in order to successfully classify these mental workload states.
- To the best of our knowledge, no previous studies in the literature distinguish between PMPGs, as we do here for the first time. This distinction is important. For instance, a line stroke is produced by motion of the wrist, with no need for precise control of left-to-right or top-to-bottom motion (or vice versa). On the other hand, producing a loop stroke requires good control of well-synchronized, precise finger movements for pen control while utilizing cognitive resources [54, 29]. Therefore, for each PMPG, the handwriting changes in a different manner. In the experiments, we show that when incorporating the PMPGs into the classification process, the average accuracy was 92.16%, as opposed to 72.90% when the PMPGs were not included.

The insights revealed in this study may be the first step toward achieving automatic activity recognition [58, 26]. They may also further the acquisition of information about the learning and reasoning processes related to individual visual-motor parameters [30].

The remainder of this paper is organized as follows. The Methods section describes the experiments performed on human subjects in order to collect the data. The Background section describes the stroke features, the PMPGs, and a brief summary of the Random Forest classifier, which is used in our system. Building on this description, we introduce our classification framework in the Handwriting Analysis Framework section. Next, a discussion of the main results and the evaluation of our framework are given in the Experiments section. Finally, conclusions and future work are given in the Future Work section.

Methods

Participants

The participants were 88 healthy students, 54 females and 34 males, aged 20–35 (mean age 25.51, SD=3.41), who were recruited at the University of Haifa in northern Israel. 70% of the participants were born in Israel, while 27% were born in the former Soviet Union and 3% in the rest of Europe. The majority (85%) of the participants had right-hand dominance, and 15% were left-handed.

The criteria for inclusion were: residence in Israel for at least 20 years; normal or corrected-to-normal vision and hearing ability; at least 13 years of education; and a minimum of three sentences in Hebrew written at least three times a week. Anyone suffering from any form of neurological/emotional disorder or physical disability was not eligible to participate in the study.

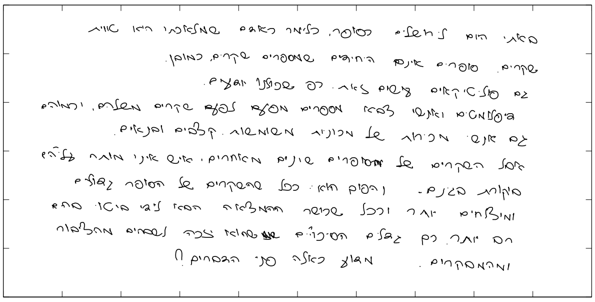
Computerized Measures of Handwriting

The handwritten text data was collected using the Computerized Penmanship Evaluation Tool (ComPET)[39]. All writing was on A4 lined paper affixed to the surface of a WACOM Intuos 2 (Model GD 0912-12X18) xy digitizing tablet, using a wireless electronic pen with a pressure-sensitive tip (Model GP-110). The x and y location and angle of the pen tip were sampled on the digitizer at 100 Hz by means of a 1300-MHz Pentium (R) M laptop computer. The digitizer provided accurate temporal measures throughout the writing, both when the pen was touching the tablet (on-paper time) and when it was raised (in-air time). It also provided accurate spatial measures when the pen was touching the tablet and/or when it was lifted above the digitizer (up to 6 mm). Beyond 6 mm, spatial measurement was not reliable.

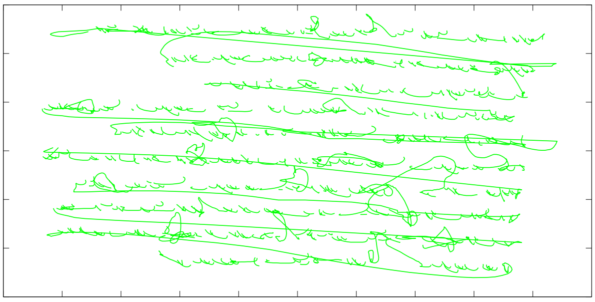
The handwriting evaluation system does not recognize letters, words, or sentences. It only analyzes segments, that is, the curves created by the movement of the pen-tip on the paper, which are represented on an x-, y-coordinate system [28]. That is, the computerized analysis recognizes points when the pen is in contact with and/or leaves the paper. Strokes were measured from when pen pressure rose above 50 (non-scaled units) at the beginning of a stroke to when the pen returned to 50 at the end of the stroke and was raised from the paper.

In conclusion, the pen collects the following raw measures:

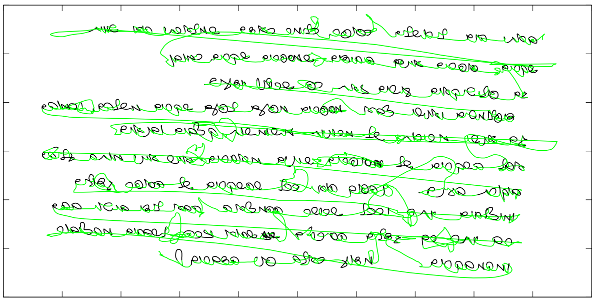
- (x,y) coordinates: the position of the pen on the tablet.
- Pressure (non-scaled units 0–1024): the pressure on the pen.
- Tilt (from 0 – 90): the angle between the pen and its projection on the tablet.
- Azimuth (from 0 – 359): the angle of the pen barrel projected on the x-y tablet plane.
- Timestamp: the time that the sample was taken.



(a) On Surface strokes



(b) In Air strokes



(c) On Surface and In Air strokes

Fig. 2 Low mental workload (class A)

Procedure

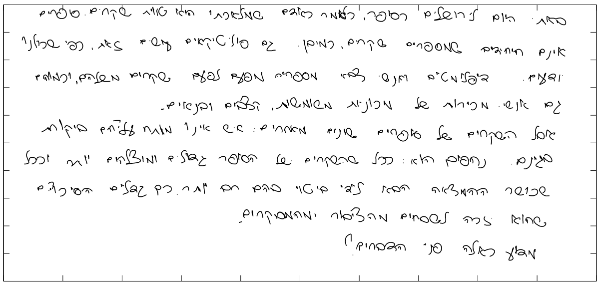
Signed informed consent was obtained from the participants following approval by the Ethical Committee of the University of Haifa. Advertisements at the university were used to recruit students to participate in the study. The socio-demographic questionnaire included gender, age and number of years of education.

As a first stage, the participants were asked by the experts to write a given text in the Hebrew language (see Fig. 2). Here, the cognitive mental workload is low since the participant is only required to write the given text. We henceforth refer to this mental workload level as class A. Then, as a second stage, the participants' cognitive workload was manipulated by asking them to hold the number 5968732 in their memory during the entire writing task (see Fig. 3). In this task the cognitive mental workload level is high and we henceforth refer to this mental workload level as class B. The participants were then asked to report the number at the end of the experiment as well as answer several hypothesis awareness questions.

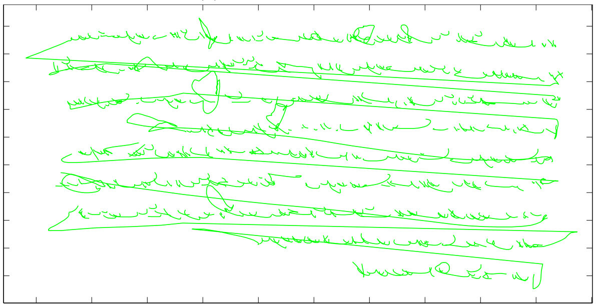
Research assistants ensured that the participants had no chance of writing down the number before being requested to recall it. The overall experimental procedure was supervised by experts from the Laboratory of Complex Human Activity and Participation (CHAP) at the University of Haifa.

In this experiment, the cognitive workload level was increased using a memory trigger task based on numbers. Numbers give better tuning control of the manipulated cognitive workload level than images and words. This specific cognitive load manipulation procedure was adopted in a number of previous studies (e.g., [13, 16, 32, 50]).

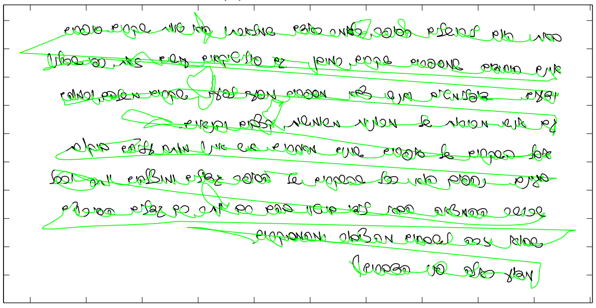
Each participant wrote two identical texts. The first text was written under low cognitive mental work load (marked as class A) and the second was written under high cognitive mental workload (marked as class B). However, as can be seen from both Figures 2 and 3, it would be difficult even for an expert to distinguish between class A and class B when looking at the handwritten text on paper. A classification algorithm that takes into account the perceptual motor and brain-hand control features of the handwriting task is required for this purpose.



(a) On Surface strokes



(b) In Air strokes



(c) On Surface and In Air strokes

Fig. 3 High mental workload (class B)

Background

In this section, we first describe the stroke features and the PMPGs. Next, we briefly describe the Random Forest classifier that we used.

Stroke Features

We used several features to analyze handwriting behavior, which were computed based on the raw measures collected for each stroke:

- **Spatial:** Width, height, displacement, curve length (all measurements were in millimeters):
  - Stroke curve length: the total path length from the starting point to the finishing point for each written stroke.
  - Stroke height (on the y-axis): the direct distance from the lowest point of the stroke to the highest point.
  - Stroke width (on the x-axis): the direct distance from the left side of the stroke to the right side.
  - Displacement: the direct distance between the starting point of stroke and its finishing point.
- **Kinematic:** Velocity (mean, median and, standard deviation (std)) and acceleration (mean, median and, std):
  - Velocity is the speed of writing the stroke, measured in millimeters per second.
  - Acceleration is the rate of change of velocity.
- **Duration time** is the time it takes the writer to write the stroke. Stroke duration in air (while the pen is not in contact with the writing surface) and on paper, both reported in seconds.
- **Pressure**: mean, median, standard deviation, skewness, kurtosis, velocity (mean, median and, std) and acceleration (mean, median and, std).
- **Angles**- tilt, azimuth and, directional angle (the direction angle of writing the stroke): mean, median, std, skewness, kurtosis, velocity (mean, median and, std) and acceleration (mean, median and, std).

The Stroke Pen Motion Pattern Groups (PMPGs)

The strokes were divided into seven PMPGs by an expert (see Fig. 4). These groups take into account the complexity level of the perceptual motor and brain-hand control required to produce the stroke. Strokes with similar complexity levels are clustered together in the same PMPG group. Strokes within the same PMPG share common properties, as follows:

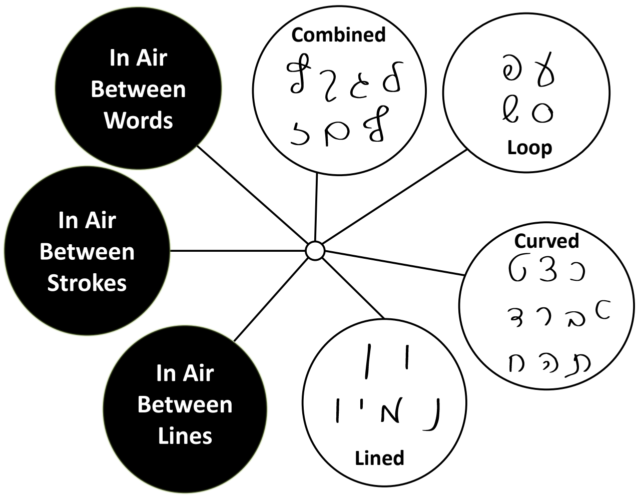
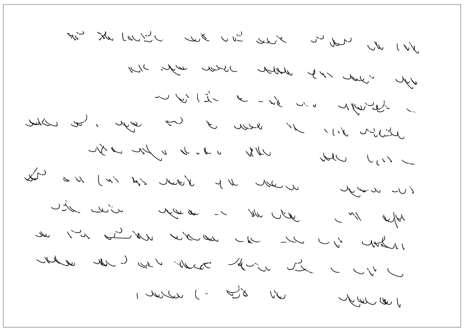


Fig. 4 The Stroke PMPGs

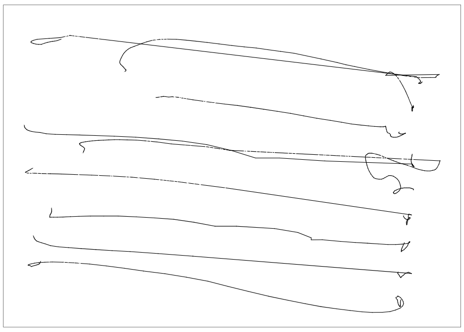
In air PMPGs

- **In Air between strokes**: the transition of the pen in air from one stroke to another (see Fig. 5(a)).
- **In Air between words**: the transition of the pen in air from one word to another (see Fig. 5(b)).
- **In Air between Lines**: the transition of the pen from one line of words to the next one. Note that the digitizer detects the actual motion only below a height of 6 mm above the tablet surface (see Fig. 5(c)).



(a) In Air between strokes

(b) In Air between words

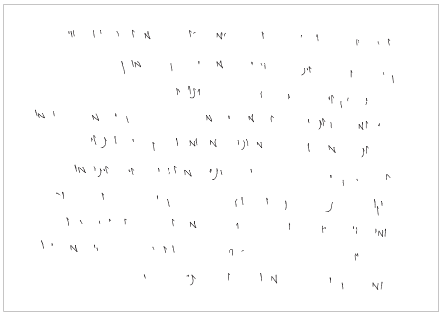


(c) In Air between lines

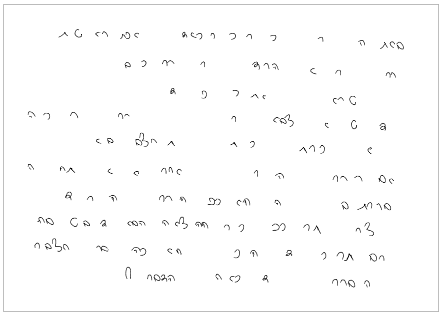
Fig. 5 In Air strokes.

On surface PMPGs

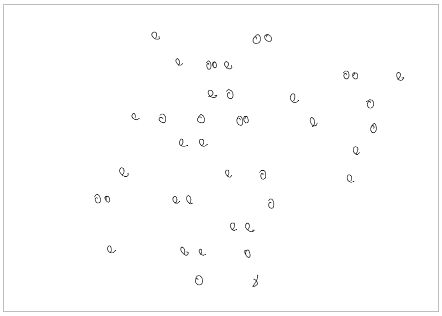
- **Lined**: the stroke consists of straight lines (see Fig. 6(a)).
- **Curved**: the stroke consists of one curve or more (see Fig. 6(b)).
- **Loop**: the stroke consists of a loop (see Fig. 6(c)).
- **Combined**: the strokes consists of two or more different parts - loop and line, curve and loop, or loop and line (see Fig. 6(d)).



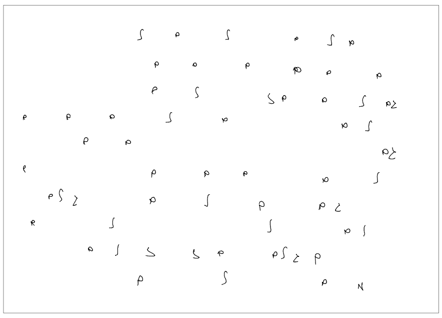
(a) Lined



(b) Curved



(c) Loop



(d) Combined

Fig. 6 On surface strokes.

The Random Forest Classifier

In our handwriting analysis framework, we used the Random Forest classifier [2]. The Random Forest grows many classification decision trees. To classify a new object from an input vector, the input vector is traversed down each of the trees in the forest. The forest chooses the classification having the most votes over all the trees in the forest.

We selected the Random Forest classifier for three reasons. First, it has the unique characteristic of being able to automatically select features, which is the automatic selection of attributes in the data that are most relevant to the predictive modeling problem [2, 20]. In the Experiments section we show the importance of this step in improving classification accuracy. Second, the Random Forest provides a natural way for understanding how each PMPG is affected under different mental workloads, as shown in The Relevance of Stroke Features section. Third, the computational complexity of the training phase of a Random Forest is quite low and its run time is relatively fast.

Handwriting Analysis Framework

The architecture of our classification framework is summarized in Fig. 1. It consists of three phases. In the first phase, we use a classifier that aims to classify each handwriting stroke to its appropriate PMPG. In our initial experiments, we tried to use one classifier for all writers but with little success because of the variability in the behavioral biometrics between them. In order to overcome this problem, we used a separate classifier for each writer.

Here, we manually labeled the PMPG of each stroke written under class A. Then, a Random Forest classifier was trained over these strokes. Using this classifier, we automatically classified the strokes of the same writer that were written under class B. In other words, a Random Forest classifier was trained on each writer's own strokes written under class A. Then, it was used to classify the same writer's strokes written under class B. We found that such a classifier succeeded to classify each stroke to its PMPG of the same writer within an acceptable error rate. However, we will see in the Experiments section that automatic classification of the PMPGs alone is insufficient and should be augmented with manual classification for more accurate overall performance.

The second phase of our framework classifies each stroke into its mental workload class. In this way, we can identify how the features of such a stroke are affected under each class and hence improve the results. A Random Forest classifier was used here as well. Again, because of the variability in behavioral biometrics, we used a separate classifier for each writer. The inputs for this classifier are the handwriting stroke feature vectors and the relevant PMPG. A separate classifier is created for each PMPG. This step is important to our classification scheme, since each PMPG is influenced differently by the writer's mental workload. In this way, each writer has a unique classifier which would classify his/her strokes into the suitable mental workload class.

Fig. 7 describes a decision tree as a part of the Random Forest classifier that reasons whether or not a given stroke is written under a particular mental workload. Given this classifier, the user can predict the mental workload under which the stroke was written and understand the behavioral characteristics of all the strokes in a particular PMPG under this mental workload.

The output of the second phase is two vectors for each writer: the first vector represents the text that was written under class A and the second vector represents the text that was written under class B. The structure of each of these vectors is the same and it contains, for each PMPG, the number of predictions that a stroke is written under class A and the number of predictions that a stroke is written under class B. We cannot expect the accuracy of the classifier to be very high since classifying a single stroke as belonging to class A or B is very challenging. We are able, however, to exploit the fact that the entire text segment was written under a single mental workload class, which is a reasonable assumption made in the experiment.

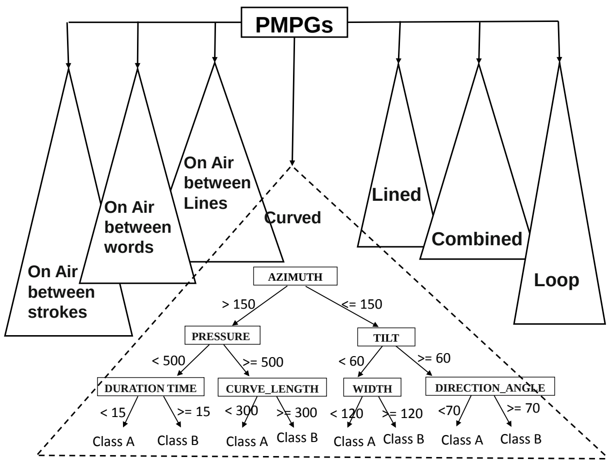


Fig. 7 An illustration for one of the decision trees that was built as a part of a Random Forest classifier.

Thus, the third phase classifies the entire handwritten text of the writer into a single mental workload class. The output of the second phase is used as the input of this phase. A Random Forest classifier was also used to classify these vectors. Here, we run a 10-fold cross-validation and accumulate the results. The output of this phase is the classification accuracy for each mental workload class.

The proposed handwriting analysis framework is general in the sense that it can be used to classify handwritten text into any two different mental classes A and B. In this paper we show how our framework can be used to distinguish between a given text written under a low mental workload (class A), where the writer is simply asked to write the text, and under a high mental workload (class B), where the writer is asked to write the text while trying to remember a number.

Table 1 The classification accuracy percent results

Classification accuracy	**Without using PMPGs**	On surface PMPGs only	In Air PMPGs only	**All PMPGs**
Class A - low workload	**75.73%**	95.23%	92.61%	**94.66%**
Class B - high workload	**70.07%**	85.80%	89.77%	**89.66%**
Average	**72.90%**	90.52%	91.19%	**92.16%**

Experiments

We carried out several experiments to demonstrate the performance of our handwriting analysis framework. We ran each experiment over 88 writers to validate our claim that PMPGs are an important factor in the correct classification of mental workloads and we tested several performance variants of our classification framework. We also show how our classification framework can be used to characterize the influence of mental workload classes on each PMPG.

The Relevance of PMPGs

We ran our analysis framework using several performance variants in order to test how the PMPGs can be used to effectively classify class A and B. As a baseline, we use only the stroke features and ignore the stroke PMPGs. The second variant includes only the on-surface PMPGs. The third test includes only the in-air PMPGs. The full system includes all the PMPGs (i.e., both the on-surface and the in-air PMPGs).

When the classification framework ignores the stroke PMPGs in the analysis, the classification accuracy for class A is 75.73% and the classification accuracy for class B is 70.07% (see Table 1). In contrast, when the PMPGs are considered, the results improve significantly. The classification accuracy for class A when including only the on-surface PMPGs is 95.23% and the classification accuracy for class B is 85.80%. When including only in-air PMPGs, the classification accuracy for class A is 92.16% and 89.77% for class B. When all the PMPGs were included, the classification accuracy is 94.66% for class A and 89.66% for class B. Therefore, the classification accuracy when including all PMPGs outperforms the other variants on average.

One reasonable explanation for the decrease in classification accuracy when the PMPGs are not considered is that each PMPG is influenced by the writer's mental workload in a different way. For instance, from a perceptual motor, brain-hand control point of view, producing a line is less demanding than producing a loop. Lines are produced while the hand makes motions from left to right or from up to down or vice versa. Such motions are mainly done by the wrist with no need for precise control and coordination in space and time of the intrinsic muscles of the hand. On the other hand, producing loops requires good control of well-synchronized, precise finger movements for pen control while utilizing cognitive resources. Thus, separating the strokes into their PMPG categories improves the classification accuracy.

The disadvantage of automatic labeling of PMPGs

As mentioned in the Handwriting Analysis Framework section, we manually labeled the PMPGs of the strokes in class A and automatically labeled them for the strokes in class B. The reason for this choice was the time and cost involved in manual labeling, but automatic labeling did negatively affect the class B results. We can improve the results of class B by manually labeling the PMPGs. Table 2 shows the result vectors for a certain writer after a manual labeling process. As can be seen, the number of correct predictions that a stroke is written under a certain workload class was much improved. For instance, in the automatic labeling process, there are 813 correct predictions that a curved stroke is written under class B as opposed to 1414 in the manual labeling process: 601 more correct predictions. We found that this manual labeling improved the results in phase III.

Why we use the Random Forest classifier in the third phase

One might claim that a simple aggregation such as averaging over the values of the feature vector of the text in the third phase is sufficient to classify the text to the correct mental

Table 2 The resulting vectors for the writer after a manual labeling process

	Combined		Lined		Loop		Curved		In Air Between Words		In Air Between Stokes		In Air Between Lines	
	#A	#B	#A	#B	#A	#B	#A	#B	#A	#B	#A	#B	#A	#B
Automatic Labeling	284	416	548	742	165	274	727	813	338	442	1663	1897	31	59
Manual Labeling	102	598	134	1156	62	377	126	1414	71	689	253	3307	8	82
Difference	-182	182	-414	414	-103	103	-601	601	-267	267	-1414	1414	-23	23

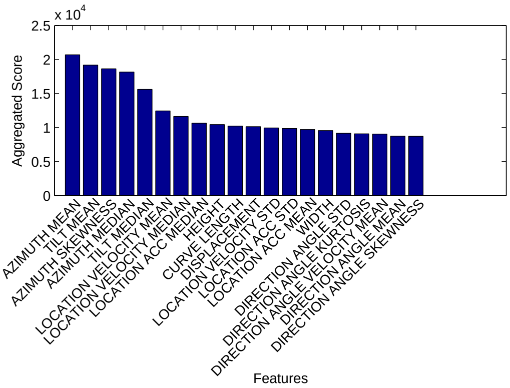
workload. We tested this idea by replacing the Random Forest classifier in the third phase with the Linear SVM classifier, which is a type of weighted averaging based classifier. As a result, the classification accuracy of class A when all PMPGs were included decreased to 89.32% while the classification accuracy of class B decreased to 86.25%.

The Relevance of Stroke Features

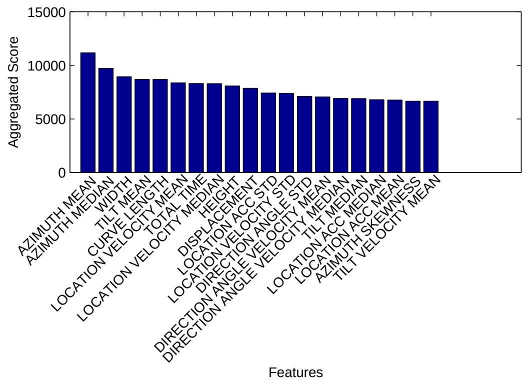
Here we investigate the relevance of the stroke features under each one of the PMPGs. The depth of a feature that is used as a decision node in a decision tree can be used to assess the relative relevance of that feature with respect to the predictability of the target variable. Features used at the top of the decision tree contribute to the final prediction decision of a larger fraction of the input samples and therefore contribute more than other features to the final classification accuracy [2]. The expected fraction of the samples they contribute to can thus be used as an estimate of the relative relevance of the features. Therefore, the relevance score of a feature is computed as $\frac{1}{2^{depth}}$. In the Random Forest classifier, there are several decision trees. Therefore, the relative relevance of feature k is the aggregated score over all of the decision trees for all the writers, computed as

$$AggregatedScore_{(feature_k)} = \sum_{i=1}^{Writers} \left(\sum_{j=1}^{DecisionTrees} \frac{1}{2^{depth(i,j,k)}} \right). \quad (1)$$

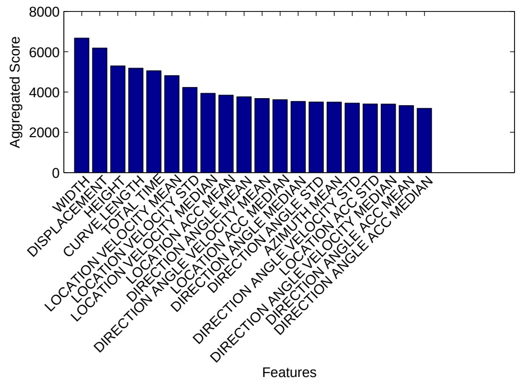
Figs. 8 and 9 show the aggregated score versus stroke features for each PMPG. As can be seen from the results, each PMPG is influenced by cognitive workload in a different way. A higher aggregated score means greater importance of the feature. Furthermore, most of the features influenced by cognitive workload cannot be discerned by the human eye. The most affected features are those related to azimuth, tilt, velocity, acceleration and pressure these features are affected differently according to the pen stroke type. For instance, the three most affected features for the line PMPG are those related to azimuth and tilt while for the combined PMPG they are those related to azimuth, pressure and velocity. Moreover, for the in-air strokes between lines PMPG, the most affected features are those related to width, displacement and height, which cannot be seen at all because the strokes occur in the air. As was demonstrated above, using only the in-air stroke types yields accurate classification results.



(a) In Air between strokes

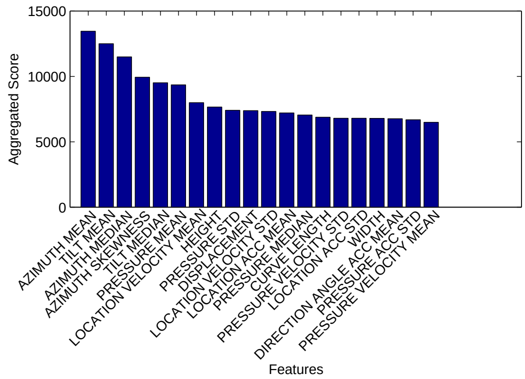


(b) In Air between words

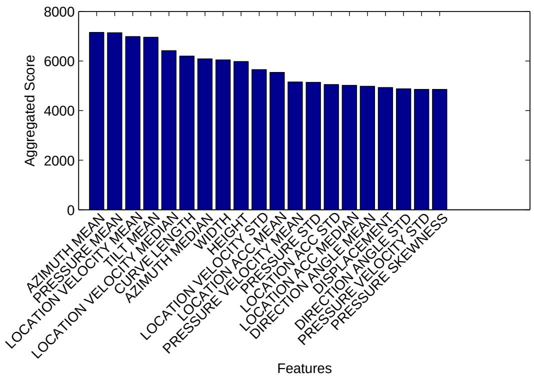


(c) In Air between lines

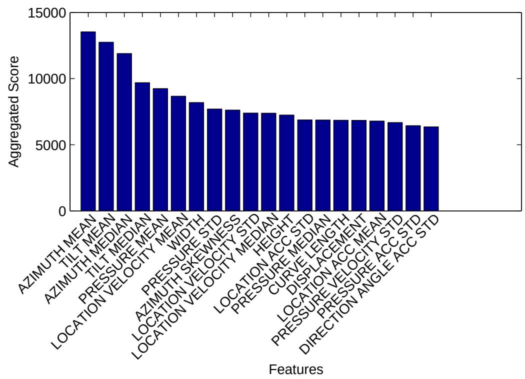
Fig. 8 The aggregated score versus stroke features for each In Air PMPG



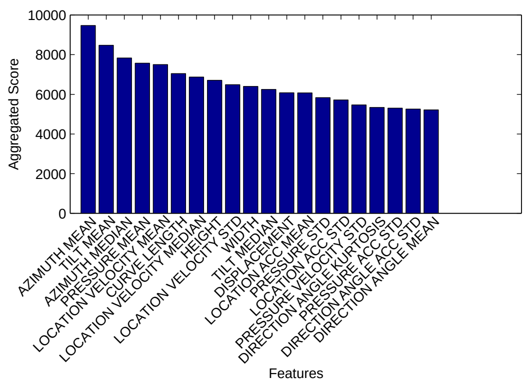
(a) Lined



(b) Combined



(c) Curved



(d) Loop

Fig. 9 The aggregated score versus stroke features for each on surface PMPG

These differences in the affected features for each PMPG may be explained by differences in the mechanical movements of the hand during the writing process under each PMPG. When producing straight lines, the motion is mainly performed by the wrist and almost no changes occur at the finger level; thus, in this case the Azimuth and tilt reflect the writer's control of the pen in space while creating up and down and right left motions. However, in order to produce circles, a very accurate sequence and timing is required from inner small hand muscles such as the lumbricales and interossei. Thus, in higher cognitive load situations, the production velocity and pen pressure towards the paper were also influenced, in addition to azimuth and tilt, because they express dexterous in-hand manipulations [57, 49, 51].

Table 3 The three most important features for each PMPG.

PMPG	**Feature 1**	**Feature 2**	**Feature 3**
In Air between strokes	AZIMUTH MEAN	TILT MEAN	AZIMUTH SKEW-NESS
In Air between words	AZIMUTH MEAN	AZIMUTH MEDIAN	WIDTH
In Air between lines	WIDTH	DISPLACEMENT	HEIGHT
Lined	AZIMUTH MEAN	TILT MEAN	AZIMUTH MEDIAN
Curved	AZIMUTH MEAN	TILT MEAN	AZIMUTH MEDIAN
Loop	AZIMUTH MEAN	TILT MEAN	AZIMUTH MEDIAN
Combined	AZIMUTH MEAN	PRESSURE MEAN	LOCATION VELOC-ITY MEAN

As we can see from Table 3, for the in air between strokes and words PMPGs, the most affected features are those related to the mechanical movement of the writers hand, such as azimuth and tilt. However, when the level of the cognitive workload is high, such as for the in air between lines PMPG, the complexity of the cognitive workload demands is reflected in other features, such as displacement and height of the strokes. For on surface PMPGs, the level of the cognitive workload demands for producing the lined, curved and loop PMPGs is consistent since for these PMPGs the cognitive workload level is low. Therefore, the affected features are only those related to the mechanical movement of the writer's hand, such as azimuth and tilt. However, when the process of producing the PMPG strokes is complex, such as for the 'combined' PMPG (see Fig. 4), the high cognitive workload demand is reflected in other features, such as pressure and velocity. In conclusion, there is a mutual relationship between motor control and the cognitive demands of each kind of written stroke. For a high level of cognitive workload, the affected features are those related to motor control, such as azimuth and tilt, in addition to pressure and velocity.

As a whole, those results support previous results that pen tilt, pen pressure and velocity were all influenced by cognitive decline among elderly people [41], those with minimal cognitive impairment or Alzheimer's disease [56], and among people with depression [42].

Limitations

In this study, all participants were university students with a small variability and range of ages. Rosenblum et al. [42, 41] have shown that the performance changes significantly in the elderly population, possibly limiting the study's results. Future research should include a sample with older participants. However, it is important to note that the relationships between the classification performance and variables such as age, gender and, hand dominance were examined in the first stage of this study. We found that these variables had no effect on the results.

Future Work

In this study we proposed a classification framework that is general in that it can be used to distinguish between mental workload classes. Given handwritten texts written under different mental workload classes, our classifier is able to distinguish between these texts. There are several directions for future work:

- In this study, we manually labeled the PMPGs of the strokes of the class A text. Then, a classifier was trained over these strokes and used to classify the class B text. As has been shown in the experiments, the performance of our framework was affected by this process. To overcome this problem, we intend to improve the classifier by manually labeling the PMPGs of a small set of selected strokes from class B. The aim of this step is to modify the current classifier to work better for class B strokes. Hence, the questions are how to choose these strokes and how many strokes are needed. The answers to these questions can be obtained using methods from the field of transfer learning [31].
- An interesting extension that we would like to test is how to classify the handwritten text of a new person (X) who was not included in our previous analyzed data set. First, we plan to find the most similar person (S) from the cognitive mental workload point of view to person (X) out of the people we already analyzed. Then, we plan to modify the classifier that was built for person (S) in order to work better for person (X).
- Currently, our classification framework is able to distinguish between two mental workload classes. However, in many fields, e.g., mild cognitive impairment or Alzheimer's disease [56] and depression [42], evaluating the effects of mental workload on handwriting behavior under different stages of the disease is also important. For a disease such as Alzheimer's, we would like to be able to discriminate between handwritten texts written under different stages of the disease. We plan to use a regression based classifier for this purpose.

Compliance with Ethical Standards

All the authors of this study declare that they have no conflict of interest. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee. Informed consent was obtained from all individual participants included in the study.

References

1. Carlos Alonso-Martinez, Marcos Faundez-Zanuy, and Jiri Mekyska. A comparative study of in-air trajectories at short and long distances in online handwriting. *Cognitive Computation*, Jul 2017.
2. Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
3. Erik Cambria and Amir Hussain. *Sentic Computing: Techniques, Tools, and Applications*, volume 2. Springer Science & Business Media, 2012.
4. P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal, and M. Faundez-Zanuy. Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(3):508–516, 2015.
5. Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, and Marcos Faundez-Zanuy. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease. *Artif. Intell. Med.*, 67:39–46, 2016.
6. M. Faundez-Zanuy, E. Sesa-Nogueras, and J. Roure-Alcob. On the relevance of age in handwritten biometric recognition. In *IEEE International Carnahan Conference on Security Technology (ICCST)*, pages 105–109, 2012.
7. M. Faundez-Zanuy, E. Sesa-Nogueras, J. Roure-Alcob, A. Esposito, J. Mekyska, and K. Lpez de Ipia. A preliminary study on aging examining online handwriting. In

5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), pages 221–224, 2014.

8. Marcos Faundez-Zanuy, Amir Hussain, Jiri Mekyska, Enric Sesa-Nogueras, Enric Monte-Moreno, Anna Esposito, Mohamed Chetouani, Josep Garre-Olmo, Andrew Abel, Zdenek Smekal, and Karmele Lopez de Ipina. Biometric applications related to human beings: there is life beyond security. *Cognitive Computation*, 5(1):136–151, 2013.
9. Julian Fierrez, Javier Ortega-Garcia, Daniel Ramos, and Joaquin Gonzalez-Rodriguez. Hmm-based on-line signature verification: Feature extraction and signature modeling. *Pattern Recogn. Lett.*, 28(16):2325–2334, December 2007.
10. Tom Froese, Keisuke Suzuki, Yuta Ogai, and Takashi Ikegami. Using human–computer interfaces to investigate 'mind-as-it-could-be' from the first-person perspective. *Cognitive Computation*, 4(3):365–382, 2012.
11. Marta Gomez-Barrero, Javier Galbally, Julian Fierrez, Javier Ortega-Garcia, and Réjean Plamondon. Enhanced on-line signature verification based on skilled forgery detection using sigma-lognormal features. In *2015 International Conference on Biometrics (ICB)*, pages 501–506. IEEE, 2015.
12. Zuzanna Górska and Artur Janicki. Recognition of extraversion level based on handwriting and support vector machines. *Perceptual and Motor Skills*, 114(3):857–869, 2012.
13. Richard H Gramzow, Lowell Gaertner, and Constantine Sedikides. Memory for in-group and out-group information in a minimal group context: the self as an informational base. *Journal of Personality and Social Psychology*, 80(2):188–205, 2001.
14. Amir Hussain and Muaz Niazi. Toward a formal, visual framework of emergent cognitive development of scholars. *Cognitive Computation*, 6(1):113–124, 2013.
15. Sonia Kandel and Sylviane Valdois. Syllables as functional units in a copying task. *Language and Cognitive Processes*, 21(4):432–452, 2006.
16. Sander L Koole, Ap Dijksterhuis, and Ad van Knippenberg. What's in a name: implicit self-esteem and the automatic self. *Journal of personality and social psychology*, 80(4):669–685, 2001.
17. C Kotsavasiloglou, N Kostikis, D Hristu-Varsakelis, and M Arnaoutoglou. Machine learning-based classification of simple drawing movements in Parkinson's disease. *Biomedical Signal Processing and Control*, 31(1):174–180, 2016.
18. S. B. Kotsiantis. Supervised machine learning: A review of classification techniques. In *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering*, pages 3–24, Amsterdam, The Netherlands, 2007.
19. L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clmenon, and G. Cordasco. Emothaw: A novel database for emotional state recognition from handwriting and drawing. *IEEE Transactions on Human-Machine Systems*, 47(2):273–284, April 2017.
20. Laurence Likforman-Sulem, Anna Esposito, Marcos Faundez-Zanuy, Bassis-Simone Clémençon, Stéphan, Anna Esposito, and Carlo Francesco Morabito. Extracting style and emotion from handwriting. In *Advances in Neural Networks: Computational and Theoretical Issues*, pages 347–355, 2015.
21. Mitchell G Longstaff and Richard A Heath. The influence of motor system degradation on the control of handwriting movements: A dynamical systems analysis. *Human Movement Science*, 22(1):91–110, 2003.
22. K. López-de Ipiña, J. B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C. M. Travieso, M. Ecay-Torres, P. Martínez-Lage, and H. Eguiraun. On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature. *Cognitive Computation*, 7(1):44–55, 2015.
23. Gil Luria, Allon Kahana, and Sara Rosenblum. Detection of deception via handwriting behaviors using a computerized tool: toward an evaluation of malingering. *Cognitive Computation*, 6(4):849–855, 2014.
24. Gil Luria and Sara Rosenblum. Comparing the handwriting behaviours of true and false writing with computerized handwriting measures. *Applied Cognitive Psychology*, 24(8):1115–1128, 2010.
25. Gil Luria and Sara Rosenblum. A computerized multidimensional measurement of mental workload via handwriting analysis. *Behavior Research Methods*, 44(2):575–586, 2012.
26. Zeeshan Khawar Malik, Amir Hussain, and Jonathan Wu. Novel biologically inspired approaches to extracting online information from temporal data. *Cognitive Computation*, 6(3):595–607, 2014.
27. James L. McClelland. Is a machine realization of truly human-like intelligence achievable? *Cognitive Computation*, 1(1):17–21, 2009.
28. Roland Mergl, Peter Tigges, Andreas Schröter, Hans-Jürgen Möller, and Ulrich Hegerl. Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives. *Journal of Neuroscience Methods*, 90(2):157–169, 1999.
29. Ruud GJ Meulenbroek, Annie Vinter, and Pierre Mounoud. Development of the start-rotation principle in circle production. *British Journal of Developmental Psychology*, 11(3):307–320, 1993.
30. Vishwanathan Mohan, Pietro Morasso, Giulio Sandini, and Stathis Kasderidis. Inference through embodied simulation in cognitive robots. *Cognitive Computation*, 5(3):355–382, 2013.

31. Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
32. Delroy L Paulhus, Peter Graf, and Mark Van Selst. Attentional load increases the positivity of self-presentation. *Social Cognition*, 7(4):389–400, 1989.
33. KW Poon, CWP Li-Tsang, TPL Weiss, and S Rosenblum. The effect of a computerized visual perception and visual-motor integration training program on improving chinese handwriting of children with handwriting difficulties. *Research in Developmental Disabilities*, 31(6):1552–1560, 2010.
34. Sara Rosenblum. Handwriting measures as reflectors of executive functions among adults with developmental coordination disorders (dcd). *Frontiers in Psychology*, 4(1):357–367, 2013.
35. Sara Rosenblum. Do motor ability and handwriting kinematic measures predict organizational ability among children with developmental coordination disorders? *Human Movement Science*, 43(1):201–215, 2015.
36. Sara Rosenblum, Liora Epsztein, and Naomi Josman. Handwriting performance of children with attention deficit hyperactive disorders: A pilot study. *Physical & Occupational Therapy In Pediatrics*, 28(3):219–234, 2008.
37. Sara Rosenblum and Miri Livneh-Zirinski. Do relationships exist between brain-hand language and daily function characteristics of children with a hidden disability? In *Recent Advances of Neural Network Models and Applications: Proceedings of the 23rd Workshop of the Italian Neural Networks Society (SIREN)*, pages 269–281, 2014.
38. Sara Rosenblum and Gil Luria. Applying a handwriting measurement model for capturing cognitive load implications through complex figure drawing. *Cognitive Computation*, 8(1):69–77, 2016.
39. Sara Rosenblum, Shula Parush, and Patrice L Weiss. Computerized temporal handwriting characteristics of proficient and non-proficient handwriters. *American Journal of Occupational Therapy*, 57(2):129–138, 2003.
40. Sara Rosenblum, Hemda Amit Ben Simhon, and Eynat Gal. Unique handwriting performance characteristics of children with high-functioning autism spectrum disorder. *Research in Autism Spectrum Disorders*, 23(1):235 – 244, 2016.
41. Sara Rosenblum and Perla Werner. Assessing the handwriting process in healthy elderly persons using a computerized system. *Aging Clinical and Experimental Research*, 18(5):433–439, 2006.
42. Sara Rosenblum, Perla Werner, Tal Dekel, Ilya Gurevitz, and Jeremia Heinik. Handwriting process variables among elderly people with mild major depressive disorder: a preliminary study. *Aging Clinical and Experimental Research*, 22(2):141–147, 2010.
43. Enric Sesa-Nogueras, Marcos Faundez-Zanuy, and Jiří Mekyska. An information analysis of in-air and on-surface trajectories in online handwriting. *Cognitive Computation*, 4(2):195–205, 2012.
44. Enric Sesa-Nogueras, Marcos Faundez-Zanuy, and Josep Roure-Alcobé. Gender classification by means of online uppercase handwriting: A text-dependent allographic approach. *Cognitive Computation*, 8(1):15–29, 2016.
45. Z. Smekal, J. Mekyska, I. Rektorova, and M. Faundez-Zanuy. Analysis of neurological disorders based on digital processing of speech and handwritten text. In *International Symposium on Signals Circuits and Systems (ISSCS)*, pages 1–6, 2013.
46. Stefano Squartini and Anna Esposito. Co-worker: toward real-time and context-aware systems for human collaborative knowledge building. *Cognitive Computation*, 4(2):157–171, 2012.
47. Stefano Squartini, Sanqing Hu, and Qingshan Liu. Advances on brain inspired computing. *Cognitive Computation*, 5(2):161–163, 2013.
48. James V Stone. Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. *Neural Computation*, 8(7):1463–1492, 1996.
49. Yu Suzuki, Misue Kazuo, and Tanaka Jiro. Stylus enhancement to enrich interaction with computers. In *Proceedings of the 12th International Conference on Human-computer Interaction: Interaction Platforms and Techniques*, HCI'07, pages 133–142, 2007.
50. William B Swann, J Gregory Hixon, Alan Stein-Seroussi, and Daniel T Gilbert. The fleeting gleam of praise: Cognitive processes underlying behavioral reactions to self-relevant feedback. *Journal of Personality and Social Psychology*, 59(1):17, 1990.
51. Taylor Brandon T., Bove Jr., and V. Michael. Graspables: Grasp-recognition as a user interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 917–926, 2009.
52. Narges Tabatabaey-Mashadi, Rubita Sudirman, Richard M. Guest, and Puspa Inayat Khalid. Analyses of pupils' polygonal shape drawing strategy with respect to handwriting performance. *Pattern Analysis and Applications*, 18(3):571–586, 2015.
53. J. G. Taylor. Cognitive computation. *Cognitive Computation*, 1(1):4–16, 2009.
54. Hans-Leo Teulings, José L Contreras-Vidal, George E Stelmach, and Charles H Adler. Parkinsonism reduces coordination of fingers, wrist, and arm in fine motor control. *Experimental Neurology*, 146(1):159–170, 1997.

55. Carlos M. Travieso and Jess B. Alonso. Special issue on advanced cognitive systems based on nonlinear analysis. *Cognitive Computation*, 5(4):397–398, 2013.
56. Perla Werner, Sara Rosenblum, Gady Bar-On, Jeremia Heinik, and Amos Korczyn. Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 61(4):228–236, 2006.
57. Bi Xiaojun, Moscovich Tomer, Ramos Gonzalo, Balakrishnan Ravin, and Hinckley Ken. An exploration of pen rolling for pen-based interaction. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, pages 191–200, 2008.
58. Zhongtang Zhao, Zhenyu Chen, Yiqiang Chen, Shuangquan Wang, and Hongan Wang. A class incremental extreme learning machine for activity recognition. *Cognitive Computation*, 6(3):423–431, 2014.