Integrating Multi-Source Data into HandSpy

Hristo Valkanov
Faculty of Sciences, University of Porto, Portugal
up200803561@fc.up.pt

José Paulo Leal
CRACS & INESC Tec LA, Porto, Portugal
Faculty of Sciences, University of Porto, Portugal
jpleal@fc.up.pt

Abstract

To study how emotions affect people in expressive writing, scientists require tools to aid them in their research. The researchers at M-BW use an Experiment Management System, called HandSpy to store and analyze the hand-written productions of participants. The input is stored as digital ink and then displayed on a web-based interface.

To assist the project, HandSpy integrates with new sources of information to help researchers visualize the link between psychophysiological data and written input. The newly acquired data is synchronized with the existing burst-pause interval model and represented on the user interface of the platform together with the already existing information.

2012 ACM Subject Classification Applied computing → Psychology; General and reference → Experimentation; Human-centered computing → Visualization

Keywords and phrases HandSpy, emotion, handwriting, psychology, psychophysiological, data, ems

Digital Object Identifier 10.4230/OASIcs.SLATE.2020.13

Category Short Paper

Funding *José Paulo Leal*: Fundação para a Ciência e a Tecnologia (FCT), within project UIDB/50014/2020.

Acknowledgements We want to thank the researchers and investigators at M-BW and José Paiva for the input and assistance on the project.

1 Introduction

The link between emotions and psychophysiological events is one of the means used in the psychological analysis of behavior. Based on this theory, the scientists in Mind-Body Interactions in Writing (M-BW) Project are studying the cognitive processes of writing[4]. Experimental Management Systems (EMS) is software oriented towards aiding researchers in their studies. One such EMS, used by the project is HandSpy, which allows scientists to visualize data together with a representation of hand-writing. The input is displayed based on intervals of interest, called *bursts*, and the pauses between them.

The project studies the real-time psychophysiological and linguistic markers of expressive writing. Expressive writing is a particular form of writing in which the study participant engages in narrating a personal, deeply charged, emotional event. Despite the numerous healing effects, the mechanism through which expressive writing operates is still poorly understood. To store and analyze the digital ink representation of the strokes, M-BW uses HandSpy, which is collected using smartpens and microdotted paper.[4] The information is stored in XML format, called InkML which is the recommended standard for storing digital ink traces.[6]

9th Symposium on Languages, Applications and Technologies (SLATE 2020).
Editors: Alberto Simões, Pedro Rangel Henriques, and Ricardo Queirós; Article No. 13; pp. 13:1–13:8
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

To further aid the research, improvements were implemented, so that the platform supports the functionalities required by the project. The new features are based on the collection and representation of psychophysiological indicators from various external sources and centralizing them in a cohesive way inside HandSpy. The newly obtained data will be stored and adapted to the Pause-Burst model and visually represented on the current interface implementation of written text. The implementation of those improvements is the focus of this paper.

2 State of the Art

Without the use of proper tools, trivial tasks can become time-consuming and can hinder the workflow of scientific experiments. While many commercial solutions that can partially manage experiment data exist, they are oriented, mostly, towards business use. They are useful for their mainstream usage but are not entirely adequate for the scientific experimental environment. Those solutions also require a significant monetary and human resource investment, not to mention the proprietary restrictions, which limit their distribution within the research community.[3]

2.1 Experimental Management Systems

One of the problems faced by the experimental sciences is the lack of proper software. Many of the studies require software tailored, specifically, to satisfy the requirements of the area of study, and the researchers involved in it. The need for such tools led to the creation of software, called Experimental Management Systems (EMS), that targeted scientific research environment, specifically. There have been many attempts to identify what are the proper requirements for an EMS. They have been narrowed down to the following requirements[3]:

- **Abstraction from the users** – The system should be as abstract as possible and, ideally, require no programming knowledge from the users.
- **System Integration** – The integration should be handled only at the highest level by the user. All the implementation details should be, previously, provided by the EMS programmers and be readily available to the researchers.
- **Data Integration** – The data should be integrated, in a manner that will allow for it to be handled, without requiring any knowledge of the source.
- **Remote Collaboration Facilities** – The platform should be available for access remotely. Web technologies are a natural choice for this goal.
- **Advanced Data Type Management** – Should handle on its own all the types of data required by the experiment.
- **Intelligent and adaptive UI** – The interface should be intuitive and without as little unneeded information as possible.

2.2 Existing Solutions

While the requirements for building an EMS seem simple, they are hard to fulfill. Currently, there is no commercial solution that stores the digital ink oriented towards studying handwriting, which can substitute an EMS. Those solutions that can perform some tasks or execute a part of the process exist, but none of them serves as a complete EMS. Other downsides to them include: not being directed towards managing scientific experiments, their cost, among others, which limit their adoption. There are other EMS that exists for handwriting analysis, but most of them are not suited for the research that HandSpy assists, namely M-BW. Therefore they were not picked as a target for the improvements in question. The specific shortcomings of each of them are explained below.

2.2.1 Eye and Pen

Eye and Pen is a system that is mainly, but not exclusively, used in the context of handwriting studies. It is oriented towards the analysis of both text and drawings. The platform is designed to record, synchronously, the gaze, and the process of handwriting. For the writing part, the researchers use a digitizing tablet, to obtain the location and state of the pen. An eye-tracking system records the writer's gaze while performing tasks. The main purpose of the system is to study the synchronization between eye and pen movements writing periods. The devices used are focused on allowing users to participate in the investigation, without hindering the activity or increasing their cognitive load.[1] Its shortcomings, compared to HandSpy, due to the usage of a tablet, which is more invasive and less scalable, than the HandSpy approach. Eye and Pen also lacks support for the psychophysiological data, required by M-BW.

2.2.2 Dictus

Dictus is a software that combines stimulus presentation, such as images, or sounds, and movement analysis. The data is gathered by a module that acquires the measures from a hand-writing movement using only a Wacom tablet. This approach has the downside of collecting information form a single source. The main advantage of this method is that the information gathering process is flexible. Because of this, the system is usable in any environment without the need for more sophisticated machinery. This advantage has many upsides. For example, it can be used in both school rooms and laboratories. Another advantage is that it does not involve any obstructing equipment which allows the experiments to be carried with the target population while maintaining the feeling of writing, as they usually do in their everyday life.[2] It shares the same disadvantages as Eye and Pen while being even more restricted due to the limitations of the tablet brand.

2.2.3 HandSpy

The HandSpy platform is a web-based EMS that excels in handwriting research studies. It supports large amounts of data sets. It is designed as a web platform to support remote use and to eliminate the need for installation on multiple interfaces and accommodate collaborative and remote work. It currently supports collections of data, related to the handwriting process such as the write/pause intervals and the distance covered. HandSpy supports input from smartpens to provide data for the studies, which is less intrusive than the input devices used by the other EMS, which is crucial since obstructions during writing affect the collected data.

The input is collected and stored in the server, and available on request. HandSpy also supports annotations such as the start and end of the burst, with a blue letter "p" and a red letter "q", respectively, as shown in Figure 1, and storing additional information, such as the number of words in the interval. [5]

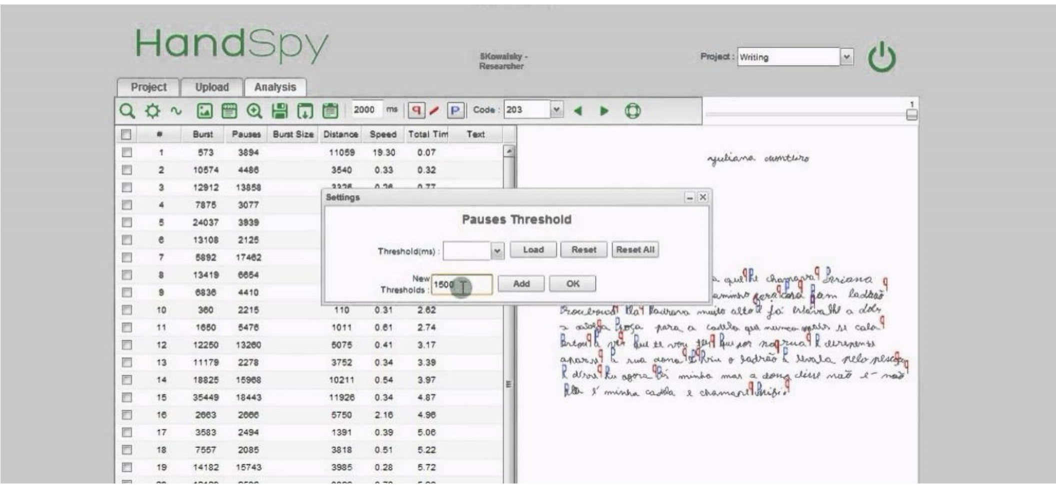

Figure 1 HandSpy showing the annotations for the start and end of the burst.

3 Improvements to HandSpy

HandSpy consists of two components: a server and a client. The server obtains data, in the form of digital ink, from the smartpen and pre-processes it. The information is made available, upon request, to the local client, which allows the users to further work with it. The client is used to filter the data set sent form the server and display it on the UI.

Until now, HandSpy had only a few forms of representation of research related data. They used to be limited to 3 markers: a trace that runs through the selected area; the beginning of a burst; and the end of a burst. This interface is insufficient for the ongoing project, and improvements were required so that HandSpy could assist the M-BW project.

The changes to the current process include the integration with new data sources and the changes to the UI, which will allow us to show more annotations, without cluttering the front-end. HandSpy is aggregating the data from the new sources, based on the previously existing pause-burst intervals.

To display the data, the UI is adapted to create overlays over the points of interest. They are calculated based on the traces of the digital ink, already stored in HandSpy. The overlays are shaped as bounding boxes and mark the separate lines of text.

3.1 Server

The server is divided into three layers, Core, Commands, and Protocol. Due to the architecture of the software, the only change required was in the last of those components. The protocol module handles the request processing and the response generation for the different controllers. Most of the business logic is stored here, inside the Protocol class. This class obtains the connection information from the core, fetches data from the database, parses the requests, generates responses, calculates intervals, and so on.

HandSpy stores the data related to the written productions. They are stored as trace entries, using InkML format, inside the database. Each trace contains a set of points and represents the movement of the pen, starting from the moment the pen touches the paper and ends when the pen leaves the paper. It consists of a set of points that describe the trajectory of the movement. Each "point" holds three values, the InkML bi-dimensional coordinates for X and Y, combined with a timestamp of when the reading was measured.

This list of coordinates is used by the platform to calculate the distance covered by the smartpen, while writing. Meanwhile, the timestamps define the intervals of interest on which the research calculations are based.

3.1.1 Extraction of Intervals

While the existing annotations, such as the beginning and the end of the intervals, served their purpose at the time, the platform lacks specific characteristics, namely the integration with additional sources of data. This improvement is significant because it is one of the requirements for a fully functioning experiment data management system[3]. To support this feature, a dedicated entity, which represents the intervals, was created. This separation was, later used, as a baseline for the implementation of all future metrics.

The code which performed the calculation of the existing intervals was extracted from the protocol class. The newly created modules, which hold the new structure, consists of separate "classes" that represent the different elements of the studies, namely pauses and bursts. An interval contains the logic required to extract the currently existing metrics and all the points for this calculation. Burst objects are identical to the previously calculated pause-bursts, and pauses represent the interval between them.

3.1.2 Lines

The previous version of HandSpy (2.3) displays only the beginning and the end of an interval. The visual representation had to change due to the new functionalities implemented for the next version of the platform. The choice for the visual representation of lines is by using rectangle shaped overlays that encapsulate the points that the interval contains. The bounding box calculation of the lines is performed by grouping the points first into a histogram and then again by using a density threshold and a tolerance range.

The method of calculation is simple since the visual representation does not require very high precision. For each burst interval, the InkML coordinate points are separated by the vertical coordinate value to obtain the number of points that share the same horizontal line. The density threshold is, then, determined by applying a sensitivity percentage to the group with the highest point count. The number obtained, determines if a horizontal line will or will not be part of the marker system. The groups are, then, compared to the threshold, and if they are above it, they are combined with the neighboring lines, until two consecutive lines are below the limit. When this happens, the last observed line is marked finalized, and the bounding box calculation occurs. The corners of the rectangle, which defines it, assume the maximum and minimum values of the X and Y values.

The entire line calculation process occurs on the server, and solely a summary of the result is passed to the client in the end. This separation exists to reduce the server's load by caching the line calculation for later use, thus avoiding unnecessary recalculation.

3.1.3 Metrics

After obtaining the lines' bounding boxes, HandSpy has all the requirements to create an overlay on top of the representation of the strokes, which will represent an interval of interest. What follows is to calculate and aggregate the newly requested data into indicators, called metrics. To obtain a metric entity, we calculate various indicators based on the universe of entries, grouped in the same time windows as the burst interval. Those entities contain data for both burst and global levels for the selected performance indicator. The calculation of every metric is implemented in a dedicated customized class, tailored to suit the necessities of the researchers.

As proof of concept, the first metric to be implemented was the user writing speed. To calculate the value, HandSpy uses the InkML coordinates of neighboring points and the timestamp of when the entry was stored. To obtain the speed metric, the calculator class iterates over the bursts and, then, calculates the aggregations on burst level using the Euclidean distance between two consecutive readings, divided by the difference in the timestamps. The result of the calculation is, then, stored inside a “double” data type value, which represents the speed in mm/ms.

3.2 Client

To allow the interaction between the user and the platform, a new section was added to the front-end component of HandSpy. The new section allows scientists to create a list of metrics that are to be requested from the server. It also allows them to select criteria, that will be later used to determine if it is to be highlighted on the representation of the strokes. The highlight is used to mark important intervals, by paining a transparent overlay over the target area as shown in Figure 2

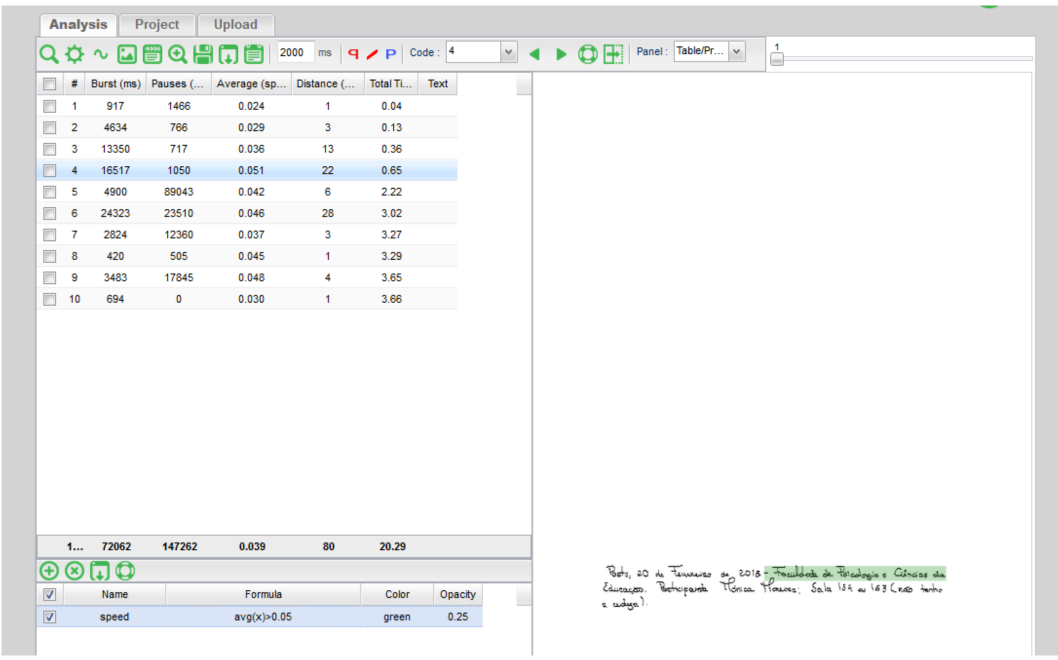

Figure 2 Example of marking the interval with speed above 0.05mm/ms in green.

The criteria for the markers consist of 3 parts: a metric selection; a pseudo programming language formula, which includes the metrics calculated by the server; and overlay options. Specifying the target metrics is as simple as writing its name in the first column. Users have to write the formula, to select the criteria for the points of interest. It has to be satisfied, and the platform will create an overlay over the corresponding stokes.

The rest of the formula represents the aesthetic part of the marking over the lines of the matched intervals. The color column determines the filling of the overlay. It can be filled-in, with either a hexadecimal code or a color name in English. For example, the color can be either “green” or “red” or “ff0000”. Since we need to be able to see the text under the overlay, we also need to select an opacity value that will be applied on the line, unless we want to hide the written input under it. The opacity should be a floating-point number between 0 and 1, with recommended values under 0.25, to keep the text, underneath, visible.

4 Future Work

With the new feature working, HandSpy is fully prepared to receive experimental data from multiple sources. This development can assist the project in the future, but it's insufficient in its current state. The existing implementation is not particularly user friendly, thus breaking the last requirement for a good EMS - Intelligent and adaptive UI [3]. The future work on the platform is going to focus on UI and usability improvements that will make HandSpy usable without training.

One of the changes will be on the metric criteria definition. This segment of the new feature was considered complex by the users. An improvement of the usability is needed so that the scientists will not be required to learn an expression syntax, to be able to use the feature. Besides this, the current information on how the metric aggregations lack proper explanations.

Another improvement discussed with the members of the M-BW project is the inclusion of relative metrics. The current criteria for them can be compared only by using static values. To overturn this, besides the currently available method, we will implement relative metrics, such as quartiles, and percentages of the global values. This change will give them more flexibility in the marking and analysis of points of interest, and will also save the users the time to calculate manually the values to which they wish to compare.

Another change requested that will be useful to the project is the HandSpy front-end separates pauses and bursts so that the users can observe psychophysiological indicators even when no points exist. This change will require the separation of the metrics on both server and client-side.

The current work introduces only some metrics to HandSpy. It is planned for more of them to be implemented in the future. The first ones will be the heart rate and the conductivity of the skin. Those metrics will be collected and stored on the HandSpy side and will be used to aid further the research done by M-BW.

The visual representation of digital ink is also going to receive an overhaul. Some of the discussed changes include the color and opacity, based on the values of the metrics and overlay representation for colorblind users.

The new minor version of HandSpy is developed together with the scientists from the M-BW project. They will validate the usability of the features and assure that the newly implemented features satisfy the needs of the project.

References

1 Denis Alamargot, David Chesnet, Christophe Dansac, and Christine Ros. Eye and pen: A new device for studying reading during writing. *Behavior research methods*, 38:287–99, June 2006. doi:10.3758/BF03192780.

2 Sonia Kandel Eric Guinet. Ductus: a software package for the study of handwriting production. *Behavior Research Methods*, 2010. doi:10.3758/BRM.42.1.326.

3 R Jakobovits, S Soderland, Ricky Taira, and James Brinkley. Requirements of a web-based experiment management system. *Proceedings / AMIA 2000 Annual Symposium. AMIA Symposium*, February 2000.

4 © 2020 M-BW. Mind-Body Interactions in Writing (M-BW) psychophysiological and linguistic synchronous correlates of expressive writing. http://m-bw.up.pt/index.php/m-bw. Accessed: 2020-05-18.

5 Carlos Monteiro and José Paulo Leal. HandSpy – a system to manage experiments on cognitive processes in writing. In Alberto Simões, Ricardo Queirós, and Daniela da Cruz, editors, *1st Symposium on Languages, Applications and Technologies*, volume 21 of *OpenAccess Series in Informatics (OASIcs)*, pages 123–132, Dagstuhl, Germany, 2012. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASIcs.SLATE.2012.123.

6 © 2011 W3C. InkML ink markup language. URL: https://www.w3.org/TR/2011/REC-InkML-20110920/. Accessed: 2020-05-18.