

**SUBJECTBOOK: DATA MANAGEMENT AND
VISUALIZATION METHODS FOR AFFECTIVE
STUDIES**

A Dissertation Presented to
the Faculty of the Department of Computer Science
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Salah Taamneh
December 2016

SUBJECTBOOK: DATA MANAGEMENT AND VISUALIZATION METHODS FOR AFFECTIVE STUDIES

Salah Taamneh

APPROVED:

Advisor Dr. Ioannis Pavlidis
Dept. of Computer Science, UH

Dr. Zhigang Deng
Dept. of Computer Science, UH

Dr. Guoning Chen
Dept. of Computer Science, UH

Dr. David Francis
Dept. of Psychology, UH

Dean, College of Natural Sciences and Mathematics

Acknowledgements

This research was supported by grants from the Texas A&M Transportation Institute. The views expressed in this work do not necessarily reflect those of the funding agencies.

I am very grateful to Dr. Ioannis Pavlidis for his guidance, encouragement, and support throughout the research process. His insightful suggestions and comments guided my research in many ways. The financial support has also been instrumental to pursuing this degree. I also want to thank my other committee members, Dr. Zhigang Deng, Dr. Guoning Chen, and Dr. David Francis for their direction, advice, and feedback.

I would like also to thank Dr. Panagiotis Tsiamyrtzis, Dr. Pradeep Buddharaju, Dr. Dvijesh Shastri, and Dr. Malcolm Dcosta for their help. I also thank each and every other lab member of the Computational Physiological Lab, who have been colleagues and friends and provided a healthy learning environment.

Most of all I have to thank God for the abundantly blessing he has bestowed upon me. I am also very thankful to my loving parents for their advice and unwavering support. Last, but certainly not least, I would like to thank my wife, Amanie, who has supported me through all the ups and downs in all aspects of life, and my son, Kareem, and my daughter, Leen, who have brought great joy into my life.

**SUBJECTBOOK: DATA MANAGEMENT AND
VISUALIZATION METHODS FOR AFFECTIVE
STUDIES**

An Abstract of a Dissertation
Presented to
the Faculty of the Department of Computer Science
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Salah Taamneh
December 2016

Abstract

Managing affective studies is very challenging for two main reasons. First, their life cycles consist of a series of cumbersome and time-consuming activities performed by different people. Second, investigators are increasingly overwhelmed by the size and complexity of data generated by affective studies. Such studies are longitudinal, and feature multimodal data, such as psychometric scores, imaging sequences, and signals from wearable sensors, with the latter streaming continuously for hours. The lack of tools for managing affective studies diminishes researchers' ability to finish collecting, analyzing, and sharing affective data sets within reasonable amounts of time and effort. Moreover, it is difficult to avoid human errors when some tasks, such as data collection and curation, are performed manually. Importantly, some critical tasks, such as quality assurance and exploratory data analysis, can not be performed efficiently unless using appropriate representations for presenting and displaying relationships among collected data. In this work, we introduce SubjectBook, an integrated tool for managing affective studies throughout their life cycles, from designing the experiments to analyzing and sharing the generated data. In this tool, data collection and curation phases have been automated and validated. This enables researchers to have access to their own data in real-time. Additionally, meaningful visual representations of data are provided. Various tools that were proposed to tackle this problem provide visualizations of the original data only; they do not support higher level abstractions. Uniquely, SubjectBook operates at three levels of abstraction, mirroring the stages of quantitative analysis in hypothesis-driven research. The

top level uses a grid visualization to show the study's significant outcomes across subjects. The middle level summarizes, for each subject, context information along with the explanatory and response measurements in a construct reminiscent of an ID card. This enables the analyst to appreciate within subject phenomena. Finally, the bottom level brings together detailed information concerning the inner and outer state of human subjects along with their real-world interactions - a visualization fusion that supports cause and effect reasoning at the experimental session level. SubjectBook was evaluated using three case studies focused on driving behaviors.

Contents

1	Introduction	1
2	Background	8
2.1	Affective Science	8
2.2	Exploratory Data Analysis	10
2.3	Data Visualization	11
2.4	Experimental Design	12
2.4.1	Independent Measures(Parallel group)	12
2.4.2	Repeated Measures (Crossover)	13
2.4.3	Randomized Block Design	13
2.5	Related Work	14
3	Methods	17
3.1	File Organization and Data Types	19
3.2	SubjectBook as a Data Management Tool	24
3.2.1	Defining the experiment	25
3.2.2	Defining the explanatory and response variables	27
3.2.3	Defining the Questionnaires	29
3.2.4	Creating the file structure	29
3.3	Real-time Data Acquisition System	32

3.4	Using SubjectBook for Already Collected Data	47
3.4.1	Experimental Design	47
3.4.2	Data Source	49
3.4.3	Explanatory and Response Variables	51
3.4.4	Covariates	51
3.5	Data Visualization	52
3.5.1	SessionPortrait	52
3.5.2	SubjectPortrait	56
3.5.3	StudyPortrait	58
3.6	Generating Ready-to-analyze Data	59
4	System Architecture	62
4.0.1	Front-end	63
4.0.2	Back-end	64
5	Case Studies	67
5.1	Simulator Study 1	67
5.2	Track study	72
5.3	Simulator study 2	76
5.3.1	Countermeasure Conditions	77
6	Conclusion	80
6.1	Summary of Contributions	80
6.1.1	Automating the Experimental Designs	80
6.1.2	Managing Affective Data Sets	81
6.1.3	Collecting, Curating, and Visualizing Data in Real-time	81
6.1.4	Novel Visualization for Hypothesis-Driven Affective Studies . .	81
6.1.5	Data Transparency	82

List of Figures

1.1	The affective study life cycle.	3
1.2	Levels of visualization in SubjectBook.	6
3.1	File Structure in SubjectBook.	20
3.2	Examples of some types of data in SubjectBook	22
3.3	Metadata for some variables in SubjectBook.	22
3.4	SubjectBook flow diagram.	25
3.5	Creating a new study: experimental unit.	26
3.6	Creating a new study: experimental design.	28
3.7	Creating a new study: explanatory and response variables.	29
3.8	An example that shows how SubjectBook generates the experimen- tal design for a parallel group study with two blocking variables (age and gender), two treatment groups (Group1 and Group2), and 3 ses- sions(S1, S2, and S3).	31
3.9	The visual interface for an empty subject.	33
3.10	SessionPortrait: tab view.	34
3.11	EQA: selecting a study group.	36
3.12	EQA: selecting a blocking variable group.	37
3.13	EQA: selecting a subject from a blocking variable group.	38
3.14	EQA: a list of questionnaires that need to be taken by a particular subject.	39
3.15	EQA: a list of filled questionnaires by a particular subject.	40

3.16	EQA: a view that shows how a subject will be prevented from taking a secession-level questionnaire until he/she finishes performing the corresponding session.	41
3.17	OTACS: recording the perinasal perspiration signal during the first session for Subject31.	43
3.18	OTACS: recording the perinasal perspiration signal during the second session for Subject31.	44
3.19	EQA: An example of a questionnaire.	45
3.20	EQA: a view that shows the progress after completing a session-level questionnaire.	45
3.21	The visual interface for Subject31 after performing all sessions.	46
3.22	Defining an existing study: experimental design.	48
3.23	Defining an existing study: study location.	50
3.24	A browse dialog box for the user's Google Driver.	50
3.25	SessionPortrait: tab view.	53
3.26	Synchronization between a physiological signal and observational streams.	54
3.27	SessionPortrait: multi-session view.	55
3.28	A SubjectPortrait example.	57
3.29	An example that shows how the response variable is visualized in SubjectBook.	58
3.30	A SubjectPortrait example.	59
3.31	The StudyPortrait of a study on driving behaviors under distractions.	60
3.32	A visual interface that shows the list of the studies the user has access to, and also shows the export option that is available for each study.	61
3.33	The structure of the file that SubjectBook generates for users to help them perform statistical analysis.	61
4.1	System Architecture.	63
4.2	A simplified picture of the local database in SubjectBook.	66

5.1	The driving simulator experimental setup.	69
5.2	SubjectPortraits for the simulator study 1.	71
5.3	The SessionPortrait for subject T029 in session LD4.	71
5.4	The track study experimental setup.	73
5.5	SubjectPortraits for the track study.	74
5.6	This figure shows how SessionPortrait enables researchers to compare between different variables in the track study.	75
5.7	A view that shows how SessionPortrait drew our attention to a hidden pattern in our data.	76
5.8	The SessionPortrait for a particular subject in simulator study 2. . .	79

List of Tables

3.1	The list of data types in SubjectBook	21
3.2	The standard list of recognized file extensions in SubjectBook	23
5.1	The list of session in the track study	72
5.2	The list of sessions in the simulator study 2	77

Chapter 1

Introduction

Affective science is the study of emotions. It combines the perspectives of many disciplines, including neuroscience, psychology, and affective computing [7]. Affective studies focus on measuring and analyzing the psychophysiological state of subjects under various stressors. Such studies illuminate the hidden factors influencing operator performance in human-machine interactions and other critical tasks. Examples include monitoring of driver performance under distractions [24], and how loss of proprioception affects sympathetic and motor responses in novice and experienced laparoscopic surgeons [14].

In recent years, there has been growing interest in collecting and analyzing multimodal affective datasets [10, 11, 17, 21]. This is due to the advent of ubiquitous devices capable of measuring different aspects and factors of human affect [15]. Managing such studies throughout their life cycles is very challenging for two main reasons. First, they consist of a series of cumbersome and time-consuming activities

performed by different people. Second, investigators are increasingly overwhelmed by the size and complexity of data generated by such studies, which typically encompass biographic, psychometric, physiological, observational, and performance sets. The longitudinal nature of affective studies, where measurements are recorded on individual subjects over a period of time [20], adds even more complexity to the collected data. This leads to data explosion, complicating analysis further.

Figure 1.1 shows the main phases in the affective study life cycle. Researchers first design and implement the experiments needed to test their hypotheses. Next, data are collected, cleaned, and passed to data analysts. At this stage, Exploratory Data Analysis (EDA) is performed first to summarize the main characteristics of the collected data, and multiple statistical tests are then run accordingly. Finally, all experimental data along with results are shared with others.

The lack of appropriate tools for helping researchers throughout all phases of their studies diminishes their ability to finish collecting, analyzing, and sharing affective data sets within reasonable amounts of time and effort. Furthermore, some human errors are hard to be avoided when some tasks, such as data collection and curation, are performed manually. Importantly, some critical tasks, such as quality assurance and exploratory data analysis, can not be performed efficiently unless using appropriate representations for presenting and displaying relationships among collected data.

Experimental design is critical for the validity of the results. Bad experimental designs lead to false conclusions and waste considerable time and resources. Assisting researchers with designing their experiments via automating such process can

greatly contribute to reduce human errors during this stage. For example, automating the process of randomly allocating the experimental units among the treatment groups can create homogeneous treatment groups, ensuring unbiased allocation of experimental units to conditions. This can not be guaranteed if this task is carried out manually by the experimenter..

Most of the problems in affective studies happen during the data collection and curation phases. This is mainly because these tasks are time-consuming and cumbersome. Moreover, people who are assigned to perform these tasks rarely receive enough training. One of the most critical tasks during these phases is data synchronization. Failing to align all channels of data properly leads to wrong conclusions as many channels in affective studies carry complementary information for other channels. For example, the perinasal perspiration signal provides information about the intensity of sympathetic responses; it does not show if responses are negative or positive. In order to get this information, the researcher needs to consult facial expression data.

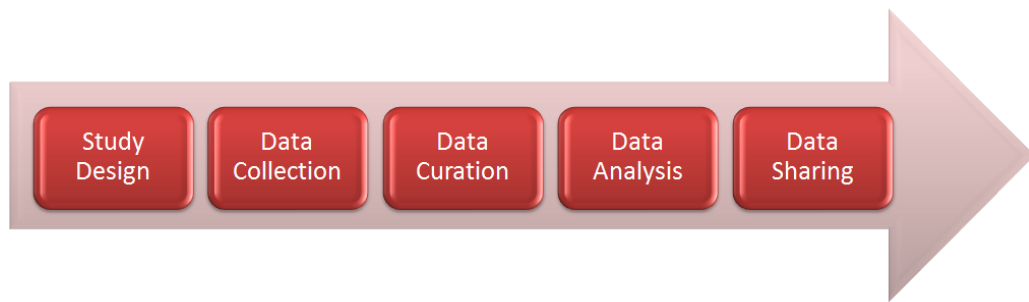


Figure 1.1: The affective study life cycle.

After being collected and cleaned, data need to be put together and organized at a single location. Moreover, an efficient mechanism to access and query such data needs

to be provided. Relational Database Management Systems (RDSMS) were found to be unsuitable to contain the multidimensional or hierarchical structure that found in scientific data sets[1]. Consequently, several software systems have been developed to handle unstructured, mostly heterogeneous data [5]. These systems are called Scientific Data Management Systems (SDMS), and their primary goal is to structure the heterogeneous data in an organized way, and make it faster and easier to access, query, and analyze such data. Data in such systems is represented in machine-independent formats that enable researchers to manipulate it using general-purpose tools and programming languages. However, very little work has been done on providing visual representations of such data.

Data visualization is indispensable in coping with complexity in affective studies. Some critical tasks, such as quality assurance and exploratory data analysis, can not be performed efficiently unless using appropriate representations for presenting and displaying relationships among collected data. Appropriate visualization facilitates insights by focusing the analysts' attention on important trends and patterns. Effective visualization can also reduce the possibility of relying on bad or corrupted data as it helps find errors.

There are a number of tools for visualizing multimodal affective datasets. Such tools focus on helping researchers analyze their data by providing visual representations of such data. Typically, these tools provide detailed visualizations of the original data. Little work has been done to communicate information at higher levels of abstraction, integral to hypothesis-driven research. Moreover, these tools are not designed to help researchers manage their data.

In my thesis, I introduce SubjectBook, an integrated tool for managing affective studies throughout their life cycle, from designing the experiments to analyzing and sharing the generated data [22]. SubjectBook assists researchers with generating their experimental designs, thus eliminating human errors. It also automates the data collection and curation phases allowing collecting accurate data in real-time. Moreover, a convenient way for ubiquitously accessing users’ data has been provided. Such data is presented to users in novel forms that support intuition and assist with quality control. In addition to that, ready-to-analyze formats for the same data, which are necessary to perform statistical data analysis, are available upon request. These formats are ready to be imported to any statistical analysis software such as R and MATLAB.

The visualizations provided in this tool were designed not only to bring together different forms of data, but also to support hypothesis-driven reasoning. Toward this end, a visualization scheme with three levels of abstraction is supported. In the first level, the *SessionPortrait*, the covariates are presented at the top, followed by the time-registered explanatory and response variables. This visualization fusion aims to bring together possible cause, effect, and context. In the second level, we extend the work from [9] to help analysts appreciate within subject phenomena. Context information along with the explanatory and response measurements per subject are summarized in a construct reminiscent of an ID card - the *SubjectPortrait*. Finally, analysts can communicate to SubjectBook statistical tests on the study’s variables, for producing the *StudyPortrait* - a grid visualization of the study’s significant outcomes across subjects. All three visualization levels are automatically created, and

a pyramidal software architecture (Fig. 1.2) is used to navigate between them.

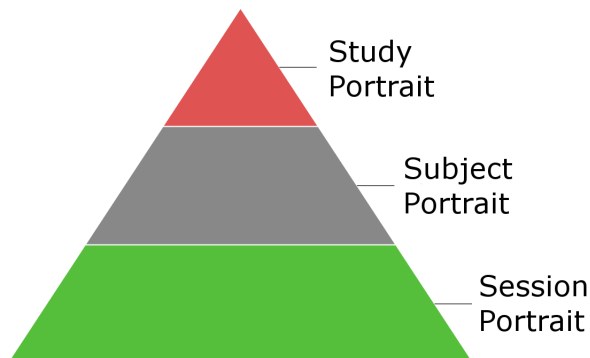


Figure 1.2: Levels of visualization in SubjectBook.

During the data collection phase, these visualizations are primarily used to perform quality assurance, and also to provide information about missing or unavailable data. Once the data set is complete, researchers use these visualizations to perform exploratory and affirmative data analysis in order to help them select the right follow-up statistical analysis. Finally, researchers have the option to share such visualizations with other people, providing them with an easy-to-comprehend version of the data.

If done manually, collecting and curating data may introduce human errors that are hard to discover in later phases. In order to avoid this problem, we propose a real-time data acquisition system that automatically collects, curates, and stores data sets in real-time. The proposed system, consists of three main components: management and visualization tool (i.e., SubjectBook), physiological signal recorder, and electronic questionnaire application. These components interact with each other to manage and visualize data in real-time, and at the same time cut human errors to the bare minimum.

SubjectBook is available for use over the Internet. In order not to impose any restriction on the amount of storage that can be used, SubjectBook was designed to read data from users' personal cloud storage. This approach is also necessary for the real-time data acquisition system to be able to operate as such storage is accessible from anywhere. Although we only support Google Drive cloud storage, other personal cloud storage services can easily be added.

The effectiveness of SubjectBook in visualizing and analyzing affective studies is demonstrated using three case studies. Each case represents a study that has been visualized and/or analyzed using SubjectBook. In the first study, the effects of cognitive, emotional, sensorimotor, and mixed stressors on driver arousal and performance with respect to baseline was studied in a simulation experiment [13]. In the second study, the same distracting stressors were studied but this time subjects were asked to drive a real car. Finally, the third study investigates the impact of biofeedback intervention on reducing drivers' level of distraction in a simulation experiment.

Chapter 2

Background

2.1 Affective Science

Affective science is the scientific study that focuses on understanding emotional and affective processes. It involves research on mood, emotion, emotion regulation, decision making, and the underlying physiology and neuroscience of emotions. The importance of this science stems from the fact that many phenomena, such as cognitive processing and social behavior, cannot be explained without studying the affective factors (i.e., motives, attitudes, moods, and emotions)[18].

Emotion or affective response is defined as *an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism* [19]. Measuring such responses is very challenging because they are

multidimensional in nature. Ideally, the comprehensive measure of emotion includes:

- Sympathetics responses that start in the amygdala and precipitates peripherally to adrenergic receptors (elevates heart rate, respiration rate, and blood pressure)
- Sympathetics responses that start in the amygdala and precipitates peripherally to cholinergic receptors (activates sweat glands on palmar, plantar, perinasal and other areas)
- Facial and vocal expressions.
- Subjective feedback (i.e., questionnaires).
- Observational data (e.g., videos, and audios)
- Performance data.

The heterogeneity of the collected data in affective studies imposes tremendous challenges for exploring and analyzing such data. In the absence of a specialized tool, researchers often need to use multiple software, synchronize data sets manually, and memorize values and time points from one data set while looking for the corresponding values in another one. Another challenge is the longitudinal nature of affective studies, where measurements are recorded on individual subjects over a period of time. This leads to data explosion, complicating analysis further.

2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach that aims to display data in a way that facilitates comprehension of casual relationships among data sets, and helps researchers summarize the main characteristics of the their data. Analysts typically use EDA to

- Perform quality assurance
- Detect outliers and abnormalities;
- Select the right statistical models
- Determine relationships among explanatory variables and between explanatory and response variables.
- Gain an insight into how covariates affect the relationship between explanatory and response variables.

EDA techniques fall under two categories: non-graphical, and graphical techniques. The former focuses on providing descriptive statistics about data such as: measures of central tendency(e.g., mean), measures of spread (e.g., standard deviation), type of distribution (i.e., exponential distribution), and the existence of outliers. The latter focus on creating visual representations of data that focus analysts' attention on important trends and patterns. These visualizations vary from simple charts to complex graphs.

2.3 Data Visualization

People are increasingly overwhelmed by the size and complexity of their data. Lacking the ability to deal with such data leads to wasted time and money, and lost scientific opportunities. The key factor in discovering the hidden knowledge in data is selecting the right methods to present it. Creating visual representation of data is by far the best way to communicate information clearly and efficiently. Effective visualization helps researchers identify interesting patterns on the fly. It can also help determine relationships among different data sets.

According to [23], there are some characteristics all effective graphical displays have in common:

- They show the data
- They encourage viewers to think about the represented data rather than the graphical design or anything else.
- They avoid misrepresenting the data.
- They use small spaces to present large data sets, and make them coherent.
- They help make comparison between different pieces of data.
- They provide several levels of abstraction and allow users to navigate between these levels.
- They have a clear purpose.

- They are integrated with the statistical descriptions of the data sets they represent.

2.4 Experimental Design

In hypothesis-driven research, researchers first set a hypothesis and then design the experiment to test that hypothesis. Experiments are designed so that changes to one item cause something else to vary. These changing items are called variables. There are three different types of variables in any experiment: Independent, dependent, and control. The independent variable (i.e., explanatory) is what the designer changes or manipulates in order to understand its impact on other variables. The dependent variable (i.e., response) is what researchers focus their observations on in order to understand how it responds to the changes in independent variable. The control variables is the one whose values are conditioned out in examining the relationship between the independent and dependent variable.

Experimental design refers to how experimental units (i.e., participants) are allocated among treatment groups or conditions. There are three types of experimental designs that are commonly used:

2.4.1 Independent Measures(Parallel group)

Different experimental units are allocated for each treatment group. In other words, each group of conditions includes a different group of participants. The allocation

should be done randomly so that all participants have the same chance to be assigned to each condition.

2.4.2 Repeated Measures (Crossover)

All experimental units will be used in each condition in the experiment. In other words, each unique condition include the same group of experimental units.

2.4.2.1 Counterbalancing

Counterbalancing is a method that is used to eliminate order effects in repeated measure experiments. In this methods, participants are separated into groups, each group performs tasks or receives treatments in a different order.

2.4.3 Randomized Block Design

Randomized block design, sometimes called matched pairs, is used when there are differences between subjects that might affect performance. In this design, experimental units are divided into homogeneous blocks before they are randomly allocated among treatment groups.

2.5 Related Work

We present an integrated tool for managing affective studies. To the best of our knowledge, there is no tool available for helping researchers throughout each phase of the affective study life cycle. However, there are some tools that researchers can use to help them perform particular tasks in such studies. Specifically, these tools focus on assisting researchers with storing and analyzing their data sets.

Several scientific data management systems were developed to handle unstructured, multidimensional, mostly heterogeneous scientific data sets [16][3][6]. The work style in these systems is to encapsulate the heterogeneous data sets and provide a platform-independent way to store and retrieve such data. Data is presented in a format that can be manipulated using programming languages. Although these systems are helpful, they are not being used extensively to store affective data. This is due to the fact that no effective visualizations are provided for data. In highly multimodal data, visualization is necessary to facilitate comprehension of casual relationships among various data sets. Visualization also helps perform quality assurance. Inability to perform this task efficiently may result in relying on bad or corrupted data, and thus drawing false conclusions.

Over the years, several tools have been proposed to visualize multimodal affective datasets [4, 8, 12, 25]. To the best of our knowledge, no tool abstracts data/results at the subject and study levels, mirroring the scientific method. Also, data sharing among researchers is either not supported by these tools or is not straightforward.

Kim et al. [8] proposed BEDA, a tool for synchronizing, visualizing, and navigating behavioral and physiological data in longitudinal studies. It creates a visual summary to help identify interesting patterns within a single experimental session or across different experimental sessions. BEDA resembles to some degree the Session-Portrait level of SubjectBook. However, no visual abstractions are provided at the subject or study levels; thus, BEDA cannot qualitatively support hypothesis-driven research that traverses these levels quantitatively. Also, it does not accommodate biographic and psychometric information.

ChronoViz [4] is another tool that supports annotation, visualization, navigation, and analysis of multimodal time-coded data. Similar to ChronoViz, ELAN [25] provides visualization and synchronization but only between audio and video data. Neither ChronoViz nor ELAN, however, support record keeping and analysis of multiple sessions in longitudinal studies.

Oefinger et al. [12] developed a web-based tool for visualizing and manipulating multi-scale physiological data. This tool, however, does not support any other data type. Moreover, no high-level visualization is provided.

There are many general purpose visualization tools available on the market. A good example of these tools is Tablue. Although they are good at showing the relationships between heterogeneous data sets, they do not provide abstract visual representations of such data. Data abstraction is of paramount importance in human studies.

SubjectBook is a data management and visualization tool. It helps users organize their data, and provides them with a convenient way to ubiquitously access and share such data. Moreover, SubjectBook provides novel visual representations of users' data. In addition to that, this tool offers two unique features. First, it assist researchers with designing their experimental design. Second, it automates the process of collecting and curating data.

Chapter 3

Methods

SubjectBook aims to help researchers at each phase of the affective study life cycle. This includes: study design, data collection, data curation, data analysis, and data sharing. However, if researchers have complete data sets they can still use SubjectBook to conduct exploratory data analysis and share such data with others. This requires data to be structured in a way that SubjectBook can understand. Using SubjectBook as an exploratory tool provides users with novel visual representations of their data. These visualizations are provided at three levels of abstraction, each reveals a certain amount of details. They were carefully designed to help summarize the main characteristics of users' data, which is necessary to select the right follow-up statistical analysis. On the other hand, using SubjectBook to manage a study throughout its life cycle enables experimenters to:

- Eliminate all human errors associated with experimental designs

- Automatically collect and curate data
- Perform quality control in a timely manner
- Check for missing data
- Find and fix problems as they appear
- Make adjustments to the experimental design based on the initial results
- Perform exploratory data analysis
- Share their data
- Get the data in a ready-to-analyze format

SubjectBook was designed either to be used as a stand-alone application, or to be part of an integrated real-time data acquisition system. Using subjectBook as a stand-alone application requires collecting and curating data manually. This is considered to be the main source of many problems in affective data. In order to overcome this, we developed an integrated real-time data acquisition system that collects and curates data automatically, and also makes these data available in real-time. This system consists of three main components: management and visualization tool (i.e., SubjectBook), physiological signal recorder, and electronic questionnaire application. It also can scale up to include as many signal producers as needed.

Our tool can read data residing on the server or on users' personal cloud storages. Although Google Drive is currently the only supported personal cloud storage service, other personal cloud storage services can easily be added to the system. Allowing

users to store their data in the cloud enables them to use as much storage as they need.

In this chapter of the thesis, we first describe how data are organized in a file structure, and how SubjectBook treats such data. Then we go over the steps that are necessary to manage and visualize an affective study. After that, we demonstrate how the real-time data acquisition system works to collect data in real-time. The novel visualizations offered by this tool are then explained. Finally, we show how SubjectBook generates ready-to-analyze data.

3.1 File Organization and Data Types

We adopted an intuitive file structure to store users' data. In order for SubjectBook to manage and visualize studies, this structure needs to be carefully followed. The way data is organized inside that file structure determines to a great extent how the visualizations will look. Figure 3.1 shows a file structure for an affective study with n participants, each of whom went through m sessions. All data that belong to a particulate study need to be organized inside a single folder. Data stored in that folder should be organized as follows:

- A separate folder should be dedicated for each subject.
- Inside the subject folder, the session-independent data, such as trend psychometrics and biographic information, need to be stored
- The data that is collected during each experimental session a subject went

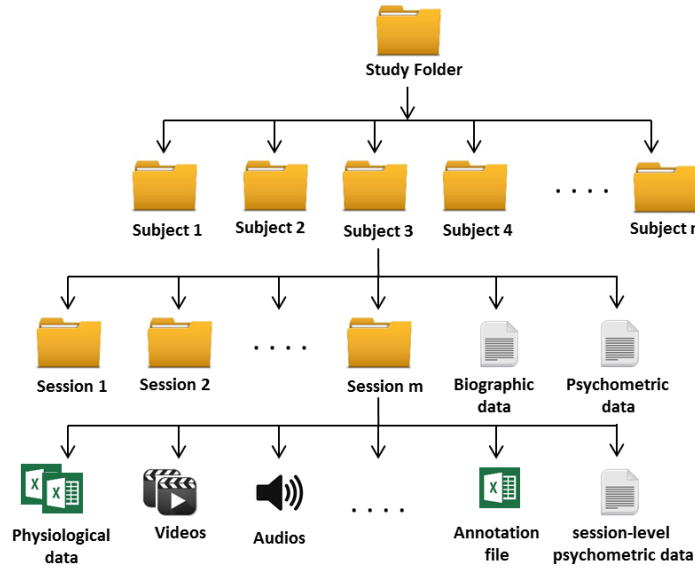


Figure 3.1: File Structure in SubjectBook.

through should be stored in a separate folder inside the subject folder.

When SubjectBook is instructed to visualize a study, it goes and scans the file structure of that study and reads the data it encompasses. Table 3.1 provides a list of all data types found in affective studies. Figure 3.2 also provides an example of each of the data types presented in table 3.1.

We use the file extension to determine the type of data that is being read. SubjectBook initially comes with a standard list of file extensions. Each of these extensions

Table 3.1: The list of data types in SubjectBook

Type	Format
Time-registered Data	Excel or csv
Biographic Data	Text file with (key:value) format
Descriptive Data	Plain Text file
Psychometric Data	Text file with (key:value) format
Annotation Data	Excel file
Rank Data	Excel file (visualized using bar chart)
Video	avi or mp4
Audio	wav or mp3
Images	JPEG, GIF, BMP, or PNG

has a data type assigned to it. We decide how to deal with each file based on these two pieces of information (i.e., file extension, and data type). Please note that these extensions are only used to help SubjectBook distinguish between different variables. The format of each file should be one of these mentioned in Table 3.1. The standard list of file extensions and the data type associated with each extension is presented in table 3.2.

As Figure 3.3 shows, users can change the way SubjectBook treats each type of data. This includes the unit of measurement, sampling rate, column number and row number from which to start reading data, minimum and maximum values to display, and whether to use logarithmic scale or not. This information also serves as metadata schemes to help researchers describe their data to others. Users also have control over which psychometrics the system should read and how it should be visualized. Users need to specify the name, minimum score, and maximum score for each psychometric test. The minimum and maximum values determine how psychometric scores are going to be visualized.

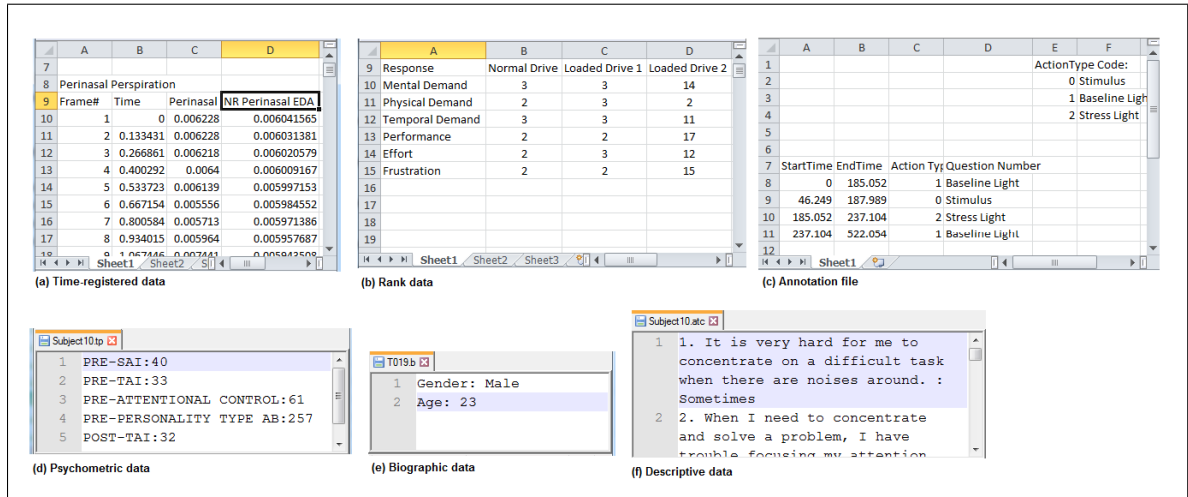


Figure 3.2: Examples of some types of data in SubjectBook

To give users even more flexibility, SubjectBook allows them to add new variables to the list presented in table 3.2. This is very helpful when users have more than one study, and want to treat data in each study differently.

Data Description	Data Extension	Data Type	Y title	Sampling Rate	Rows	Col	Ymin	Ymax	Log. Scale	
Perinasal Perspiration	pp	Time	perinasal	1	9	2				Update
Nasal Perspiration	np	Time	nasal	8	9	2				Update
Palm EDA	peda	Time	Palm EDA	32	9	2				Update
Finger EDA	feda	Time	eda	8	9	2				Update
Plantar EDA	pleda	Time	eda	8	9	2				Update

Figure 3.3: Metadata for some variables in SubjectBook.

Table 3.2: The standard list of recognized file extensions in SubjectBook

Name	Type	Extension
Perinasal Perspiration	Time-registered data	pp
Nasal Perspiration	Time-registered data	np
Palm EDA	Time-registered data	peda
Finger EDA	Time-registered data	feda
Plantar EDA	Time-registered data	peda
Heart Rate Variability	Time-registered data	hrv
Heart Rate - Chest	Time-registered data	hrc
Heart Rate - Wrist	Time-registered data	hrw
Breathing Rate - Chest	Time-registered data	brc
Breathing Rate - Thermistor	Time-registered data	brt
Breathing Rate - Thermal	Time-registered	brtl
Facial Expression	video data	facs
Motion - Core	Time-registered data	mc
Motion - Hand	Time-registered data	mh
Motion - Feet	Time-registered data	mf
Eye Track	Time-registered	et
Face Video	Video data	avi/mp4
Stimuli	Annotation	stm
Trend Psychometrics	Psychometric Data	tb
State Psychometrics	Psychometric Data	sp
Biographic	Text data	b
Rank Data	Rank Data	bar
Images	images	JPEG/GIF/BMP/PNG

3.2 SubjectBook as a Data Management Tool

As demonstrated by Figure 3.4, the first step in creating a new study is to define the experiment. Experimenters need first to provide information about the experimental units and experimental design. Information about experimental units include number of participants and number of blocking variables (e.g., age), while information about the experimental design include the type of experiment, number of sessions, type of each session, and the order in which these sessions should be performed. After that, data that are expected to be produced by the study need to be defined.

The information provided by experimenters will be used by SubjectBook to create an empty file structure that will be used to host their data. The created file structure will also determine how subjects are going to be allocated among treatment groups, and the order in which tasks should be performed within each group of units. This will contribute to reduce all human errors that may be introduced if such task is to be performed manually.

The data that have been defined during the process of creating the study will be visualized and delivered to users as soon as they become available (i.e., stored in its designated file structure). This approach helps researchers perform quality assurance in a timely manner, and also generate reports about the current status of their experiments. Moreover, such early access to visualized data can greatly facilitate finding problems and avoiding them early in the study life cycle. Another important feature SubjectBook offers is the ability to transform the currently available data into a ready-to-analyze format that can be used to perform in-depth analysis, a task that

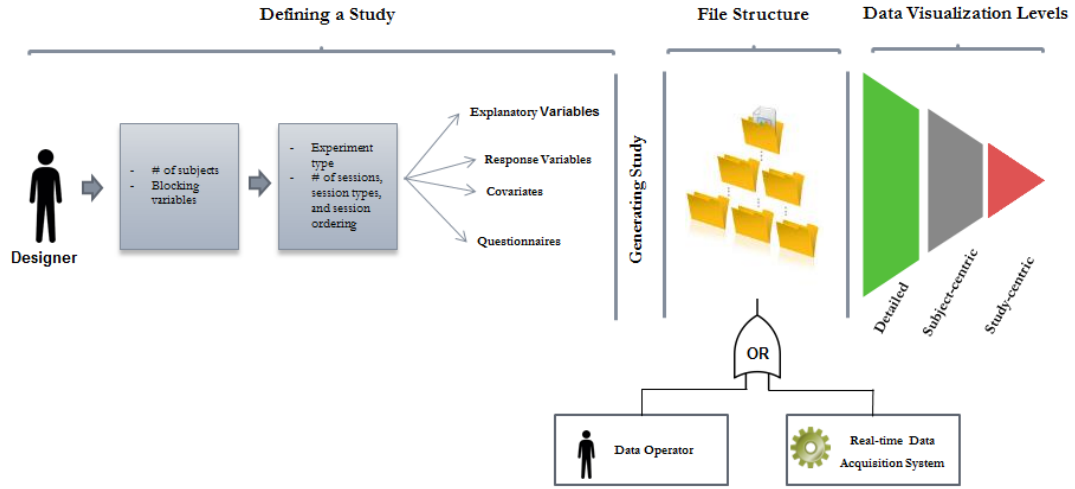


Figure 3.4: SubjectBook flow diagram.

is considered to be very cumbersome if done manually. Finally, the provided visual representations and ready-to-analyze formats can be shared with other collaborators over the web, supporting joint data analysis.

Next, we explain in detail all the steps that are required to initiate a new study using SubjectBook:

3.2.1 Defining the experiment

In this step, the study name and the number of recruited subjects need to be determined (see Figure 3.5). Next, the experimenter needs to define the experimental design. This includes: the grouping variable, number of experiments, type of each experiment, and number of sessions (i.e., intervention) in each experiment.

There are several types of experimental designs used in affective studies, but the

The diagram illustrates a five-step process for creating a new study. The steps are: EXPERIMENTAL UNITS, EXPERIMENTAL DESIGN, EXPLANATORY VARIABLES, RESPONSE VARIABLES, and COVARIATES. The first step, EXPERIMENTAL UNITS, is highlighted with a blue square. Below the steps is a 'STUDY STRUCTURE' form. The form contains two input fields: '* Study name:' with the value 'My study' and '* Number of subjects:' with the value '48'. A blue 'Next' button is located at the bottom of the form.

Figure 3.5: Creating a new study: experimental unit.

most common ones are: parallel groups, crossover, and randomized block design. SubjectBook was designed to support all these designs.

A session is any task a human subject was asked to perform in a specific period of time. We distinguish between 4 types of sessions:

- **Baseline:** Totally neutral condition(e.g., subjects sitting and doing nothing).
- **Cross-baseline:** Basic experimental activity without additional treatments.
- **Cross-Intervention:** Same experimental activity as in cross-baseline but with treatments superimposed. Subjects' responses in such a session are compared against their responses in the cross-baseline session.
- **Intra-Intervention:** Within the same session, subjects' responses during the treatment phase are compared with subjects' responses before the treatment

phase.

It is very important to select the correct type for each session as this largely affects the provided visualizations. In addition to session type, users need also to provide the following information for each session:

- Session name and its acronym.
- Arm or group: if the experiment is parallel group, this option determines to which group this session belongs to. If the user selects "All" that means this session belongs to all groups. This option will be disabled if the experiment is crossover.
- Order: this attribute determines if the order of performing this session is fixed or not. If fixed, subjects need to perform the task in the order it appears in the list. Otherwise the task's order will be randomized.

3.2.2 Defining the explanatory and response variables

After defining the experimental design, experimenters need to define the data that are going to be generated from the experiment. In this step, the key and backup explanatory variables need to be determined. Also, users can choose as many secondary explanatory variables as they want. The same thing also apply for the response variable (see Figure 3.7).

EXPERIMENTAL
UNITS
EXPERIMENTAL
DESIGN
EXPLANATORY
VARIABLES
RESPONSE
VARIABLES
COVARIATES

EXPERIMENTAL DESIGN

Blocking Variables

How many experiments do you have?

Select Variables ▾

Age
Gender

1

(1) Experiment type : Parallel Group ▾

Number of groups : 2

Group name : Group 1

Group name : Group 2

Number of sessions : 3

Session Name	Acronym	Session Type	Arms	Order
Baseline	BL	Cross Base ▾	All ▾	Random ▾
Treatment 1	T1	Cross Inter ▾	All ▾	Random ▾
Treatment 2	T2	Cross Inter ▾	All ▾	Random ▾

Previous
Next

Figure 3.6: Creating a new study: experimental design.

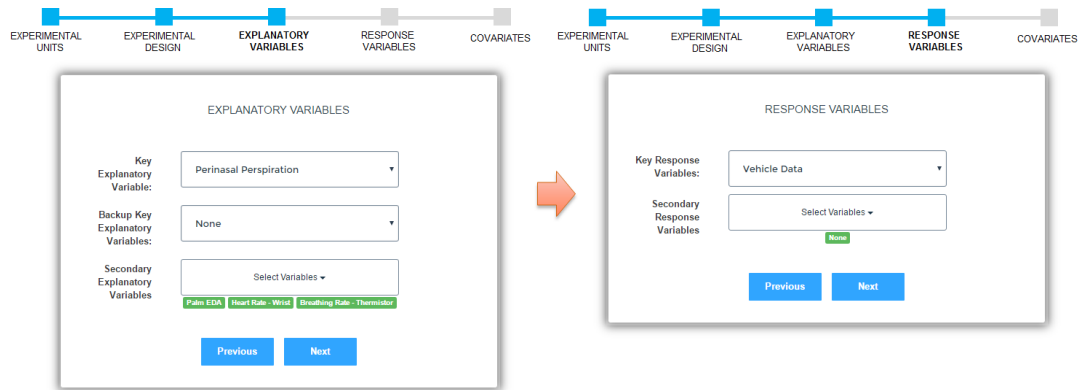


Figure 3.7: Creating a new study: explanatory and response variables.

3.2.3 Defining the Questionnaires

Finally, users need to identify the questionnaires that are supposed to be taken by subjects. We distinguish between three types of questionnaires:

- **Subject-level Pre-study Questionnaires:** The questionnaires that are taken by subjects before performing any session.
- **Subject-level Post-study Questionnaires:** The questionnaires that are taken by subjects after performing all sessions.
- **Session-level Questionnaires:** These are session-dependent questionnaires and aim to get subjects' feedback about the task they performed.

3.2.4 Creating the file structure

Using the information entered by experimenters, SubjectBook creates empty file structures in users' personal storages that are ready to host their data. SubjectBook

starts by randomly allocating the experimental units across the treatment groups. To do that, SubjectBook needs first to check if the study has blocking variables or not. Blocking is the arrangement of experimental groups or units into blocks in order to reduce sources of variability within each group. Examples of blocking variables are age (old/young), and gender (male/female). Based on the selected blocking variables, subjects will automatically be split into homogeneous groups. If no blocking variable was selected, all subjects will be organized in a single group. Next, treatments will be allocated across the groups of subjects. If the experiment is crossover, subjects in all groups will be assigned the same treatments, but in random order within each group. If the experiment is parallel groups, each group will be split into a number of subgroups equal to the number of treatment groups in the experiment, and each of these subgroups will be assigned a different treatment group. Subjects in each subgroup will receive the treatments in random order. Figure 3.6 provides an example of creating a new parallel group study in SubjectBook. This study has two blocking variables (age and gender), two treatment groups (Group1 and Group2), and 3 sessions. Figure 3.8 shows how SubjectBook uses this information to generate the experimental design. Please note that both groups have the same sessions. Therefore, we expect that some or all sessions will be treated differently among groups.

After finishing developing the experimental design, SubjectBook will start creating the file structure that will be used to store users' data. Initially, this file structure will be empty (i.e., no data), and only contains information needed by experimenters. This information is saved in three separate files that are found inside the root folder

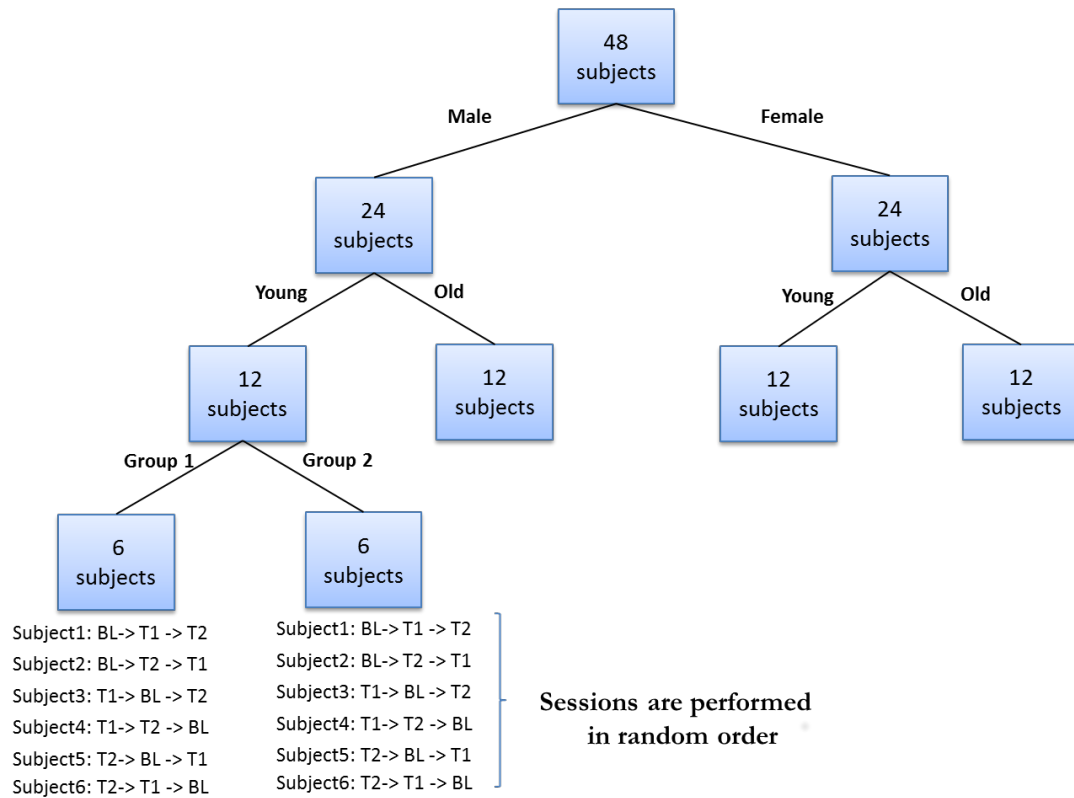


Figure 3.8: An example that shows how SubjectBook generates the experimental design for a parallel group study with two blocking variables (age and gender), two treatment groups (Group1 and Group2), and 3 sessions(S1, S2, and S3).

(i.e., the study folder):

- **Blocks File (BF)**: the file that contains information about the experimental design.
- **Sessions File(SF)**: the file that contains information about the type of each session, and the questioners that need to be taken by subjects.
- **Progress File(PF)**: the file that is used to keep track of the progress that has been made in the study.

These files are necessary for the real-time data acquisition system that will be explained in the next section. The generated study will have as many subjects as specified by the experimenter. Inside each subject folder, an empty folder for each session the subject will go through will be found. The name of each session will be preceded by a number that determine the order of performing this session. The selected key and secondary variables for both explanatory and response variables will be saved in the database. SubjectBook will be looking for these data whenever a request is made to scan the study. Data will be inserted into the study either manually by data operators or via the real-time acquisition system. Figure 3.9 demonstrates how the visual interface for an empty study will look like in SubjectBook.

3.3 Real-time Data Acquisition System

SubjectBook was designed to collaborate with other software components to collect and visualize the data in a real-time manner. This real-time system consists of three

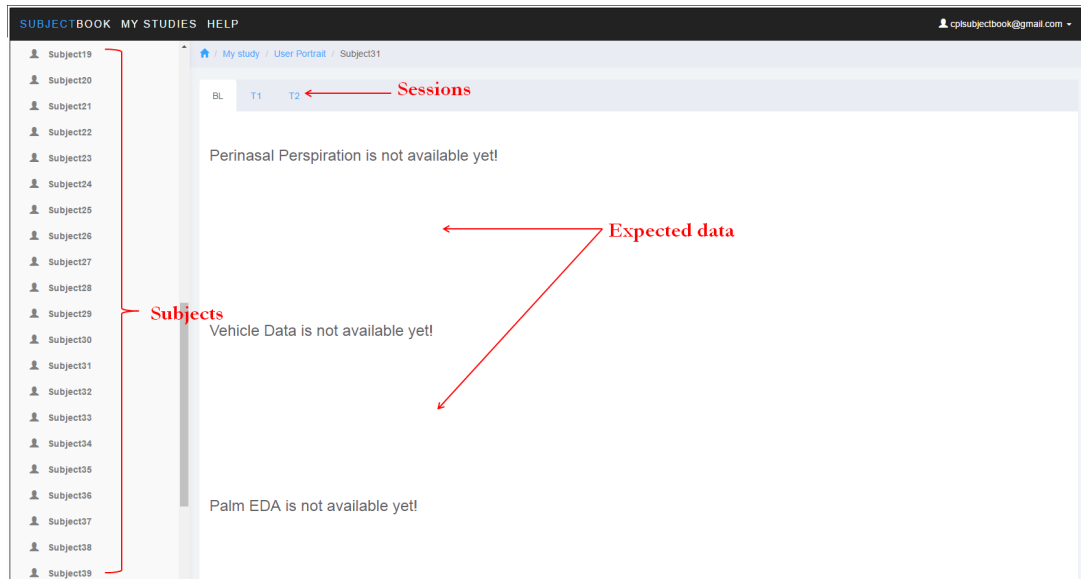


Figure 3.9: The visual interface for an empty subject.

main components.

- **SubjectBook:** it will be used by the experimenter to develop the experimental design and create an empty file structure to store the expected data. It will also be used to visualize, analyze, and share such data.
- **Electronic Questionnaire Application (EQA):** a tablet-based application that was developed at the Computational Physiology Lab (CPL) at the University of Houston (UH). This application will be used by experimenters to start new subjects, and will also be used by users to fill in the required questionnaires that were specified by experimenters.
- **Open Thermal Action Coding System (OTACS):** an application that was also developed at CPL Lab. This application uses the thermal imaging

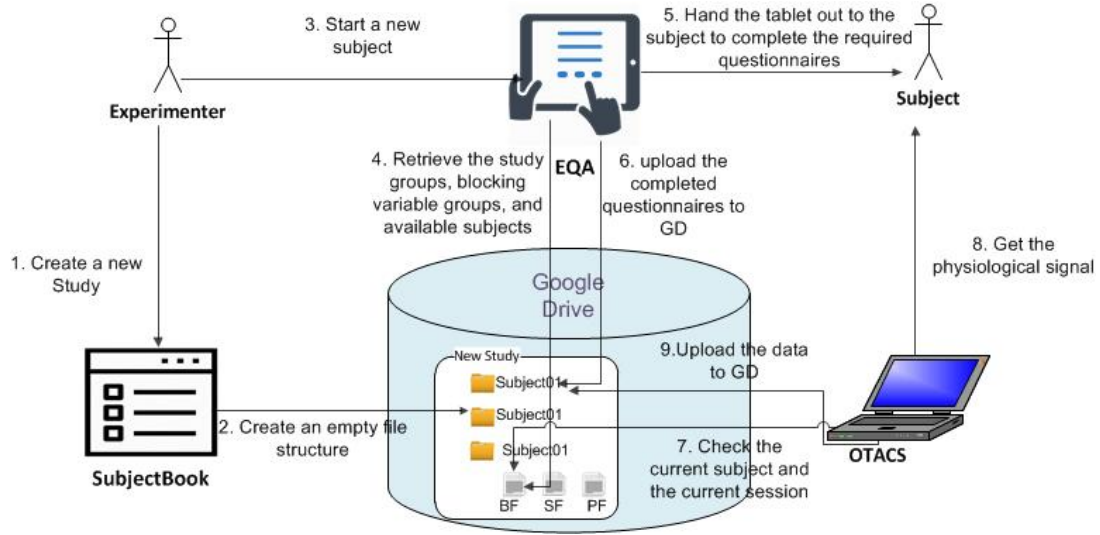


Figure 3.10: SessionPortrait: tab view.

data coming from a thermal camera attached to it to extract the perinasal perspiration signal for a subject. OTACS coordinates with EQA to make sure that questionnaires are taken in their proper order, and also to make sure that sessions are performed in the order that was specified by SubjectBook.

Figure 3.10 demonstrate how these components interact with each other to collect the data in a real-time fashion. First, the experimenter initiates a new study using subjectBook. This, as explained earlier, will create an empty file structure for the new study (see Figure 3.9), and will also create three files that regulate how all components interact with each other to complete the job. When a new subject comes in, the experimenter uses EQA to identify the treatment group for that subject. Next, the experimenter needs to select the blocking group that the new subject belongs to. This is determined based on the characteristics of each subject (e.g., age or

gender). Consequently, a list of available subjects inside the selected group will appear, allowing the experimenter to select a subject. Figures from 3.11 to 3.14 demonstrate how EQA is used to select a young-male subject (i.e., Subject31) from study group "Group 1". The number of blocking groups, and number of subjects inside each group is found in the BF file. After that, the experimenter will hand the tablet out to the subject to fill in the pre-study questionnaires. Figure 3.15 shows that Subject31 finished all the pre-study questionnaires, while Figure 3.16 shows that he/she can not fill in the session-level questionnaire for the baseline session until he/she performs the required task in that session. EQA uses the information in the SF file to know which questionnaires users need to take at each stage. The completed questionnaires will be uploaded to Google Drive, and the PF file will be updated to show that a subject is ready to start his/her first session.

OTACS will then be used by the experimenter to extract the perinasal perspiration signal from the subject while performing a specific task. The information in the PF file will be used to determine what session the current subject should perform at the moment. Figure 3.17 demonstrate how OTACS collects the perinasal perspiration data for Subject31 during the baseline session. The list of sessions on the left side will be disabled until the subject finishes filling in all the pre-study questionnaires. The generated data during each session will be uploaded to Google Drive in real-time, and the PF will be updated to show that one session has been completed. Data generated by OTACS include the perinasal perspiration signal, facial video, region of interest video, and operating theater video. Next, EQA will be used to check if there are session-level questionnaires that need to be taken by the subject or

iPad 2:05 PM 36%

< Back Subject home screen

Select a study

My Study

test

Select a study group X

24/24

Group 1

24/24

Group 2

Select a blocking variable group

Select a subject

Start

Figure 3.11: EQA: selecting a study group.

iPad 2:06 PM 36%

< Back Subject home screen

Select a study

My Study

test

Select a study group X

24/24

Group 1

Select a blocking variable group X

6/6 6/6 6/6 6/6

Old-Male

Old-Female

Young-Male

Young-Female

Select a subject

Start

Figure 3.12: EQA: selecting a blocking variable group.

37

iPad 2:07 PM 36%

< Back Subject home screen

Select a study

My Study

test

Select a study group X

24/24

Group 1

Select a blocking variable group X

6/6

Young-Male

Select a subject

Subject31

Subject32

Start

Figure 3.13: EQA: selecting a subject from a blocking variable group.

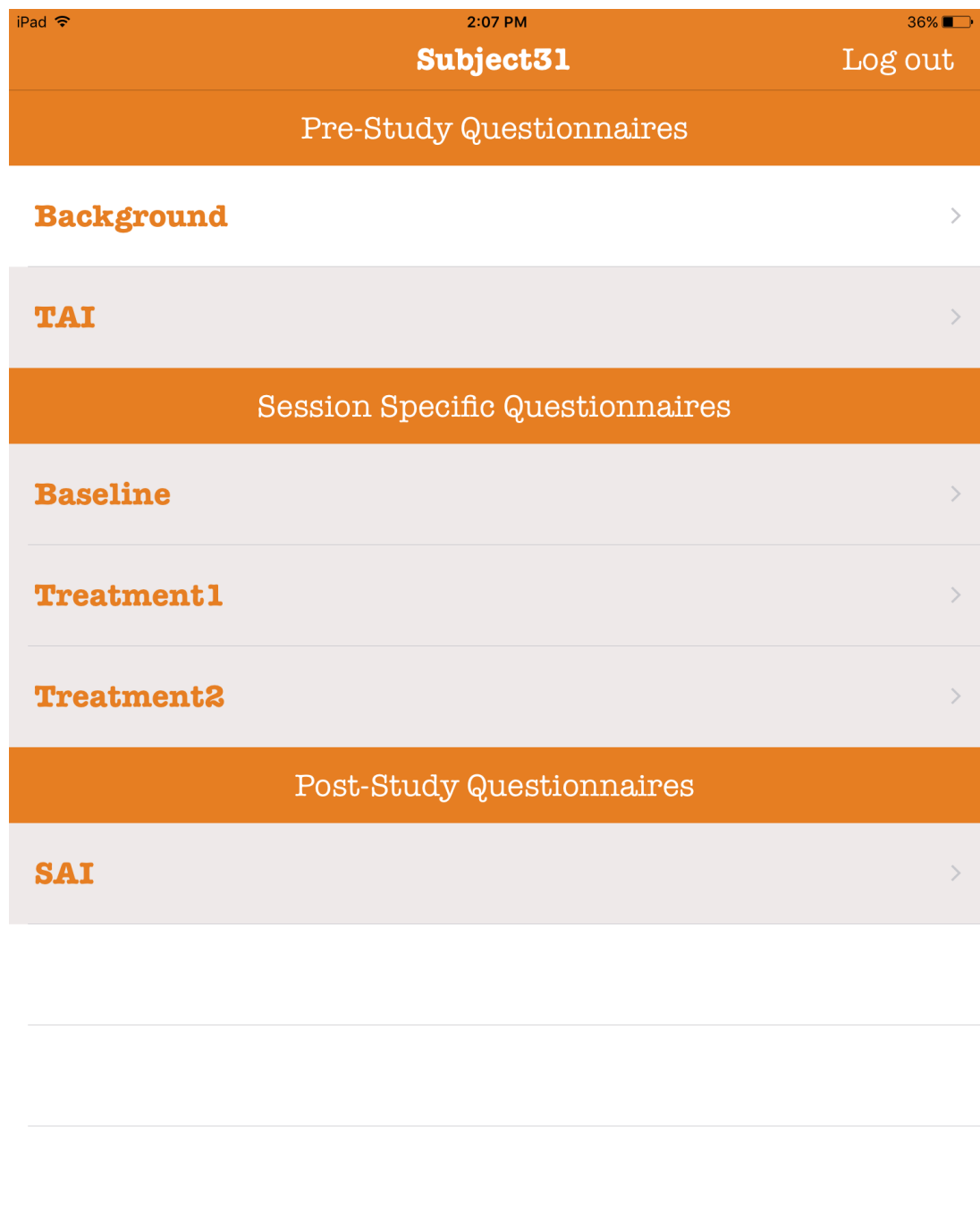


Figure 3.14: EQA: a list of questionnaires that need to be taken by a particular subject.



iPad	
2:13 PM	
100%	
Subject31	
Log out	
Pre-Study Questionnaires	
Background	
TAI	
Session Specific Questionnaires	
Baseline	
Treatment1	
Treatment2	
Post-Study Questionnaires	
SAI	

Figure 3.15: EQA: a list of filled questionnaires by a particular subject.

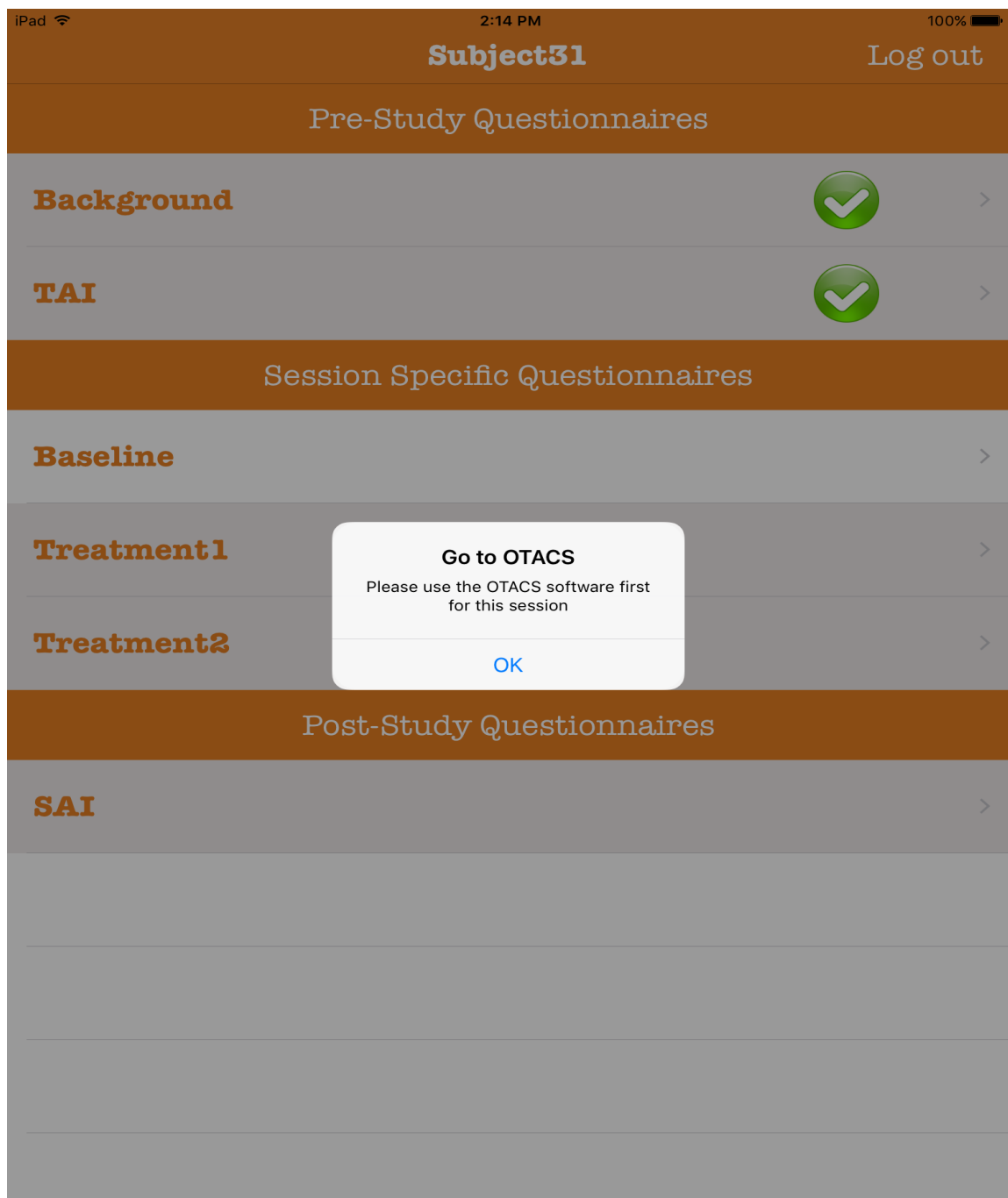


Figure 3.16: EQA: a view that shows how a subject will be prevented from taking a secession-level questionnaire until he/she finishes performing the corresponding session.

not. If not, permission will be granted for OTACS to repeat the same thing for the next session. Figure 3.19 provides an example of a questionnaire that needs to be taken by Subject31 after finishing the baseline session, while Figure 3.20 shows the progress made by the same subject after taking that questionnaire. OTACS can now collect the data for the next session (see Figure 3.18). Once all sessions are done, PF needs to be updated to show that one subject has been completed.

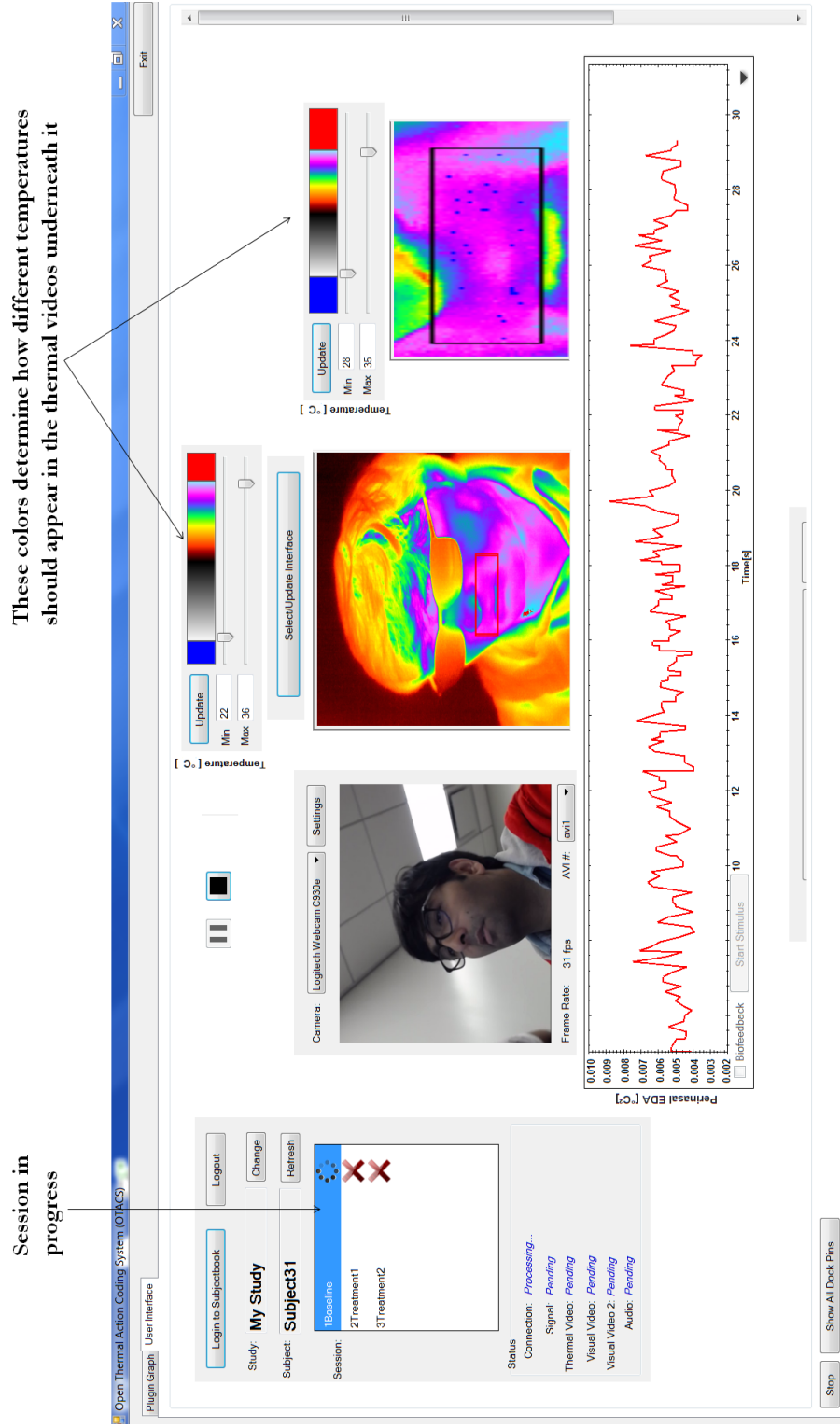


Figure 3.17: OTACS: recording the perinasal perspiration signal during the first session for Subject31.

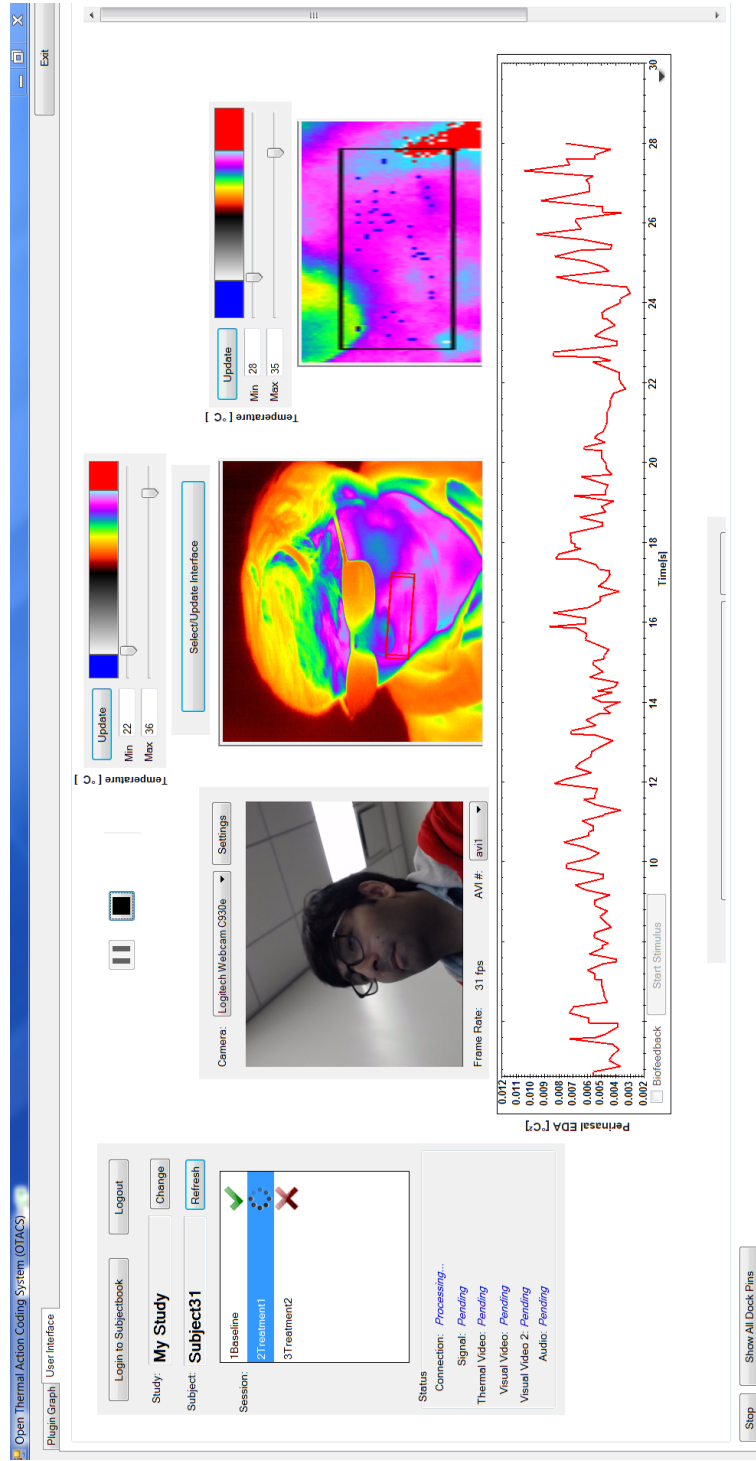


Figure 3.18: OTACS: recording the perinasal perspiration signal during the second session for Subject31.

Mental Demand: How mentally demanding was the task?
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
 Very Low Very High

Physical Demand: How physically demanding was the task?
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
 Very Low Very High

Temporal Demand: How hurried or rushed was the pace of the task?
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
 Very Low Very High

Performance: How successful were you in accomplishing what you were asked to do?
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
 Perfect Failure

Figure 3.19: EQA: An example of a questionnaire.

Subject31 Log out

Pre-Study Questionnaires

Background ✓

TAI ✓

Session Specific Questionnaires

Baseline ✓

Treatment1

Treatment2

Post-Study Questionnaires

SAI

Figure 3.20: EQA: a view that shows the progress after completing a session-level questionnaire.

Data that are uploaded to Google Drive either by OTACS or IQA do not need any further processing. Such data are ready to be visualized and analyzed using SubjectBook. Figure 3.21 shows how Subject31 will appear in SubjectBook after finishing all sessions.

Although only the perinasal perspiration signal can be collected in real-time using our system, the scenario that was just explained allows any signal producer software to coordinate with the existing components to collect and upload data in real-time without having to perform any intermediate processing. The only required thing is to store data in proper formats.

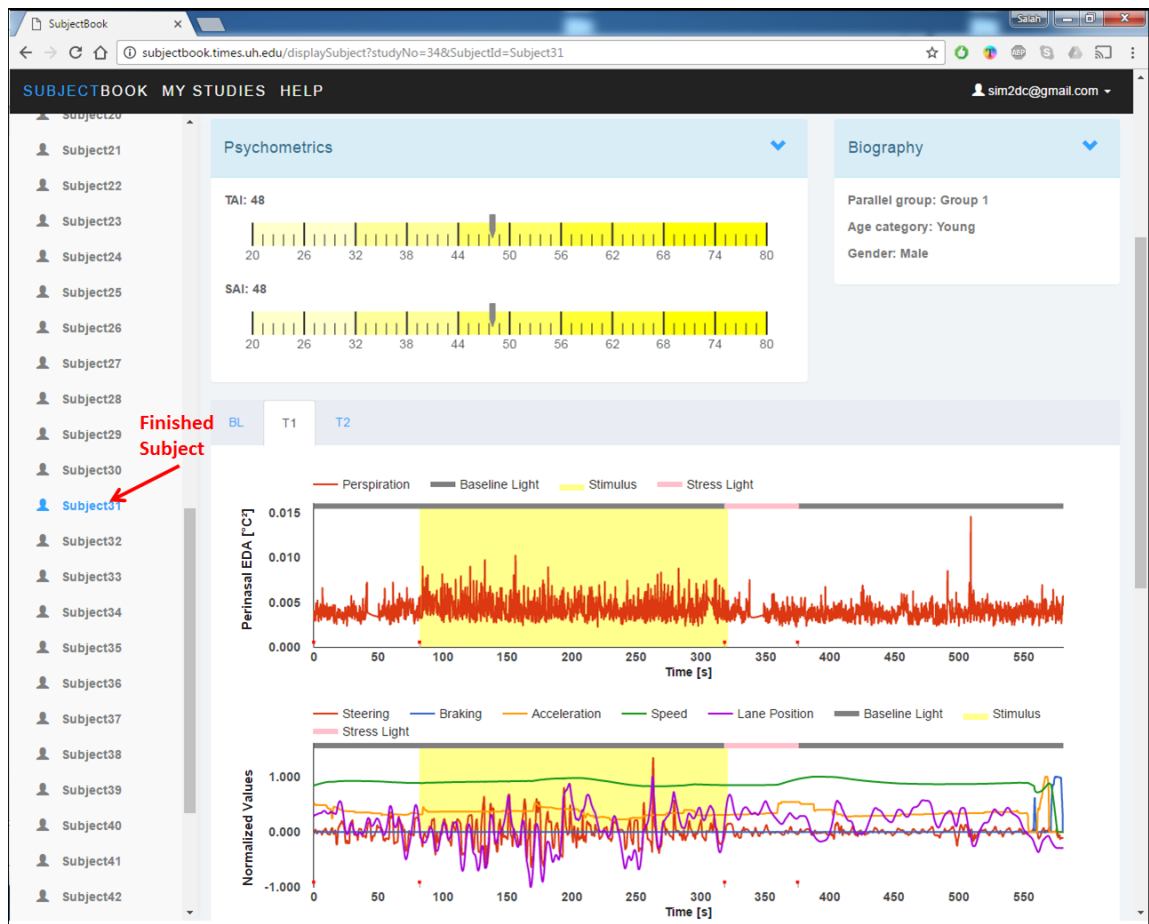


Figure 3.21: The visual interface for Subject31 after performing all sessions.

3.4 Using SubjectBook for Already Collected Data

Researchers who have finished collecting their data are also allowed to visualize and share their data using SubjectBook. In this case, data are assumed to be completed and stored in a file structure similar to the file structure explained earlier. Users then need to go through the following steps that are necessary for SubjectBook to decide how data is going to be visualized.

3.4.1 Experimental Design

Users can skip this step as SubjectBook can get the required information from the file structure. However, it is highly recommended to provide such information especially if users want to have control over how data should be visualized.

The experimental design can be communicated to SubjectBook in two different ways. First, via a GUI provided by SubjectBook (see Figure 3.22). Second, using a text file saved in the root folder of the study. This file should have ".descriptor" file extension. Using the graphical interface, experimenters need to enter the number of sessions, and number of treatment arms in their experiments. If the experiment is a crossover, this value should be one. If there are some sessions that were not distributed evenly between subjects this value should reflect that. For example, Figure 3.22 indicates that the study has two arms of experiments. The first arm includes the first three sessions in addition to session FDN, while the second arm includes the first three sessions in addition to session FDL.

EXPERIMENTAL DESIGN

How many unique sessions does your study have?

5

How many treatment arms does your study have?

2

Session Name	Acronym	Session Type	Treatment Arms
Normal Drive	ND	BaseLine ▼	Crossover ▼
Loaded Drive 1	LD1	Intervention ▼	Crossover ▼
Loaded Drive 2	LD2	Intervention ▼	Crossover ▼
Failure Drive-No Loading	FDN	Intervention ▼	Arm1 ▼
Failure Drive-Loading	FDL	Intervention ▼	Arm2 ▼

Next

Figure 3.22: Defining an existing study: experimental design.

The following information needs to be provided for each session

- The name of the session.
- The preferred acronym to use for visualization purposes.
- The session type. This is very important as it will greatly determine how the second level of visualization will look (i.e., SubjectPortrait)
- The treatment arm. If all subjects went through the session this should be Crossover. Otherwise, users should specify the treatment arm.

Please note that the order in which these sessions are entered in will be followed when the second and third level of visualization are generated.

3.4.2 Data Source

As Figure 3.23 shows, the experimenter needs to provide the following information in this step:

- Source type: whether data is stored on the user's Google Drive or on the local server. The second option is only enabled for the system administrator.
- Study name.
- Study location: if the study is on Google Drive, a browse dialog box will appear asking the user to select the desired folder from his/her own Google Drive(see Figure 3.24). Similarly, if the source type was the local server, a dialog will appear but this time to browse the server.

At this stage, users can also specify whether to make the study available for the public or not, and whether to create higher levels of visualization or not. Users may also instruct the system to quickly create the study without going over the last three steps. In this case, the explanatory and response variables will be unknown, and thus only the first level of visualization will be offered by SubjectBook (i.e., SessionPortrait).

If a user chooses not to provide information about the experimental design, SubjectBook will scan the selected study and come back to the user with a list of found sessions. Next, the user will be asked to specify which of them is the baseline session. This piece of information is necessary for all levels of visualization. Failing to choose

1 2 3 4 5
EXPERIMENTAL DESIGN DATA SOURCE EXPLANATORY VARIABLES RESPONSE VARIABLE COVARIATES

DATA SOURCE

* Source type: Google Drive

* Study name:

* URL:

☐ Public study ☐ Generate Portrait ☐ Quick

Previous Next

Figure 3.23: Defining an existing study: study location.

Select a file

Google Drive

test

Name	Owner	Last modified ↓
test	me	Oct 4, 2016
Video Processed Study	me	Sep 30, 2016
TechSmith	me	Sep 28, 2016
Quality Control Archive	me	Sep 20, 2016
Sim2	me	Sep 15, 2016
Reporocessed Subjects	me	Sep 2, 2016
SIM2 Pilot	me	Jun 24, 2016

Select Cancel

Figure 3.24: A browse dialog box for the user's Google Driver.

the right baseline session will affect the whole visualization as we will see in later sections.

Because SubjectBook will be checking users' data several times and coming back to them with questions about it, we use the WebSocket Protocol. This protocol provides full-duplex communication channel between the client and the servers, enabling real-time data flow from and to the server.

3.4.3 Explanatory and Response Variables

During this step, users will be asked to select their key and backup explanatory and response variable. The back up variable will only be used if the key is missing. The explanatory and response variables will be selected from the list of time-registered variables provided by the system or created by users. Once a variable is selected, the system will ask users to specify which column of data should be used in that file. This is necessary if the file has multiple columns.

3.4.4 Covariates

Finally, users need to specify the questionnaires that were taken by subject before or after sessions. all variables with psychometric type will be shown to users to select from. Additionally, users will be asked to determine where the biographic information can be found. This information will be used to create the SubjectPortriat.

3.5 Data Visualization

SubjectBook was designed to visualize longitudinal affective studies where numerous human subjects are involved. In this type of study, baseline measurements are initially collected for subjects under normal conditions. These measurements are then compared with corresponding measurements taken under interventional conditions. The analysts' aim is to identify any significant changes in the response variables and associate them with any significant changes in the explanatory variables. For example, a significant increase in sympathetic arousal (explanatory variable) due to driving distractions may negatively influence driving performance, resulting in a significant increase in lane deviations (response variable).

The analysts' investigation takes place either at the intra-subject level or at the inter-subject level, using data from the experimental sessions. Accordingly, we designed a three-level visualization scheme: SessionPortrait, SubjectPortrait, and StudyPortrait.

3.5.1 SessionPortrait

SessionPortrait is the grassroots visualization level, where data collected during each experimental session, including surveys, signals, and videos, are combined into a visual report. As illustrated in Fig. 3.25, SubjectBook puts session-specific data under a tab dedicated to that session. Specifically, the session's temporal signals are synchronized and presented in the form of stacked line charts. This arrangement brings to the fore relationships among physiological (explanatory) and performance



Figure 3.25: SessionPortrait: tab view.

(response) channels in a particular session. Analysts can switch from one session to the other by simply clicking the appropriate tab.

Session-independent data (i.e., covariates), such as trend psychometrics and biographic information, are placed above the session tabs, providing constant access to such information. Trend psychometric scores are visualized within the corresponding scales. Biographic information is presented in a separate panel, next to trend psychometrics.

One important feature of SessionPortrait is stimulus annotation. This annotation boils down to a colored region on the chart, marking the duration of an intervention.

For example, the two colored regions in Fig. 3.25 mark two similar events - texting distractions. Different events are assigned different colors.

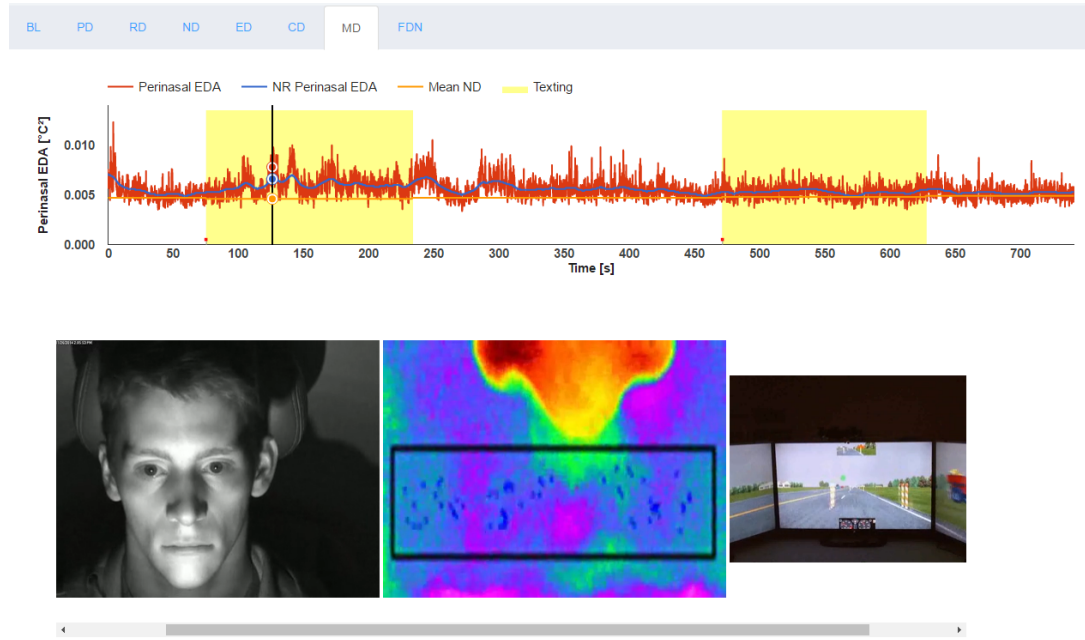


Figure 3.26: Synchronization between a physiological signal and observational streams.

In order to establish association between the subjects' internal and external states, synchronization between physiological signals and observational data (e.g., facial videos) is needed. SessionPortrait provides a convenient way for analysts to perform such a task. By clicking at any data point in a line chart, the associated videos will appear, and start playing from that point onward in sync with a scanning play-line in the line chart (Fig. 3.26).

The tab view supports qualitative analysis within each experimental session. However, comparing measurements across different sessions is also desirable. To

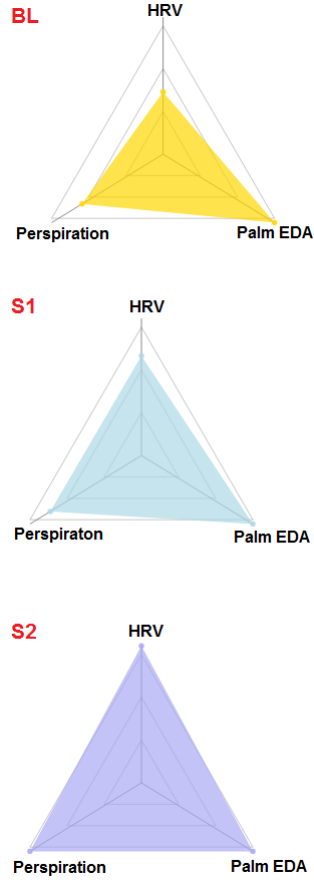


Figure 3.27: SessionPortrait: multi-session view.

address this problem, we provide a view that brings together corresponding measurements from all sessions. In this view, each session is represented by a radar graph, where each axis represents one variable and takes values between 0 and 1. For each variable, we compute the mean values in all sessions and normalize them with respect to the largest mean obtained. The range of values each variable assumes determines the extent of coloring. Figure 3.27 gives an example of a multi-session view. It is easily noticeable that the highest heart rate variability (HRV) was reached

in session S2. This figure also shows that there were no significant differences in palmar electrodermal activity (EDA) among sessions.

3.5.2 SubjectPortrait

SubjectPortrait is the second level of visualization; it enables analysts to identify intra-subject patterns, by condensing subject data into its essence, showcased in a single figure. For each subject, summary information for the explanatory variable is juxtaposed to summary information for the response variable, framed by context (covariates), in a construct reminiscent of an ID card. Specifically, the key explanatory variable in each session is represented by a horizontal colored bar, with the proportion of red color reflecting the proportion of time during which the subject's level of affect rose above her/his mean baseline. Figure 3.28 shows how to visualize the explanatory variable for a particular session.

Next to the affect bars, there are disks with different filling levels, representing the key response variable (typically performance) in each session. Filling levels are: empty, lightly-filled, heavily-filled, and fully-filled, corresponding to the first, second, third, and fourth quartile with respect to the proportion of time during which the subject's performance dipped below her/his mean baseline. Figure 3.29 provides an example of how to visualize the response variable for a study with 5 subjects, each performed three sessions in addition to the baseline.

Finally, subjective feedback collected from subjects at the end of each session is represented with color-coded semicircles that appear between the affect bars and the

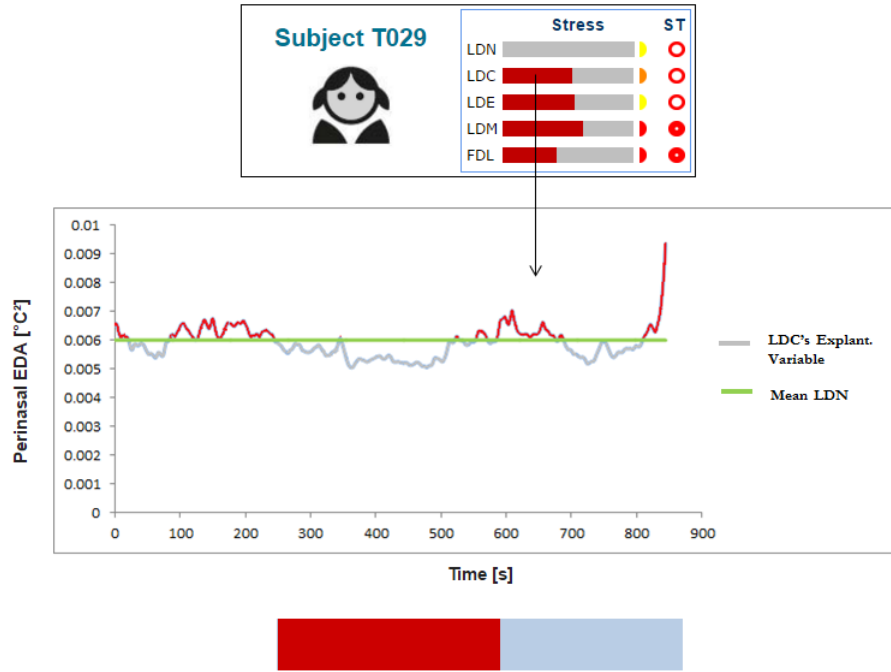
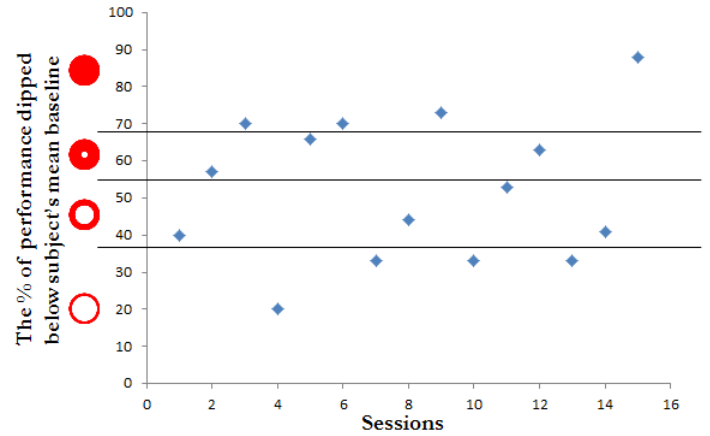


Figure 3.28: A SubjectPortrait example.

performance disks. Quartiles are used again to encode the intensity of the perceived affect, mapping to four color shades: green, yellow, orange, and red, corresponding to the first, second, third, and fourth quartile with respect to the maximum range of the psychometric scale.

Figure 3.30 provides a SubjectPortrait example. This abstraction illuminates how the subject's affect changes in each session, and how his performance appears to be influenced by it. As the sessions progress from top to bottom, it is apparent that the interventions have a stronger catholic effect on the subject's physiological state. In sync, his performance seems to deteriorate, as indicated by the increasingly filled disks. His perceived affect, represented by the increasingly redder semicircles, is in agreement with the objective (physiological) measurements represented by the

Subject	Session	The % of performance dipped below subject's mean baseline
Subject1	S1	40
	S2	57
	S3	70
Subject2	S1	20
	S2	66
	S3	70
Subject3	S1	33
	S2	44
	S3	73
Subject4	S1	33
	S2	53
	S3	63
Subject5	S1	33
	S2	41
	S3	88



The 25th percentile: 36.5%

The 50th percentile: 53%

The 75th percentile: 68%

Figure 3.29: An example that shows how the response variable is visualized in SubjectBook.

bars. The subject is male, as indicated by the contextual icon. Analysts can click on the subject's icon to bring up his SessionPortrait, thus accessing the grassroots visualization level. This is often needed to perform quality control, confirm patterns, or disambiguate unexpected observations.

3.5.3 StudyPortrait

StudyPortrait is the third level of visualization, aiming to communicate the study's outcome in a nutshell. It provides a high-level visual summary, illustrating whether the explanatory and response variables during each phase of an interventional session are significantly different with respect to baseline or not.

For the subject set, we compute the distributions of paired differences between:

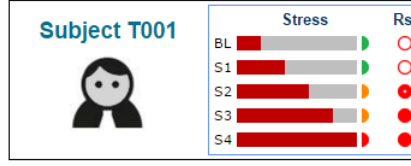


Figure 3.30: A SubjectPortrait example.

1. The mean of the key explanatory variable E in S_j (interventional session) and S_ϕ (baseline session) in phase P_i ,

$$\Delta \overline{E}(\cdot, j, P_i) = \overline{E}(\cdot, S_j, P_i) - \overline{E}(\cdot, S_\phi, P_i). \quad (3.1)$$

2. The mean of the key response variable in S_j and S_ϕ using an equation similar to (3.1).

Paired t-tests are then used to determine whether the mean of each distribution is different than 0 or not at a significance level chosen by the analyst. If the difference is significant, the corresponding phase is represented by a red rectangle. Otherwise, the rectangle is left blank. Figure 3.31 shows the outcome of a study featuring three interventional sessions.

3.6 Generating Ready-to-analyze Data

Researchers often spend considerable amount of time converting their data sets into ready-to-analyze formats. Moreover, sharing such data with remote collaborators can sometimes be difficult due to the size of that data. SubjectBook provides a convenient way for researchers to get their data in a ready-to-analyze format. As

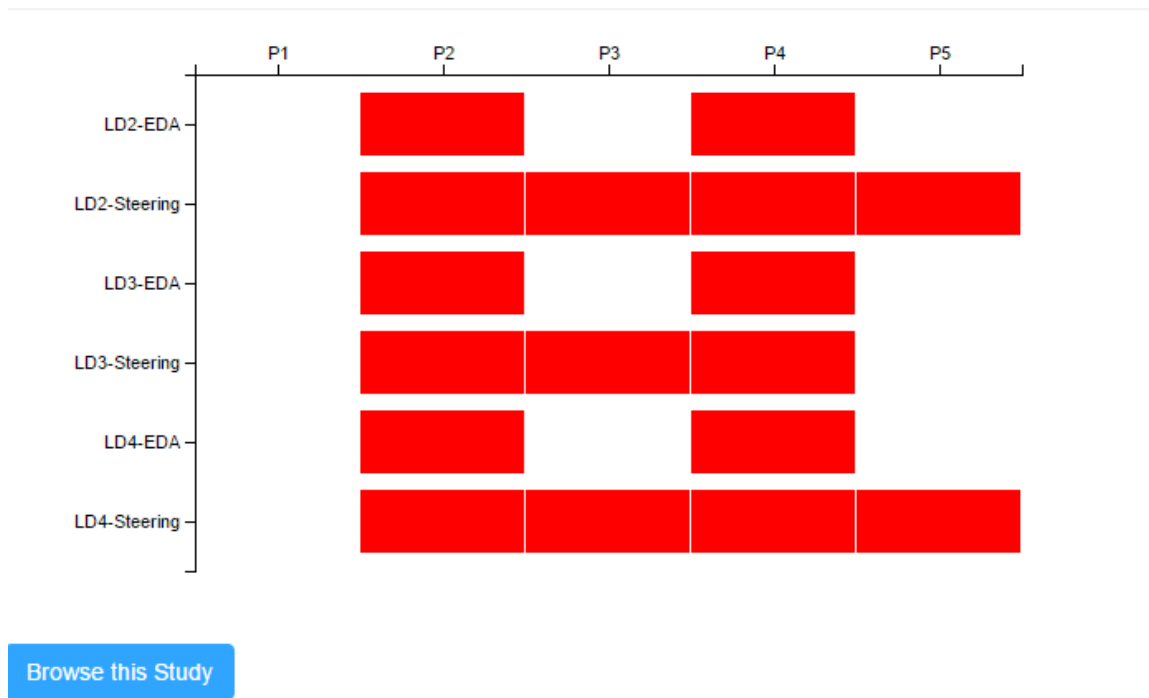


Figure 3.31: The StudyPortrait of a study on driving behaviors under distractions.

shown by Figure 3.32, next to each study, an export button appears, giving users the ability to download all data sets into their local machines. Figure 3.33 Demonstrate how data are organized in the file provided by SubjectBook. Data will be saved in a CSV file. Each row represents a data point recorded for a subject while performing a specific task. Each applied intervention is represented by a separate column. The values that such columns take at each point of time are either the name of the applied intervention or "NA" if no intervention was applied at that point. The biographic information will also be included in this file if they are available.

<input type="checkbox"/>	Study Name	Owner	Edit	Export	Delete
<input type="checkbox"/>	Simultor Study 1	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	Simulator Study	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	Track Study	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	Children Study	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	Simulator Study	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	Simulator Study	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	sa	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	SIM2 Pilot	sim2dc@gmail.com	<input type="button" value="Edit"/>	Export	<input type="button" value="delete"/>
<input type="checkbox"/>	SIM2 Dryrun July18	cplsubjectbook@gmail.com		Export	
<input type="checkbox"/>	Sim2	sim2dc@gmail.com	<input type="button" value="Edit"/>	Export	<input type="button" value="delete"/>

Figure 3.32: A visual interface that shows the list of the studies the user has access to, and also shows the export option that is available for each study.

Subject Name	Task	Time [s]	Data Type 1	Data Type n	Intervention 1	...	Intervention m
Subject01	LD1	0	70		15	Cognitive Load		NA
Subject01	LD1	0.134	72		17	Cognitive Load		NA
Subject05	FD	33	81		12	NA		Emotional Load
Subject05	FD	33.134	81		18	NA		Emotional Load

Figure 3.33: The structure of the file that SubjectBook generates for users to help them perform statistical analysis.

Chapter 4

System Architecture

SubjectBook is available for use over the Internet. It requires users to log in to the system using their Google's login information. However, If a user decides to share his /her data with the public, he/she can generate a direct link to that data, and people no longer need to be logged in to see such data. Oauth2 protocol is used to gain access to users' Google Drives. When a user logs in to the system for the first time, a message appears telling the user that SubjectBook is asking for off line access to your Google Drive. Once that privilege is granted, users can start visualizing and managing their own data using SubjectBook.

As shown in 4.1, SubjectBook consists of two main parts: Front-end, and Back-end. The Front-end is the part of the system that users can see and interact with. Whereas the Back-end is the part of the system that is responsible for processing users' data and prepare it for visualization. This involves interacting with users' Google Drives, and also operating a local database for saving some metadata about

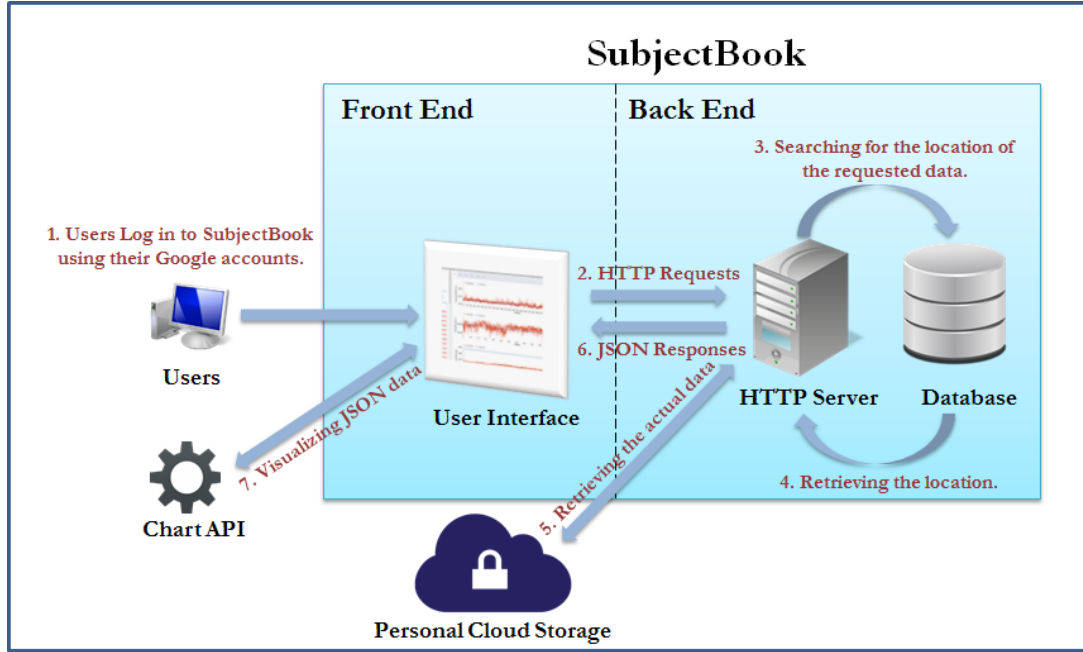


Figure 4.1: System Architecture.

each created study.

4.0.1 Front-end

The Front-end is the interface that allows users to interact with the system in order to access or visualize their data. This part of the system was built using HTML 5, Java script, J Query, Google Charts, and D3.JS.

When a request is made to explore the data for a particular subject at the lowest level of abstraction (i.e., SessionPortrait), users do not directly receive visualizations for that subject but rather an HTML page containing the required GET requests to

retrieve such visualizations from the server. After that, AJAX requests will automatically be made to visualize only the first session (i.e., the first tab). Requests to visualize other sessions will only be made upon users' request (i.e., when clicking on the corresponding session's tab). This approach is necessary to prevent users from making a lot of requests that may harm the performance of their machines, and also to keep our server from being overwhelmed by unnecessary users' requests.

Each single data set needs a separate AJAX call to be retrieved. For example, if a session has two types of data, breathing rate, and hear rate, two AJAX calls need be made to the server. The benefit of having multiple calls is to reduce the time needed for users to start seeing the visualizations. Making users wait until all data is available can be frustrating for many of them.

When an AJAX call is made to visualize a particular data set, that data will be downloaded from the user's Drive, converted into a predetermined JSON format, and then sent to the user. At the client side, the received JSON data will be visualized using either Google Chart or D3.JS.

4.0.2 Back-end

The Back-end consists of 3 main parts:

- HTTP Server: it is the heart of our system, and it serves all users' requests.

After a user logs in to the system, a full access will be granted for SubjectBook to his/her Google Drives. This access will be used either to visualize existing studies or to create a study from scratch. When visualizing an existing study,

the server will use the study's location received from users to start collecting some metadata about the structure of that study, and the location of each data set on the Drive. This information will be saved in the database, and will later be used to display data to users quickly and properly. When creating a study from scratch, the information received from users about the structure of the new study will be used to create a new file structure inside Google Drive. Again, some metadata will be saved about this study. The locations of data sets will be left empty as there are no data available yet. These locations will later be filled once the data are generated.

- MySQL Database: we use a local database to store user account information, and also to store metadata about users' studies. Figure 4.2 shows a simplified picture of our database schema. The relationships between the tables at the bottom of the graph is one-to-many relationship from left to right. It is worth mentioning here that table "signal" has a field called "signal json". This field is used to store the intermediate JSON format that is sent to users to be visualized. Saving this intermediate format helps optimize the performance of the system as the corresponding file is less likely to be changed. If, however, a file was changed SubjectBook will automatically update the JSON data.
- User's Cloud Storage: It is the place where users keep their data in. We currently support Google Drive. Other personal cloud storage can be easily integrated.

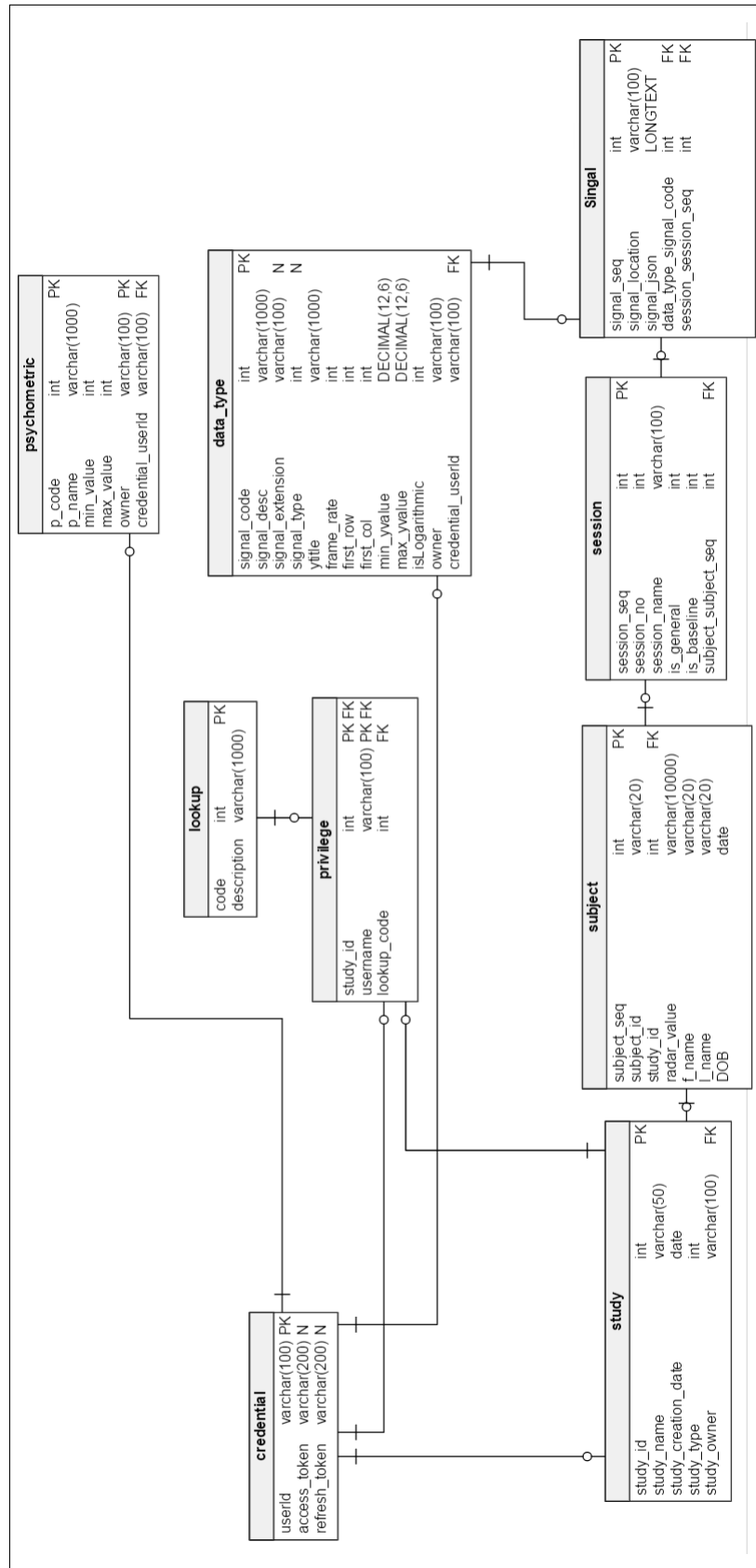


Figure 4.2: A simplified picture of the local database in SubjectBook.

Chapter 5

Case Studies

5.1 Simulator Study 1

We used SubjectBook in a study of driving behaviors under distracting stressors. A central goal of this study was to understand how driver-based and vehicle-based data might be used in tandem to detect high risk driving scenarios. In a sample of 59 drivers, balanced in terms of age and gender, the effects of cognitive, emotional, sensorimotor, and mixed stressors on driver affect and performance with respect to baseline was studied in a simulation experiment. Each subject was asked to perform all of the following drives:

- Loaded Drive - no Stressor (LD1): Driving with no secondary activity (no additional stressor).

- Loaded Drive with Cognitive Stressor (LD2): Driving under a cognitive stressor. The cognitive stressor was mathematical questions in one phase of the drive and analytical questions in another phase of the drive, posed orally by the experimenter.
- Loaded Drive with Emotional Stressor (LD3): Driving under an emotional stressor. The emotional stressor was emotionally stirring questions posed orally by the experimenter in two phases.
- Loaded Drive with Sensorimotor Stressor (LD4): Driving under a sensorimotor stressor. The sensorimotor stressor was texting back words, sent one by one to the subjects smartphone; this texting exchange took place in two phases.

After that, each subject was asked to perform one of the following two drives:

- Failure Drive-Loaded (FDL): Driving while performing secondary tasks and experiencing a vehicle-based failure.
- Failure Drive-Not Loaded (FDN): Driving while only experiencing a vehicle-based failure (no secondary activity).

The instantaneous perspiration at the perinasal area was used as proxy of the subject's affective state, thus forming the study's explanatory variable. The instantaneous steering angle was used as proxy of the subject's driving performance, thus forming the study's response variable. To ascertain that the experiment's challenging drives were perceived as such, we asked subjects to complete the NASA Task Load



Figure 5.1: The driving simulator experimental setup.

Index (TLX) after each drive. The experimental setup for this study is shown in Figure 5.1

Figure 3.31 shows the case's StudyPortrait. LD1 is the baseline session and is not represented in the StudyPortrait's grid, because it is absorbed in the paired comparisons. FDL and FDN are special sessions that have no across session baseline (only within session) and for this reason cannot be displayed at the StudyPortrait level. Emotional, cognitive, and sensorimotor stressors were used in LD2, LD3, and LD4, respectively. These standard interventional sessions are represented in the grid by two rows each - one for the explanatory variable, and the other for the response variable. Each row is split into 5 columns (i.e., phases), depending if a stressor was present or not: P1 (no stressor), P2 (with stressor), P3 (no stressor), P4 (with stressor), and P5 (no stressor). The results show that all stressors incurred significant increases ($p < 0.001$) in mean sympathetic arousal accompanied by significant

deterioration in mean absolute steering.

Analysts can delve into the next visualization level (SubjectPortrait) by clicking at "Browse this Study". Figure 5.2 shows a snapshot of the case's SubjectPortrait level. Remember that erratic steering performance is represented by a fully-filled disk, while the best steering performance is represented by an empty disk. This view helps in identifying subjects with high levels of physiological arousal, such as subject T025. Additionally, the subjective feedback after each session can be used to explain unexpected behaviors. For example, the green semicircle in session LD3 of subject T024 means that she did not perceive the intervention as challenging, thus, explaining the low arousal indicated by the short color bar.

SubjectPortrait helped us easily find a serious problem in our study's experimental design. The problem was that the FDN task was only assigned for male subjects, while FDL was assigned for female subjects. Figuring that out early in the study life cycle enabled us to fix that problem by simply recruiting more subjects.

To gain a deeper insight as to how a particular subject responded to stimuli in the experiment, analysts can click on the subject icon to delve into the next visualization level - the SessionPortrait. Data collected during session LD4 for subject T029 is presented in Fig. 5.3. The line chart at the upper portion of the screen represents the explanatory variable (i.e., arousal), while the line chart to the bottom represents the response variable (i.e., steering). The two yellow backgrounds in each chart mark the interventional phases, that is, texting. It is evident from the line charts that as soon as the subject starts texting, the physiological arousal increases, and steering performance deteriorates.

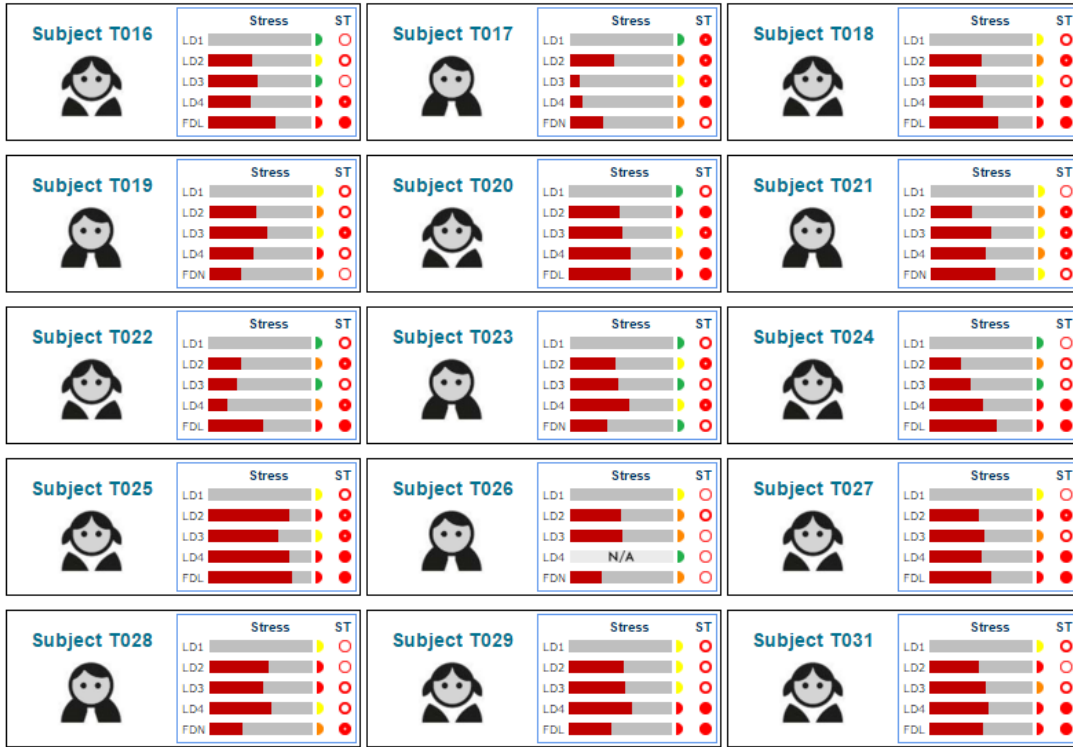


Figure 5.2: SubjectPortraits for the simulator study 1.



Figure 5.3: The SessionPortrait for subject T029 in session LD4.

Table 5.1: The list of session in the track study

Session Name	Acronym	Type	Group(Arm)	Length(s)
Practice Drive	PD	Baseline	All	240
Normal Drive	ND	Cross-Baseline	All	480
Cognitive Drive	CD	Cross-Intervention	All	480
Motoric Drive	MD	Cross-Intervention	All	480
Failure Drive-Cognitive	FDC	Intr-Intervention	Arm 1	90
Failure Drive-Motoric	FDM	Intr-Intervention	Arm 2	90

5.2 Track study

SubjectBook was also used to manage another driving study that has the same goal as the previous study. However, a real car was used this time. Twenty-six subjects have participated in this experiment, each was asked to perform 5 real drives. These drives and their attributes are presented in Table 5.1.

PD drive is used to allow drivers to become familiar with the driving environment. CD and MD drives are intended to load drivers cognitively and motorically, respectively. ND drive serves as a baseline against which the loading drive data will be compared. FDC and FDM drives aim to understand how drivers respond to that additional load while cognitively and motorically loaded, respectively. The additional load is a vehicle-based failure, which is a sudden uncontrolled acceleration of the participant's vehicle. Figure 5.4 shows the experimental setup for this study.

The following are the data that have been collected in this study:

- **Key explanatory variable:** perinasal perspiration
- **Secondary explanatory variables:** palm EDA, heart rate, and breathing



Figure 5.4: The track study experimental setup.

rate

- **Key response variable:** steering
- **Secondary response variables:** speed, acceleration, braking, and lane deviation
- **Questionnaires:** Type AB, TAI, and NASA TLX

Figure 5.5 shows a snapshot of the case’s SubjectPortrait level. One quick observation that can be made using this visualization is that women generally perform less well than men. Additionally, there is a correlation between subjects’ perceptions of task difficulty and their performance.

Figure 5.6 provides an example of a subject’s SessionPortrait. This view helps

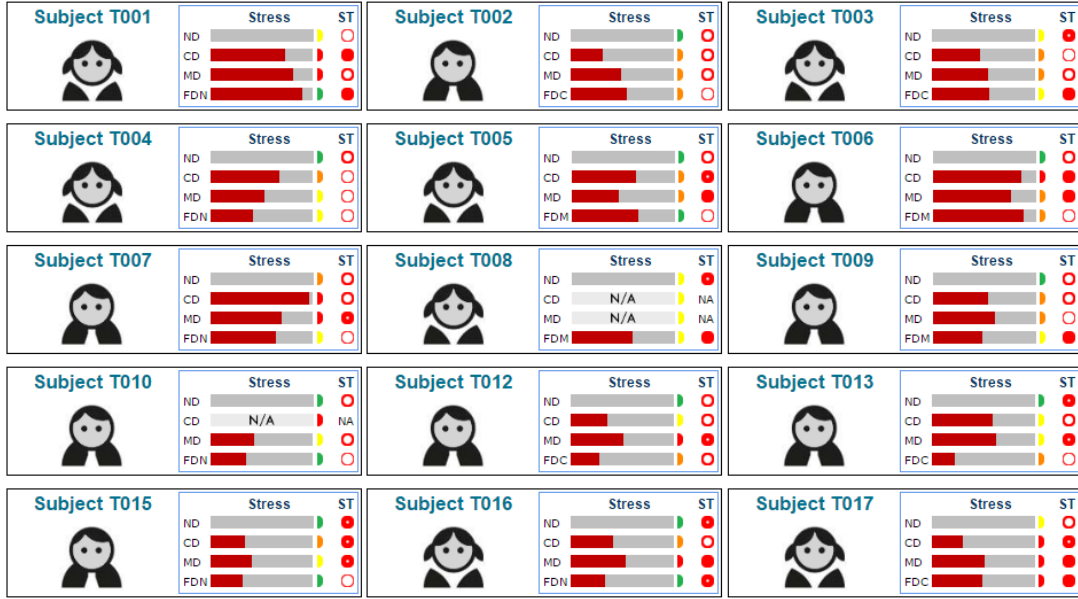


Figure 5.5: SubjectPortraits for the track study.

compare between different variables within each session, and also helps compare between each variable and its corresponding one in the baseline session. SessionPortrait was used to identify a particular pattern that occurs in almost all sessions. As Figure 5.7 demonstrates, there is a sudden increase in subjects' sympathetic responses (i.e., perinasal perspiration) between every two consecutive interventions. This increase happens when a subject makes a U turn, causing the tracker to be placed on a different area other than the perinasal area. This is obviously not legitimate data, and the signal needs to be cleaned before starting to analyze the data. This observation was greatly facilitated by the visualizations offered by SubjectBook. It would have been very difficult to notice this problem using the traditional statistical methods.



Figure 5.6: This figure shows how SessionPortrait enables researchers to compare between different variables in the track study.

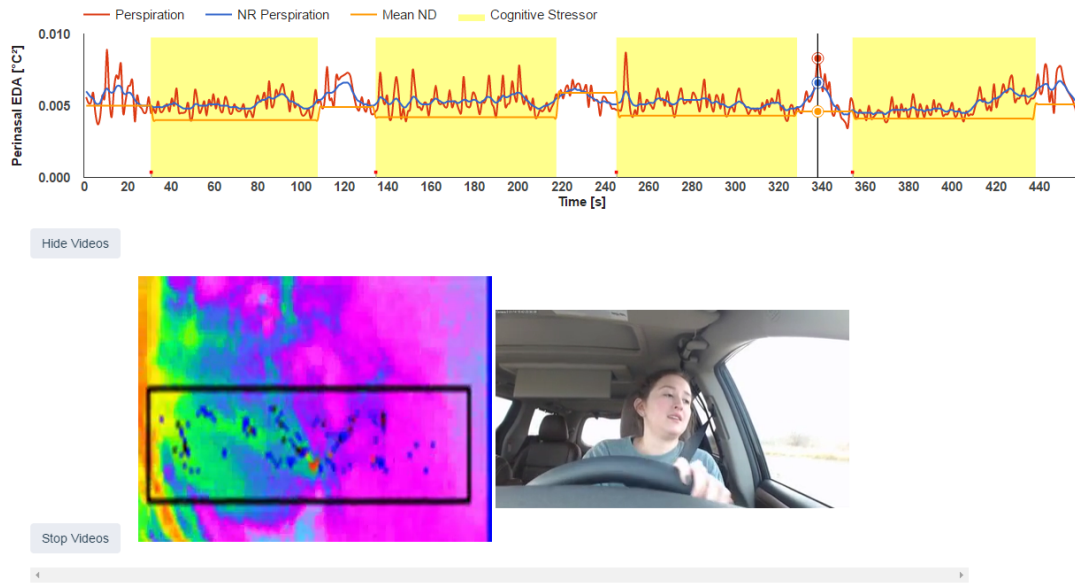


Figure 5.7: A view that shows how SessionPortait drew our attention to a hidden pattern in our data.

5.3 Simulator study 2

In this study, a simulator study was used to test and validate two countermeasure conditions (control, SIT) on driver response (e.g., stress and performance) under different driving loads (i.e., Normal, Cognitive, and Motoric).

The goals of this study are:

- To understand how the countermeasures affects driver performance and safety.
- To identify the vehicle state patterns/dynamics associated with problematic driver performance.
- To determine the interaction between car and driver condition.

Table 5.2: The list of sessions in the simulator study 2

Session Name	Acronym	Type	Group(Arm)	Length(s)
Baseline	BL	Baseline	All	280
Practice Drive	PD	Baseline	All	540
Normal Drive	ND	Cross-Baseline	All	540
Cognitive Drive	CD	Cross-Intervention	All	540
Motoric Drive	MD	Cross-Intervention	All	540
Final Drive	FDC	Intr-Intervention	Arm 1	180

5.3.1 Countermeasure Conditions

Two countermeasure conditions were used in this study: Control, and Stress Indicator Technology (SIT). Participants were exposed to different loading drives as well as a failure drive to further evaluate physiological response and driving performance under each countermeasure condition. A description of each condition is listed below:

- **Control:** participants received no training and experienced no countermeasures during their loaded and failure drives.
- **Stress Indicator Technology(SID):** for all conditions, the center wheel of the simulator was equipped with an orange light. The light was activated when participants physiological levels of stress reached a pre-determined value. Participants were given information on the stress indicator light (see Procedure section), but no training was provided.

Fifty-six subjects participated in this study. The list of drives that each subject performed in this study is presented in Table 5.2. BL and PD drives aim to allow the driver to become familiar with the operational characteristics of the driving

environment simulator and the task of driving. The intent of the ND drive is to allow time for the participant to become as relaxed as possible before starting the Loading drives and to serve as a baseline against which the Loading drive data will be compared. CD and MD drives are intended to load drivers cognitively and motorically respectively. Finally, FD is intended to examine driver responses to a vehicle-based failure. All participants will complete the failure drive while performing a secondary task (i.e., cognitive).

The following are the data that have been collected in this study:

- **Key explanatory variable:** perinasal perspiration
- **Secondary explanatory variables:** palm EDA, wrist EDA, heart rate-chest, heart rate-wrist, and breathing rate-chest.
- **Key response variable:** steering
- **Secondary response variables:** speed, acceleration, braking, and lane deviation
- **Pre-study questionnaires:** SAI, TAI, ATTENTIONAL CONTROL, and PERSONALITY TYPE A/B
- **Post-study questionnaires:** TAI, POST STUDY, and USABILITY
- **Session-level questionnaires:** NASA TLX

Unlike in the previous two studies, data in this study was automatically collected and curated. Such data were made available for five collaborating teams from across

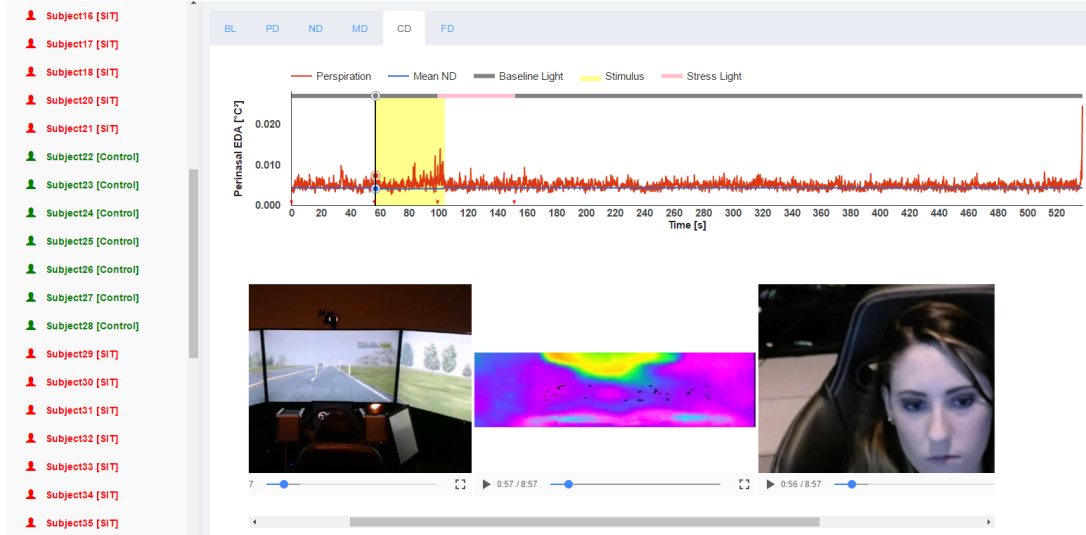


Figure 5.8: The SessionPortrait for a particular subject in simulator study 2.

the country in real-time. This scenario enabled our team to perform quality control in a timely manner despite the fact that we were 200 miles away from where the experiment was taking place. We were able to make quick decisions regarding what subjects needed to be replaced either because they did not follow the rules, or because data was corrupted due to technical issues. Finally, the algorithm that was used to turn the light on and off was evaluated and adjusted based on the data collected for the first few subjects. This allowed us to use the best algorithm possible early in the study life cycle. If the data was not available in real-time that would be very difficult to achieve.

Figure 5.8 is the SessionPortrait for a subject from the SIT group. This view shows when the light was off (gray) and when it was on (pink). Also, the yellow color shows when the stressor was applied. The list of subjects to the left helps users easily distinguish between different groups.

Chapter 6

Conclusion

We introduced SubjectBook, a software package that integrates our data management and visualizations methods for affective studies. These methods aim at helping researchers manage their data sets throughout their life cycles.

6.1 Summary of Contributions

6.1.1 Automating the Experimental Designs

The process of allocating experimental units among treatment groups is performed automatically using SubjectBook. Additionally, the order in which subjects perform tasks is automatically randomized to avoid order effects.

6.1.2 Managing Affective Data Sets

SubjectBook helps researchers store their data whenever it is available, and provides them with a convenient way to access such data from anywhere at anytime. This assists them with checking progress and performing quality control. Moreover, SubjectBook enables researchers to generate ready-to-analyze data that can be used to perform in-depth statistical analysis.

6.1.3 Collecting, Curating, and Visualizing Data in Real-time

SubjectBook cooperates with other software components to automatically collect, curate, and visualize data in real-time. This helps avoid human errors, and also enable researchers to perform quality control in a timely manner.

6.1.4 Novel Visualization for Hypothesis-Driven Affective Studies

We presented a novel approach to visualizing multimodal affective datasets. Three levels of data abstraction mirror the scientific method. First, data from different sources are combined and presented in a way that supports cause and effect association within context. Second, information for each subject is abstracted into a figure reminiscent of an ID card, enabling analysts to appreciate within subject phenomena. Third, the inter-subject outcomes of the study are represented through a grid

visualization. Importantly, SubjectBook is a web-based tool that enables researchers to share data with other collaborators or with the research community.

6.1.5 Data Transparency

SubjectBook gives researchers the capability to share their data with other collaborators or with the research community. This is considered very important in today's science that is hypothesis-driver.

Bibliography

- [1] S. A. Brown, M. Folk, G. Goucher, R. Rew, P. F. Dubois, et al. Software for portable scientific data management. *Computers in Physics*, 7(3):304–308, 1993.
- [2] S. R. Brown and L. E. Melamed. *Experimental design and analysis*. Number 74. Sage, 1990.
- [3] M. Folk, G. Heber, Q. Koziol, E. Pourmal, and D. Robinson. An overview of the hdf5 technology suite and its applications. In *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases*, pages 36–47. ACM, 2011.
- [4] A. Fouse, N. Weibel, E. Hutchins, and J. D. Hollan. ChronoViz: A system for supporting navigation of time-coded data. In *CHI’11 Extended Abstracts on Human Factors in Computing Systems*, pages 299–304, Vancouver, British Columbia, May 7-12 2011. ACM.
- [5] J. Gray, D. T. Liu, M. Nieto-Santisteban, A. Szalay, D. J. DeWitt, and G. Heber. Scientific data management in the coming decade. *ACM SIGMOD Record*, 34(4):34–41, 2005.
- [6] R. J. Hanisch, A. Farris, E. W. Greisen, W. D. Pence, B. M. Schlesinger, P. J. Teuben, R. W. Thompson, and A. Warnock III. Definition of the flexible image transport system (fits). *Astronomy & Astrophysics*, 376(1):359–380, 2001.
- [7] R. J. D. W. James et al. *Handbook of affective sciences*. Oxford University Press, USA, 2002.
- [8] J. Kim, M. Snodgrass, M. Pietrowicz, K. Karahalios, and J. Halle. BEDA: Visual analytics for behavioral and physiological data. In *Workshop on Visual Analytics in Healthcare*, pages 23–27, Washington DC, November 16 2013.
- [9] K.-A. Kwon, D. Shastri, and I. Pavlidis. Interfacing information in affective user studies. In *Proceedings of the 2014 ACM International Joint Conference*

- on *Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 87–90, Seattle, Washington, September 13–17 2014. ACM.
- [10] M. Mahmoud, T. Baltrušaitis, P. Robinson, and L. D. Riek. 3d corpus of spontaneous complex mental states. In *4th International Conference on Affective Computing and Intelligent Interaction*, pages 205–214. Springer, Berlin, 2011.
 - [11] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 3(1):5–17, 2012.
 - [12] M. Oefinger, W. Zong, M. Krieger, and R. Mark. An interactive web-based tool for multiscale physiological data visualization. In *Computers in Cardiology*, pages 569–571, Chicago, Illinois, September 19–22 2004. IEEE.
 - [13] I. Pavlidis, M. Dcosta, S. Taamneh, M. Manser, T. Ferris, R. Wunderlich, E. Akleman, and P. Tsiamyrtzis. Dissecting driver behaviors under cognitive, emotional, sensorimotor, and mixed stressors. *Scientific reports*, 6, 2016.
 - [14] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddhharaju, R. Joseph, A. Mandapati, B. Dunkin, and B. Bass. Fast by nature-how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports*, 2:305; doi: 10.1038/srep00305, 2012.
 - [15] M.-Z. Poh, N. C. Swenson, and R. W. Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5):1243–1252, 2010.
 - [16] R. Rew and G. Davis. Netcdf: an interface for scientific data access. *IEEE computer graphics and applications*, 10(4):76–82, 1990.
 - [17] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG 2013)*, pages 1–8, Shanghai, China, April 22–26 2013. IEEE.
 - [18] D. Sander and K. Scherer. *Oxford companion to emotion and the affective sciences*. OUP Oxford, 2009.
 - [19] K. R. Scherer. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, methods, research*, 92:120, 2001.

- [20] H. A. Simon. Invariants of human behavior. *Annual Review of Psychology*, 41(1):1–20, 1990.
- [21] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, 2012.
- [22] S. Taamneh, M. Dcosta, K.-A. Kwon, and I. Pavlidis. Subjectbook: Hypothesis-driven ubiquitous visualization for affective studies. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1483–1489. ACM, 2016.
- [23] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.
- [24] A. Wesley, D. Shastri, and I. Pavlidis. A novel method to monitor driver’s distractions. In *CHI’10 Extended Abstracts on Human Factors in Computing Systems*, pages 4273–4278. ACM, 2010.
- [25] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. ELAN: A professional framework for multimodality research. In *Proceedings of The 5th International Conference on Language Resources and Evaluation (LREC)*, volume 2006, pages 1556–1559, Genoa, Italy, May 22-28 2006. ELRA.