Visual Analysis of Relationships between Behavioral and Physiological Sensor Data

Jennifer G. Kim[1], Melinda Snodgrass[2], Mary Pietrowicz[1], and Karrie Karahalios[1]
Department of Computer Science[1] and Special Education[2]
University of Illinois at Urbana-Champaign
Email: {jgkim2, mrsnodg2, mpietro2, kkarahal}@illinois.edu

Abstract*—More and more people are collecting, organizing, and interpreting health data through off-the-shelf products such as Apple's HealthKit. People use these systems to collect steps taken, calories ingested, etc. These ecosystems also support the collection of physiological data. For example, users collect heart rate data during exercise, and even electrodermal activity data to help detect the onset of seizures. Analyzing physiological data, however, and relating it to specific behaviors or events, is challenging. In this paper, we present an 11-week, multi-session, participatory design case study, identify challenges in analyzing physiological and behavior data, and present BEDA, an analytics tool we developed to mitigate the challenges. The two primary data analysis challenges include: (1) interfacing multiple software programs required for capturing and analyzing the different data sources, and (2) extending the limited data analysis functionality within and across these software programs to support a wide range of physiological data analyses. BEDA resolves the fragmented analysis pipeline by integrating closely-related analysis tasks into a common interface. It also addresses the extensibility problem by integrating scripts that apply any custom or publicly-available function written in MATLAB or R. These scripts extend basic analytic capability, provide the analytic bridge between physiological and behavior data, and incorporate machine learning algorithms to highlight behaviors associated with physiological data. BEDA's capabilities mitigated the challenges of signal analysis and fragmented data sources, and motivated behavioral scientists to combine physiological measures with behavioral analysis. Although we developed this tool for a domain-specific case study, the use of the tool can be generalized to analyze any time-based data source or sources.

***Keywords*—Behavior analysis, physiological sensor data analysis, visual analytic tool.**

I. INTRODUCTION

The development of wearable sensor technologies allows researchers, clinicians, and non-experts to easily collect their subjects' or own physiological data in everyday activities and to analyze specific data for a wide range of health monitoring purposes [1], [2], [3]. For example, heart rate monitoring can identify excessive exertion [1] and stressful situations [3] in everyday life. Researchers have used electrodermal activity (EDA) to identify and predict seizure conditions [2] so that the affected person can take appropriate emergency measures. Associating these behaviors or events with physiological data helps users better understand the context in which the behaviors and events occur, which can lead to the development of effective therapies and increased personal behavior knowledge. However, analyzing many physiological data sessions with respect to specific events is very challenging [4].

We conducted a case study of a participatory design practice [5] with five behavioral scientists, one clinician, and three computer scientists to identify challenges in analyzing twenty-five longitudinal sessions of physiological data and behavioral observation data. We developed a visual analytic tool, Behavioral and External Data Analysis (BEDA), to mitigate the difficulties identified. The two primary challenges with the data analysis revolved around requiring the use of multiple software programs to analyze each data source (i.e., physiological data and observational behavior data), and the lack of extensibility of existing data analysis software to analyze a wide range of physiological sensor data and apply state-of-art algorithms. BEDA resolved the fragmented analysis pipeline problem by integrating interfaces for closely related analysis tasks. The physiological data analysis became extensible and easy for novice programmers by integrating MATLAB [6] and R [7] scripts into BEDA. The scripts are designed to access any physiological or behavioral data visualized in BEDA and apply any MATLAB or R function to analyze the data. Thus, a wide variety of the state-of-art machine learning algorithms can be used in BEDA to predict behaviors of interest using physiological data. BEDA motivated behavioral scientists to continue combining physiological measures with behavioral analysis. Because the tool solved the challenges in the data analysis process, it encouraged the users to continue collecting and analyzing data.

Although this tool was developed in the context of a domain-specific case study, the use of the tool can be generalized when analyzing similar data sets. A growing number of individuals who are interested in collecting physiological data in their everyday lives using wearable sensors [8], [9] can use our tool to understand their physiological data changes in relation to their behaviors or events (e.g., having a very stressful meeting, driving, taking a class, or sleeping).

The main contributions of this paper include:

- The application of commonly used visual analytic techniques to a new specialized area: behavior analysis augmented with physiological measures.
- The design a unified pipeline for visual analysis of multiple, time-based data sources to better find relationships between the different data sources and trends across sessions.
- An environment where non-expert programmers can use MATLAB and R functions to analyze both the behaviors and physiological sensor data visualized in the tool.

- An extensible environment that integrates machine learning algorithms to highlight key time periods of physiological data where behaviors of interest might occur.

II. Related Work

In this section, we compare our work to the existing software tools developed for behavioral coding and sensor data analysis. Although we include commonly-used visualization techniques in BEDA, our contributions are making the tool highly extensible, integrating a fragmented analysis pipeline, and applying a set of visualizations to a new domain, the combination of behavioral and physiological measures and analysis.

Elan [10], VCode/VData [11], Noldus[12], and Chronoviz [13] currently support behavioral coding from video recordings by allowing users to create annotations for different types of behaviors and insert the annotations into the video time frame. These tools, however, do not support sensor data analysis, which means users have to use other software programs to analyze sensor data, making it difficult to compare sensor data analysis results and behavioral coding annotations.

Conversely, Ledalab[14] and AcqKnowledge[15], have specific physiological data analysis algorithms embedded in them which provide graphical user interfaces that allow users to change the parameters of an algorithm, run an analysis, and review a visualization of the analysis results. However, AcqKnowledge is very expensive to purchase, and Ledalab only supports algorithms for EDA analysis. These tools limit researchers to the sensor data analysis algorithms that the tools provide, and cannot apply other state-of-art algorithms. Furthermore, these tools do not support behavior coding from video. Other software tools, such as MATLAB [6] and R[7], allow users to write algorithms to import, analyze, and visualize results for their sensor data streams. However, these programs have steep learning curves for people who are not familiar with programming.

Behavior annotations and sensor data are, respectively, categorical and continuous time-series data; and previous research has explored visualization of time-series data to help quickly identify patterns within time-series data. A common approach is to represent time as a linear ordered axis and plot data cases by their timestamps [16]. This linear representation helps identify overall trends and peaks in time. LiveRACs [17] and Line Graph Explorers [18] both show multiple views of time-series graphs in a grid-based layout to compare different time-series data side-by-side at multiple levels of detail. We use a similar approach to help behavioral scientists better identify the relationship between the continuous and categorical time-series data.

None of the currently available tools allow users to view and analyze sensor and behavior data captured on video simultaneously. To address this and the other drawbacks to the previously existing tools listed above, we designed BEDA as a free and open-source software package to address these issues.

III. Understanding the Problems

In this section, we describe the purpose of the behavioral science study in which BEDA was piloted and the challenges in data analysis using traditional analysis tools.

BEDA was designed to support behavioral scientists' data analysis of a single-case design study [19] in which twenty-five sessions were conducted, with physiological sensor data and behavioral data captured on video during each session. Three computer science researchers collaborated with five behavioral scientists and one clinician to identify challenges in the data analysis process, and, at their request, concurrently develop a tool, BEDA, to support the analysis process. The computer science researchers were involved in this study for five months, during which the team formulated research questions and methods, collected video and sensor data over an 11-week period, and analyzed the obtained data.

A. The Purpose of the Behavior Science Study

The purpose of the behavioral science research study was to examine the efficacy of pressure vests – a tight vest that gives the wearer the sensation of receiving a hug – that are commonly used in occupational therapy for individuals with disabilities [20]. In such Sensory Integration (SI) therapy, an area of occupational therapy, it is hypothesized that deep pressure from the vest will help the wearer better process sensory inputs by providing a sense of calm [21]. The calming effect is then hypothesized to lower arousal levels leading to improved engagement in on-task behaviors and a decrease in undesired or problem behaviors [22], [23]. However, few studies explore the relationship between the use of a pressure vest and positive changes in both the physiological arousal levels and observable behavior of the wearer [20], [24]. Thus, the goal of this research study was to explore the following two research questions.

RQ1: Do changes in arousal correlate with observable changes in the level of engagement with a task and/or the level of undesired behavior?

RQ2: Does wearing a pressurized vest produce measurable changes in arousal level, the level of engagement with a task, or the level of undesired behavior?

We conducted a longitudinal study with a child whose occupational therapist (OT) specifically recommended pressure vests for his therapy. To measure arousal levels, the child wore a Q-sensor [25] that measured his EDA, a measure of change in electrical skin conductance produced by changes in sweat production. The sweat glands are controlled by the sympathetic nervous system, and the changes in EDA may therefore reflect changes in arousal [26]. Specifically, gradual changes in EDA, called tonic EDA, are known to reflect general arousal level [27] [26] [28]. Rapid changes in EDA, called phasic EDA, are known to measure more immediate changes in arousal levels (i.e., sensory reactivity) of the individual who is wearing the sensor [27] [26] [28]. To measure levels of engaged and undesired behavior, we video recorded each session and analyzed these videos to determine the frequency with which these behaviors occurred. Each of our 25 intervention sessions included EDA data (for measuring arousal), and video data (for observing behavior). The full explanation of this pressure vest study [29] is presented in a separate paper. Here, we focus on describing the challenges that we identified in the data analysis procedures for the study and how our tool (BEDA) supports our data analysis for this and similar studies.

B. The Current Data Analysis Procedure

The two major analysis steps we used to answer our research questions are as follows: (1) the comparison of coded behaviors and EDA signal analysis results within each session, to see how behaviors and arousal levels are related, and (2) the comparison of each session's behavior and EDA signal analysis results across sessions or conditions, to assess how wearing a pressure vest produces changes in behaviors and arousal levels.

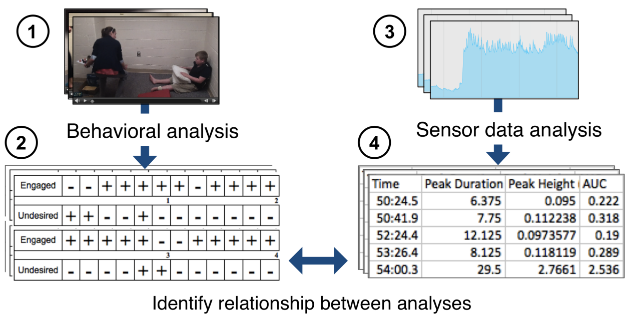

Fig. 1. The first step of the analysis: While watching (1) a video of each intervention session using QuickTime Player, (2) code behaviors of interest in a table form created using Microsoft Word. Then, (3) analyze each session's EDA signal using Q software and (4) verify the analysis results using Microsoft Excel. Finally, compare each session's coded behaviors and EDA signal analysis results.

In the first step of the analysis, researchers first defined the child's undesired behaviors (e.g., yelling, or hitting the table) and engaged/on-task behaviors (e.g., manipulating materials, looking at the teacher) which the participant's OT hypothesized would be affected by wearing a pressure vest. The defined behaviors were coded using momentary time sampling [30] – a commonly used technique for behavioral analysis to quantify whether a behavior occurs during specified time periods – by watching the video on a QuickTime player (Fig. 1-1) and documenting occurrences on a table form (Fig. 1-2) that the researchers created using Microsoft Word. Researchers coded '+' in the corresponding cell on the table (Fig. 1-2) if the defined behaviors occurred during the last one second of every 10-second interval in the video recording of the intervention sessions.

To analyze the EDA signal, we attempted to use the Q software (Fig. 1-3) and Excel (Fig. 1-4). However, Q software did not support the separation of the tonic and phasic portions of EDA signal. Thus, we developed an analysis script for separating tonic and phasic of EDA signal using MATLAB. Then, the researchers individually noted the time points when undesired or engaged behaviors occurred via behavioral coding (Fig. 1-2) and looked at the corresponding time point of sensor data analysis results (Fig. 1-4) to examine how occurrences of behaviors were associated with changes in EDA signal. This comparison process continued for each of 25 intervention sessions.

In the second step of the analysis, researchers first calculated each session's behavioral measures and EDA signal

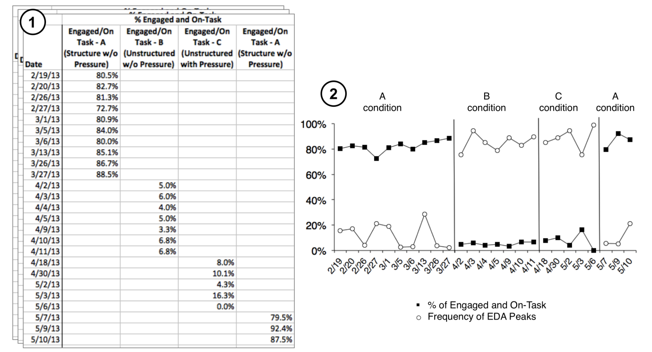

Fig. 2. The second step of the analysis: (1) Calculating each session's behavioral and physiological data features (2) Visualizing the features across different sessions using Microsoft Excel

features. To summarize a behavioral measure per session, they counted the number of intervals in which the undesired and engaged behaviors occurred out of the total number of intervals in a session. To summarize EDA signal features per session, they calculated the average tonic EDA to measure the general arousal level and the average phasic EDA to approximate the short-term arousal level [28]. Additionally, the average number of peaks per minute and the average peak amplitude of phasic EDA were calculated to measure the frequency and the level of short-term arousal, respectively. Finally, researchers organized the results into a single Excel file (Fig. 2-1) to allow for manual comparison of the feature vectors across sessions and/or conditions. An example of this graph with data that approximate the data generated in the study is presented in Fig. 2-2. Sometimes, physiological features were highly correlated with each other. For example, the phasic EDA average and the phasic EDA average peak amplitude had a correlation of 1.0. One of the highly correlated features, phasic EDA average, was manually removed in the Excel file because this measure did not contribute additional information.

C. The Problems of Current Data Analysis Procedure

As we have mentioned, the two major problems with the current data analysis procedure are (1) a fragmented analysis pipeline, and (2) the lack of extensibility of the existing data analysis tools.

1) Fragmented data analysis pipeline: The behavioral scientists' data analysis procedure required four different software programs. They used Quick-Time video player and Word for behavioral analysis, and the Q software and Excel for physiological sensor data analysis. This fragmented analysis pipeline created quite a burden for the researchers who had to organize 25 sessions of video files, 25 Word documents containing behavioral analysis, 25 physiological Q sensor data files, and 25 Excel files containing the physiological data analysis results. The separation between the data sources (video and physiological sensor data) and analysis results posed additional difficulties in examining the relationship between behavior and physiological data. Because sensor data analysis results were separated from the corresponding video and behavioral coding, it was also difficult to know the context of the exact behavior under which the sensor data analysis results occurred. For example, to determine if a peak in EDA data was generated from movement or from an undesired behavior, researchers had to individually note the time of the peak in the sensor data analysis results in the Q software, open the corresponding video file in QuickTime, and attempt to find the same time in the video to verify the exact behavior. This comparison process required using multiple software programs, organizing and managing the numerous data files, and introduced the opportunity for error (e.g., observing different events in sensor data than in video data because the timelines of these sources were not linked).

2) Limited extensibility of existing analysis tools: Although behavioral scientists may have extensive knowledge of the physiological data analysis measures, computational physiological data analysis is not common in this field; and the unfamiliar analysis algorithms were challenging. As a result, most behavioral scientists use existing data analysis tools provided in a software package that simply allows them to select and apply pre-programmed algorithms to their physiological data. These tools are limited in their analytic capabilities, and do not contain all of the state-of-the-art analysis algorithms that support the range of data that we and other researchers want to examine. Furthermore, since the pre-programmed algorithms are typically hidden in the tools, it is difficult to verify whether the tool used the same analytic algorithms that were used in previous research studies. For example, one of our behavioral scientists had to ask the Q software engineers for the exact procedures and parameters used in their "finding peak" algorithm to know how to report results accurately.

IV. Design Rationale

To resolve the major difficulties in the current data analysis, we applied the proximity compatibility principle (PCP) [31]. This principle suggests that information relevant to a common task be rendered in close proximity to each other. Thus, in this section, we combined the closely related analysis tasks into three major tasks. We describe the design motivations for the development of BEDA below.

A. Comparing Coded Behaviors and Physiological Data

To easily identify whether changes in EDA signal were associated with behaviors coded from the videos, the following three sub tasks need to be accomplished.

1) Aligning video and physiological data over time: The timing accuracy between video footage and the physiological data is critical for accurately examining associations between behavior and physiological response. The accurately synchronized video and physiological data also helps examine the context in which a particular behavior of physiological response occurred. A discrepancy of one frame in the video footage and the physiological data can lead to inaccurate research conclusions.

2) Coding behaviors: We coded behaviors from video data to annotate the occurrence of behaviors of interest in each session. To do this, two researchers were trained to identify specific, operationally defined behaviors in the videos. Initially, each researcher individually coded the behaviors. We then compared the behavioral codes to verify reliability of the coding system. If a coding discrepancy occurred, the two researchers watched the corresponding time window of the video together, verified the exact behavior, and determined the accurate behavioral codes by consensus. For this coding, it is critical that each entered behavior code corresponds to the same exact time point in the video for each coder and across coders. BEDA guaranteed this consistency. [1] Connecting the annotations to the video footage allowed researchers to easily and accurately return to examine the context in which a behavior occurred and identify the particular behavior (e.g., tapping the table versus yelling).

3) Processing physiological sensor data: The purpose of analyzing physiological sensor data is to accurately capture the features of physiological data that are relevant to the individual and/or to the occurrences of particular behavior. Sensor data collected in the real world, however, often contain outlier data, noise, or invalid data (e.g., because of movements of a sensor wearer or disconnection of the sensor). Identifying these instances and removing them from the data is an important step to improve the quality of physiological sensor data [32]. Furthermore, finding an appropriate parameter setting from the analysis algorithm that can best capture the features of physiological data of each individual is crucial because each person often has different physiological responses [28].

B. Comparing Features Across Sessions

Visualizing high-level trends of both behavioral measures and physiological data features can help researchers understand if analysis results represent a significant change different study conditions (such as Fig. 2-2). To do this, the following three sub tasks need to be tied together.

1) Calculating each session's features: Each session's behavioral measures and physiological data features need to be calculated to easily compare the measures and features across sessions.

2) Finding a concise set of features: A goal of feature selection is to discover a concise set of essential features that represent the main characteristics of the original EDA signals accurately [33]. We strive to identify physiological features that are uncorrelated with each other but highly correlated with behavioral measures. Uncorrelated EDA features capture different characteristics of the original EDA signal. On the other hand, highly correlated EDA features imply that these features capture similar characteristics from the original data. Removing overlapping EDA features minimizes redundancy in the feature set.

3) Visualizing features across sessions: When calculated features for a session are higher or lower than other sessions, researchers should be able to easily and concurrently inspect the sessions data analysis results for both physiological and behavioral data and the session video. This helps researchers understand what might have caused the change in trend. For example, if the average number of peaks per minute of EDA in a session is higher than other sessions, the researchers may discern that the participant was moving a lot in that particular session, possibly generating frequent peaks in EDA. In another instance of a higher-than-usual number of peaks, the researcher may identify that the participant was engaging in a lot of

[1] In earlier procedures, the coders manually advanced frames in a Quicktime video every 10 seconds.

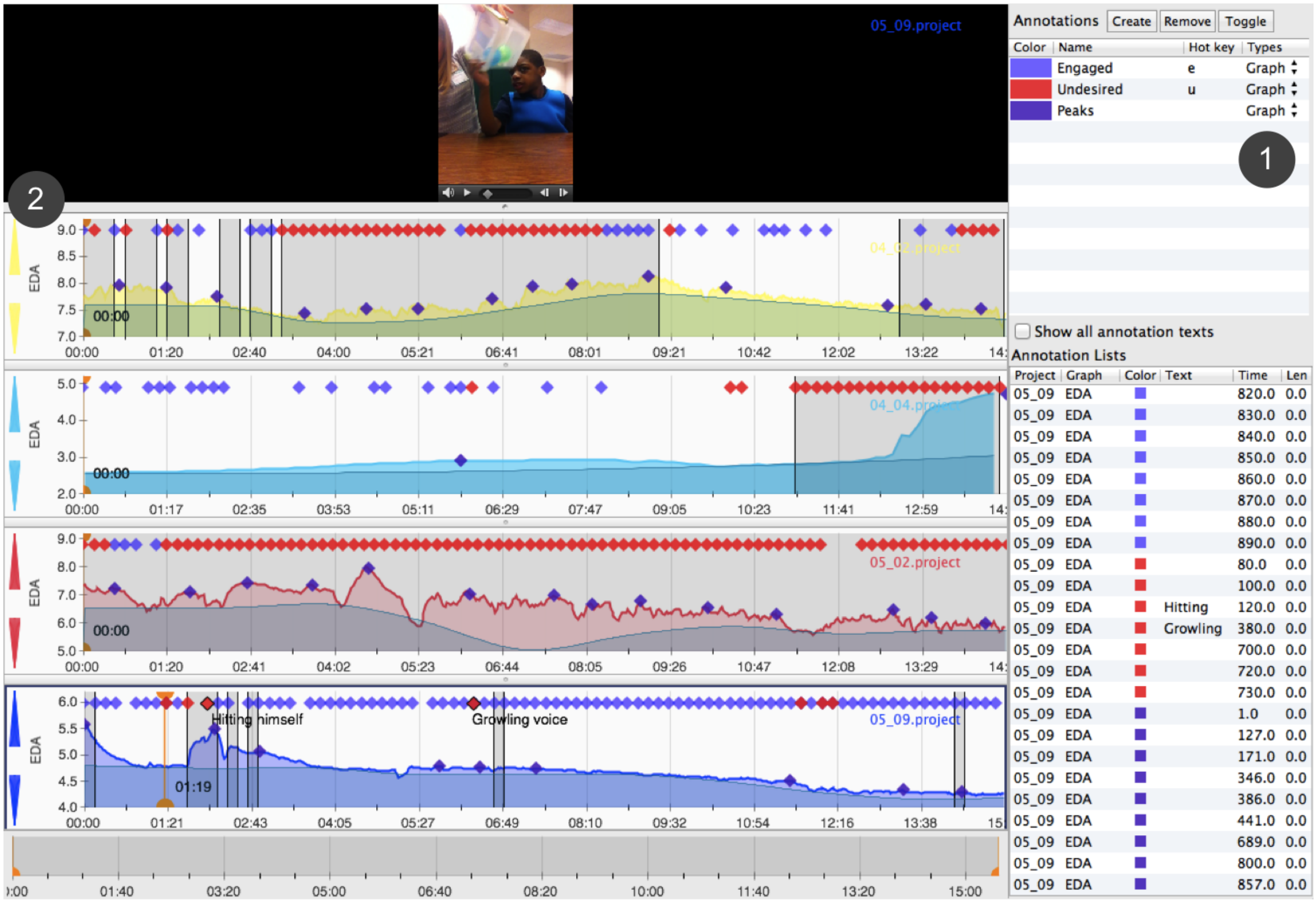

Fig. 3. The *Timeline View* is designed to compare behavior codes with physiological data. (1) Users can define the name, color, and hot key for behaviors of their choice. (2) Users code the behavior of each session by pressing a corresponding hot key while watching the session's video. Each session's raw EDA data is visualized in different colors beneath the session's behavior coding. Processed EDA data (smoothed EDA data, tonic and phasic EDA data, and peaks in EDA data) are visualized above of the raw EDA data. Thus, users can observe patterns between behavior coding and processed physiological data (e.g., most of undesired behaviors are associated with peaks in EDA data.) Finally, machine learning algorithms predict undesired behavior times from EDA features. These time intervals are highlighted via vertical bars. The use of this functionality is described in more detail in Section V-C2.

undesired behaviors (e.g., yelling) during this session, possibly generating the higher number of peaks in EDA data. BEDA needed to give researchers the capability to explore such high-level trends interactively, see the actual data analysis results, and corresponding video, and examine and explain changes in trend and correlations between data sources.

C. Extensible Data Analysis Methods

First, the data analysis methods in BEDA need to be flexible enough to cover all of the analysis tasks required for research that integrates physiological and behavioral measures. As we identified in the previous section, the analysis methods are needed for processing physiological data (e.g., separating tonic and phasic EDA) and calculating features for both coded behaviors and physiological data. Although we were creating BEDA for EDA analysis, the tool needed to be flexible to allow analysis of other physiological measures that may be taken in other studies, such as heart rate.

Second, the data analysis needs to be easy enough for novice programmers to use, while still allowing the embedded algorithms to be easily visible so that users can verify the details of the procedures and parameters. This transparency is important to ensure reliable analysis methods for research and to encourage reproducibility of the research.

V. BEDA System

In this section, we present how BEDA supports the tasks identified in the Design Rationale section. We do this in the context of a research study that explored the efficacy of wearing a pressurized vest. The term *participant* refers to the child who participated in the pressurized vest study. The term *user* refers to any prospective user of our tool (BEDA) including the behavioral scientists who conducted the pressurized vest study with us. Using PCP principles, BEDA's *Timeline View* (Fig. 3) integrates the interfaces needed for comparing coded behaviors and processed physiological data; The *Data Analysis View* (Fig. 6) integrates the interfaces needed for selecting a concise set of features and comparing the features across sessions. Data analysis functionality is extensible and supports the application of a wide range of data analysis algorithms in MATLAB and R to both coded behavior and physiological data visualized in the *Timeline View*. BEDA is implemented in Objective C [34] and we use the Core-plot library [35] for OS X for visualization. We connect MATLAB and R for categorical and continuous time-series data analysis.

A. *Comparing Behavioral Codes and Processed Physiological Data Aligned over Time*

The *Timeline View* (Fig. 3) visualizes physiological data right under the coded behaviors to help users identify how physiological data changes when a particular behavior occurs. Visualizing processed physiological data on top of the original physiological data visualization helps users find the relationships between the processed data and the coded behaviors. For example, Fig. 3 shows that most of undesired behaviors (coded in a red diamond) are associated with peaks in phasic EDA (marked in a purple diamond). Multiple sessions of data visualized in the *Timeline View* verify that the same association applies across the other sessions visualized in the same view. In the following section, we describe how BEDA supports each sub-task to make this integrated visualization.

1) Aligning video and physiological data: Video and physiological sensor data are typically not aligned in time because they are collected from different sources, such as a video camera and a wearable sensor. These two data sets often have unequal lengths, use different sampling rates, or have different start times. The *Timeline View* has two modes: *individual session mode* for aligning video and physiological sensor data, and *multiple session mode* for investigating temporally synchronized data. In the *individual session mode*, users can control each video and sensor timeline individually by moving each data source to the time point where they want to align the two data sources. After aligning the data, they can go back to the *multiple session mode* where the video and the physiological sensor data streams will be controlled in a unified time frame.

2) Coding behaviors: BEDA simplified behavior coding by allowing users to watch a session video and press a corresponding hot key that the user has assigned to each behavior of interest (Fig. 3-1). This connects a coded behavior to a corresponding moment of video. Different colors highlight the frequency of undesired and engaged behaviors in each session, and users can annotate the behavioral coding with text. To demonstrate, the fourth row of Fig. 3, session 05_09 shows that the child demonstrated many engaged behaviors (blue diamonds) and few undesired behaviors (red diamonds). Some of the undesired behaviors that occurred in the beginning of the session were associated with higher peak amplitude of EDA and others were not. To identify what may account for this difference, users move the orange header of EDA data to the undesired behavior coding in the beginning of the session and it synchronously moves the video time frame to the corresponding time. Users found that an undesired behavior tagged in the video *hitting himself* corresponded with a higher peak amplitude of EDA data. A different undesired behavior *growling voice* was associated with a lower peak amplitude of EDA data. Finally, users suspected that more severe undesired behaviors, such as hitting himself, might be associated with higher peak amplitudes in the EDA data.

3) Processing physiological data: Visualizing processed physiological data on top of the original physiological data visualization helps BEDA users identify noise or outliers in the original sensor data intuitively, and easily remove those to improve data quality. Fig. 4 shows the result of applying the ”smoothing.m” script (shown in Fig. 6-1) to remove noise from the selected range of the raw EDA signal via the *Timeline View*.

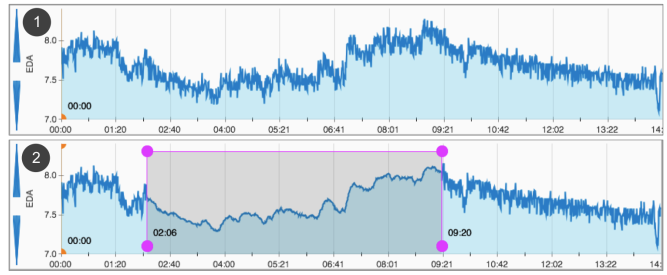

Fig. 4. (1) The raw EDA signal. (2) Selecting a range of EDA signal to clean noise in the signal by running a smoothing script.

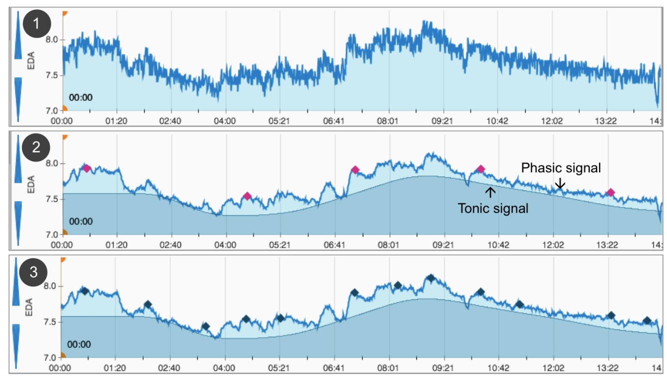

Fig. 5. A sequence of EDA data analyses: (1) The raw EDA signal. (2) The result after running a signal smoothing script, a script separating the EDA signal into tonic and phasic signals, and a script counting peaks with a parameter of 500 for the *minimum peak distance*. (3) The results with this *minimum peak distance* parameter adjusted to 100.

The visualization of sensor data analysis also helps users find the right parameter for physiological sensor data while iteratively experimenting with how the analyses alter the original sensor data source. In Fig. 5, the second image shows the resulting visualization using a parameter of 500 for the *minimum peak distance*. The third image shows the result after changing this parameter to 100. Using the different visualizations based on the parameters chosen, users can decide how granular they want their analysis to be. We provide interfaces in BEDA (Fig. 6-1) that allow users to manipulate algorithm parameters to attain more accurate data analysis results. To facilitate this process in a more intuitive way, we created an interactive horizontal slider bar interface next to the *Script Selection View* (Fig. 6-1) that appears every time the user selects a script that requires parameter modifications. Users can directly manipulate the parameter setting using the slider bar rather than editing the parameter in the script.

B. *Comparing Features Across Sessions*

The *Data Analysis View* (Fig. 6) visualizes the entire sessions' features in three different ways. The *Table View* (Fig. 6-2) displays the calculated feature vectors to verify whether each feature is calculated accurately. The *Heat Map View* (Fig. 6-3) visualizes the correlation values between pairs of features to help select a concise set of EDA features that best describe the original EDA data. The *Trends View* (Fig. 6-4)

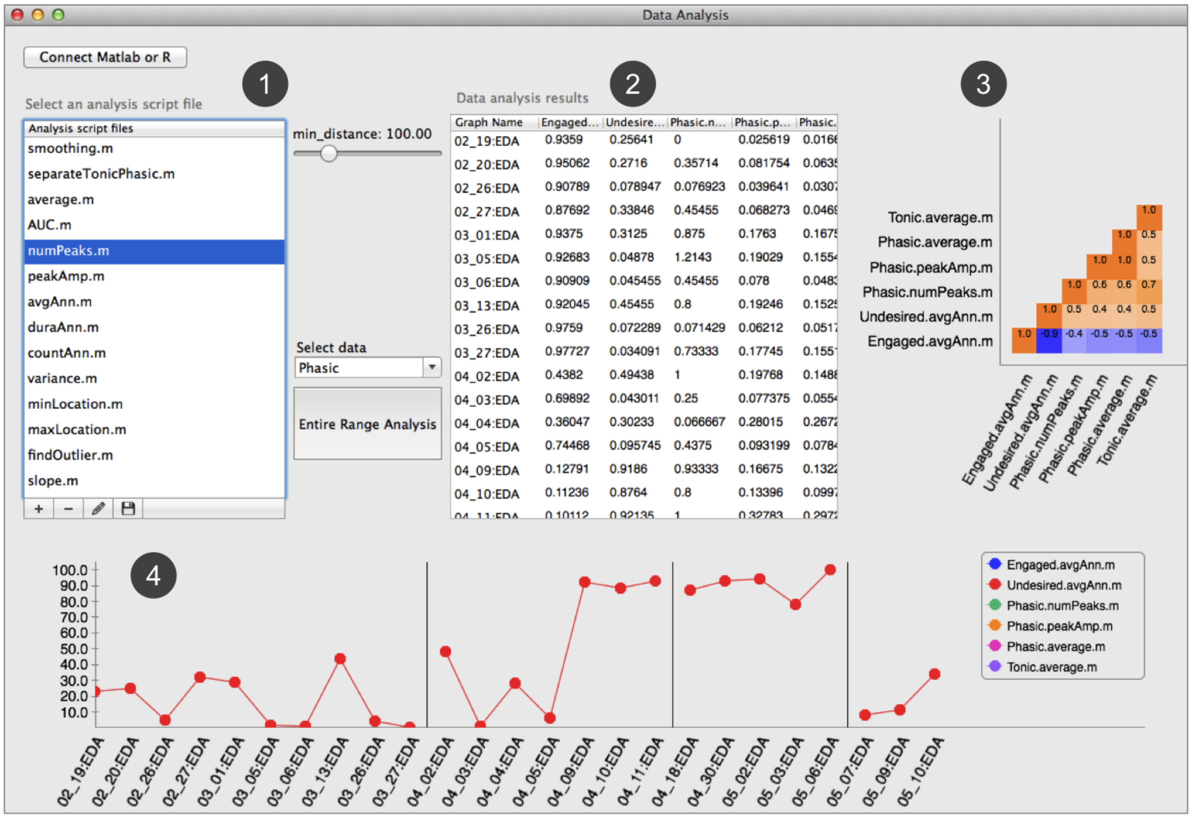

Fig. 6. Data Analysis View: (1) Scripts Selection View. (2) Table View. (3) Heat Map View. (4) Trends View.

shows how each feature changes across sessions or conditions. In the following section, we describe how the sub-tasks were conducted in more detail.

1) Calculating features: As we identified in the Section III-B, the proportion of undesired behaviors and engaged behaviors of each session are calculated for the behavior measures. EDA features are: tonic average, phasic average, phasic average peak amplitude, and phasic average number of peaks per minute for each session. To do this in BEDA, the scripts access each coded behavior and processed EDA signal (tonic EDA, phasic EDA, and phasic EDA peaks) visualized in the *Timeline View* (Fig. 3). Then, using MATLAB functions, the scripts calculate each feature and show the feature vectors in the *Table View* (Fig. 6-2).

2) Identifying a concise set of features: A heatmap is a popular graphical representation for visualizing the relationships between features [36]. Every time features are calculated in BEDA, Pearson correlation coefficients between pairs of features are visualized in the *Heat Map View* (Fig. 6-3). The correlation coefficients range from -1 to 1, where a value of 0 indicates no relationship between the two features. We used a diverging blue-orange color scheme [37] that emphasizes the zero midpoint in white, positive correlation in orange, and negative correlation in blue. A darker orange corresponds to a higher positive correlation; a darker blue corresponds to a higher negative correlation. For example, Fig. 7-1 shows that the phasic EDA average and the phasic EDA peak amplitudes have a strong positive correlation of 1.0, implying the two features represent the same characteristics from the original data. After removing one of the two strongly correlated features (phasic EDA average), lighter orange squares appear in the *Heat Map View* (Fig. 7-2), indicating these features have moderate correlation with each other. Finally, Fig. 7-3 shows the result of adding the behavior measures to the *Heat Map View*. The undesired and engaged behavior measures have a strong negative correlation of -0.9. Each EDA feature has moderate positive correlation (0.4 or 0.5) with the undesired behavior measure, depicted in light orange, and negative moderate correlation (-0.4 or -0.5) with the engaged behavior measure, depicted in light blue.

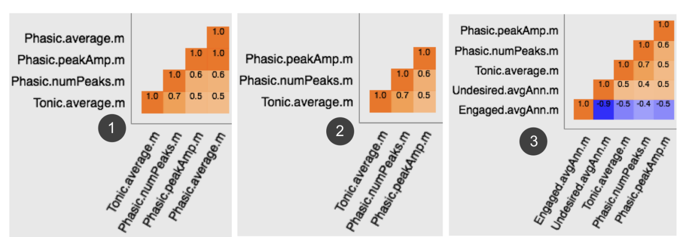

Fig. 7. The *Heat Map View*: (1) Correlation values between every pair of EDA features are visualized. (2) The phasic EDA average feature is removed because the correlation between the phasic EDA average and the phasic peak amplitude was very high 1.0. (3) Adding behavior features to understand which physiological features are highly correlated with the behavioral features.

3) Visualizing trends of features: The *Trends View* allows users to examine how the level of undesired behaviors and EDA features change when a child wears a pressurized vest. For example, Fig. 8-1 shows that wearing a pressurized vest did not decrease undesired behaviors, contrary to our original hypothesis. The EDA features, however, did not show any consistent pattern across sessions. To understand the inconsistent pattern of EDA features, users explore the outlier sessions identified in the *Trends View* (Fig. 8) in more detail using the *Timeline View* (Fig. 9).

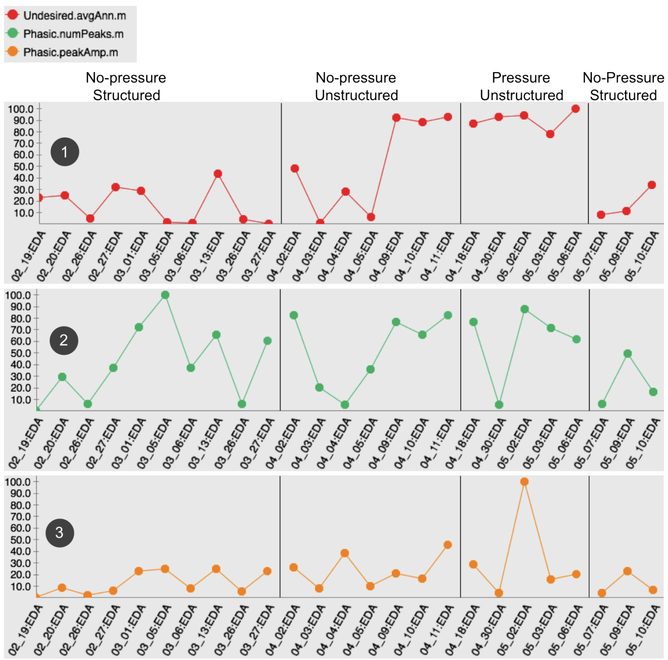

Fig. 8. The first *Trends View* showing level of undesired behaviors across sessions. The second and third *Trends View* each shows phasic EDA average number of peaks per minute and average peak amplitude across sessions. Clicking the legend interactively shows each graph.

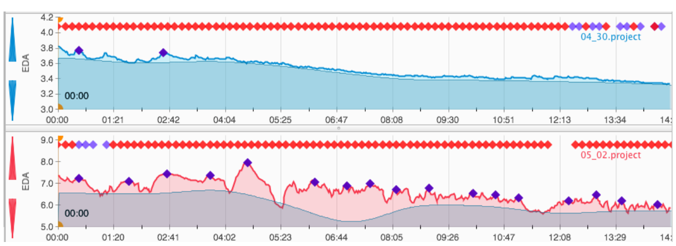

Fig. 9. The *Timeline View* of session 04_30 (blue graph on the top row) and 05_02 (red graph on the bottom row). Although the participating child showed frequent undesired behaviors in both sessions, the phasic EDA of 04_30 session has fewer frequent peaks and lower peak levels, compared to the 05_02 session.

Users identified that the participant's severe undesired behaviors (e.g., yelling) were associated with higher peaks in EDA compared to mild undesired behaviors (e.g., tapping knuckles on the table). For example, in Fig. 8-2, the phasic EDA average number of peaks per minute (i.e., frequency of sensory reactivity) in 04_30 session was significantly lower than in other sessions in the same condition of wearing a pressure vest. The presence of fewer frequent peaks suggests that the participant reacted less to immediate events, which may indicate fewer undesired behaviors. The proportion of undesired behaviors in that session, however, was very high. By returning to the *Timeline View* (Fig. 9), the users discovered that most of undesired behaviors in the 04_30 session were very mild undesired behaviors (e.g., tapping knuckles on table; saying words in growling voice), which explained why a small number of peaks were identified. In contrast, Fig. 8 shows that the 05_02 session had a significantly high peak amplitude compared to other sessions. By returning to the *Timeline View*, the researchers identified that the child engaged in more aggressive undesired behavior, such as screaming, explaining the higher peak average in phasic EDA (Fig. 9).

C. Extensible and Transparent Data Analysis Methods

We resolved the limited extensibility of previously available software by making use of MATLAB and R scripts. Scripts are a series of MATLAB and R functions that can process the input data to get desired outputs. As we described in Section V-A and Section V-B, the scripts can access any physiological or behavior data visualized in the *Timeline View* as input and visualize analysis results as output on either the *Timeline View* or the *Data Analysis View*. Sharing analysis scripts is a common practice among computer science researchers, as it helps to ensure the reliability of their research by using the same analysis method.

1) Transparent and extensible data analysis: We make the analysis script visible to help users better understand the exact process of each algorithm. For example, clicking the name of "numPeaks.m" script (Fig. 6-1) shows the script content in Fig. 10. Further, all of the scripts are easy to find and appear in the same display (Fig. 6-1). Novice programmers can easily select and apply the analysis scripts on their data.

```
%-*- input: vector
%-*- output: number annot
%-*- parameter: min_distance 0 500 1000

data = beda_input('vector');
min_distance = beda_parameter(1);
[y,x]= findpeaks(data, 'MINPEAKDISTANCE', round(min_distance));
beda_output(length(x));
beda_annot([x, y]);
```

Fig. 10. The content of "numPeak.m" script: *beda_input* reads EDA data visualized in the *Timeline View* and analyzes the input data using the findpeak function in MATLAB. The script writes the functions output value in the *Table View* (Fig. 6-2) through *beda_output* and visualizes the identified peaks using *beda_annot* in the *Timeline View* (Fig. 3)

BEDA currently embeds MATLAB and R scripts that are specifically required for analyzing EDA data and behavior coding for momentary time sampling. Because many open source MATLAB and R communities exist on-line, users can download, modify, and apply these functions in their sensor data analysis.

2) Extensible data analysis with machine learning: In our case study, two researchers manually coded the behavior data and compared the coded behaviors to verify for accuracy. In practice, training additional researchers in the coding process, manually coding behaviors, and validating the coded behaviors is time intensive. Recently, machine learning algorithms have been explored to predict target behaviors or events using physiological data [2], [4]. Such algorithms may be a viable option for accelerating the validation process of the coded behaviors. To illustrate, we trained a supported vector machine (SVM) in MATLAB to predict the coded undesired behaviors using EDA features. As input of SVM, we used EDA features (e.g., tonic EDA average, phasic EDA average etc.) in every

10-second time window. The 10-second window was chosen because behavior was coded every 10 seconds in this study. Each window was classified into one of the following two classes: 1) annotated as an undesired behavior and 2) no undesired behavior code. Then, all EDA features and classes are randomly partitioned into ten folds and the behaviors were cross-validated by the trained SVM. We achieved 84.03% accuracy (precision: 0.88, recall: 0.81). The predicted behaviors are highlighted in the *Timeline View*. Interestingly, the false positive predictions – the time windows where the machine learning algorithm predicts an undesired behavior will occur, yet no undesired behavior code exists – can help users explore additional behaviors that exhibit EDA characteristics similar to those of the explicitly defined undesired behaviors. For example, the participant in this study would often push his chair back from the table where he was working on assigned tasks. This behavior was not included in the definition of undesired behaviors, but, using SVM, we were able to identify this as a frequently occurring behavior that shows a similar pattern of changes in EDA as the EDA changes that occurred with the coded undesired behaviors.

VI. User Feedback

We invited two scientists to review how BEDA presents and facilitates analysis of the data the research team collected for the pressurized vest study. Then, we conducted an in-depth interview to collect feedback about the efficacy of our tool. Both participants were very familiar with the pressurized vest study — both participated in a weekly meeting to discuss data collection and analysis. P1 is a graduate student in Special Education. The pressure vest study was her first experience combining a physiological measure with behavioral analysis. P2 is a graduate student in Speech and Hearing Science who had previously conducted two studies combining physiological measures with behavioral analysis prior to participating in the pressure vest study.

A. Motivated Continue Research Combining Physiological Measures with Behavioral Analysis

P2 said our tool motivated her to continue conducting the studies combining physiological measure with behavioral analysis, "It [this tool] would motivate me to continue conducting the type of work that I was previously conducting. I was really frustrated by that whole syncing process between the EDA data and video. Because our research team decided to code emotional valence every 30 seconds, I wanted to make sure that we were coding the same 30 seconds [of sensor data that we were] watching in the video. [Using BEDA], I would be more likely to use the sensors again with video data and then conduct behavioral coding as well."

She also emphasized how this type of study is especially beneficial for people who have limited verbal abilities. Getting additional information from physiological data could inform behavioral scientists, clinicians, parents, and caregivers about the emotions of individuals who have difficulties in communicating their feelings.

B. Raised Confidence when Analyzing Sensor Data

P1, who was not previously familiar with sensor data analysis, found that she was more confident in conducting sensor data analysis research when using BEDA. Given that the sensor analysis features are embedded in BEDA, she stated that she would be, "more comfortable, interested, and willing to conduct research with sensor data" if she were using our tool. "I don't know anything about signal processing and how to use R or MATLAB to write code to process signals. With BEDA, I can do sensor data analysis on my own, see how this analysis altered the sensor data, and understand them while looking at the behavioral coding and identify things that are happening. It makes multimodal data analysis doable for me as a behavioral scientist who has no skills or training in sensor signal processing or analysis."

C. Validated Behavioral Coding and Facilitated Secondary Analysis

P1 said BEDA is useful for, "understanding underlying causes for behavior and validating some of the behavioral coding that we already have." The behavioral scientists had defined the categories of behaviors that they coded in this study based on behavioral observation and teacher and therapist reports. For example, undesired behaviors for this child (screaming, leaving his seat, tapping the table with his knuckles, using a growling voice, etc.) were hypothesized to be correlated with high arousal level. However, they did not know to what extent these behaviors would affect his internal status and/or correspond to internal levels of arousal. Using BEDA's *Timeline View*, P1 easily evaluated whether the behavioral codes the team had defined were actually related to EDA data changes. Many times, an increase in EDA levels corresponded with the predefined undesired behaviors. However, when she found a similar increase in EDA levels when no behavioral coding coincided, BEDA allowed her to access the corresponding video directly and identify what behavior was occurring when the EDA level increased (e.g., the child refusing his task) or what environmental conditions may have been affecting arousal (e.g., loud noise). She indicated that including this newly identified behavior in subsequent analyses may help the team to better examine under what conditions the child engaged in on-task and undesired behaviors.

D. Saved Time in Organizing and Cleaning Data

Both P1 and P2 identified that BEDA was useful for organizing and aligning different modalities of data across multiple sessions. P1 said "Instead of having twenty Microsoft Word files containing the behavioral data for each session, the twenty video files representing each session, and the twenty sensor data files representing each session, BEDA allowed us to have a single file containing all of the data for the study aligned and synchronized. This reduced both the confusion and amount of time required for managing the data."

Previously, cropping sensor signal data to delete invalid data or outlier data was overwhelming. Previously, the researchers opened a ".csv" file using Microsoft Excel and manually deleted unwanted in the signal. With BEDA, she appreciated that she could easily crop the sensor data visually by directly demarking the start and end points.

VII. Conclusion

We presented a tool that supports visual analysis of behavior and physiological sensor data within and across multiple sessions. With a participatory design study team that included five behavioral scientists, a clinician, and three computer scientists, we identified the challenges behavioral scientists encountered in behavior and physiological sensor data management and analysis during a case study exploring the efficacy of pressure vests. Throughout this study, we developed a tool, BEDA, to support the management and analysis challenges. BEDA visualizes video footage, sensor data streams, behavioral codes, and sensor data analysis together in the *Timeline View* to facilitate the identification of patterns between analyses and within and across multiple study sessions. Data analysis became extensible by integrating scripts that apply any personally written or publicly available function in MATLAB and R to analyze physiological data with respect to behavior. Our results from a pilot study of BEDA found that BEDA motivated behavioral scientists to continue combining physiological measures with behavioral analysis. Without this tool, they would have stopped collecting these data. It also helped researchers identify anomalies in the data that may have otherwise been overlooked.

ACKNOWLEDGMENT

This work is supported by NSF grant CCF-1029679. We would like to thank the children and their families who made this study possible.

REFERENCES

[1] M. S. Lauer, P. M. Okin, M. G. Larson, J. C. Evans, and D. Levy, "Impaired heart rate response to graded exercise prognostic implications of chronotropic incompetence in the framingham heart study," *Circulation*, vol. 93, no. 8, pp. 1520–1526, 1996.

[2] M.-Z. Poh, T. Loddenkemper, C. Reinsberger, N. C. Swenson, S. Goyal, M. C. Sabtala, J. R. Madsen, and R. W. Picard, "Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor," *Epilepsia*, vol. 53, no. 5, pp. e93–e97, 2012.

[3] J. Choi, B. Ahmed, and R. Gutierrez-Osuna, "Development and evaluation of an ambulatory stress monitor based on wearable sensors," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 2, pp. 279–286, 2012.

[4] H. Banaee, M. U. Ahmed, and A. Loutfi, "Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges," *Sensors*, vol. 13, no. 12, pp. 17 472–17 500, 2013.

[5] D. Schuler and A. Namioka, *Participatory design: Principles and practices*. L. Erlbaum Associates Inc., 1993.

[6] *MATLAB, http://www.mathworks.com/products/matlab/*.

[7] *R, http://www.r-project.org/*.

[8] L. Nachman, A. Baxi, S. Bhattacharya, V. Darera, P. Deshpande, N. Kodalapura, V. Mageshkumar, S. Rath, J. Shahabdeen, and R. Acharya, "Jog falls: a pervasive healthcare platform for diabetes management," in *Pervasive Computing*. Springer, 2010, pp. 94–111.

[9] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor-based systems for health monitoring and prognosis," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 40, no. 1, pp. 1–12, 2010.

[10] P.-E. Aguera, K. Jerbi, A. Caclin, and O. Bertrand, "Elan: a software package for analysis and visualization of meg, eeg, and lfp signals," *Computational intelligence and neuroscience*, vol. 2011, p. 5, 2011.

[11] J. Hagedorn, J. Hailpern, and K. G. Karahalios, "Vcode and vdata: illustrating a new framework for supporting the video annotation workflow," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2008, pp. 317–321.

[12] *Nolduc, http://www.noldus.com/*.

[13] 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*. ACM, 2011, pp. 299–304.

[14] *Ledalab, http://www.ledalab.de/*.

[15] *Acqknowledge, http://www.biopac.com/acqknowledge-data-acquisition-analysis-software-win*.

[16] W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski, "Visual methods for analyzing time-oriented data," *Visualization and Computer Graphics, IEEE Transactions*, vol. 14, no. 1, pp. 47–60, 2008.

[17] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, "Liverac: interactive visual exploration of system management time-series data," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2008, pp. 1483–1492.

[18] R. Kincaid and H. Lam, "Line graph explorer: scalable display of line graphs using focus+ context," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2006, pp. 404–411.

[19] A. E. Kazdin, *Single-case research designs: Methods for clinical and applied settings* . Oxford University Press, 2011.

[20] B. Reichow, E. E. Barton, L. Good, and M. Wolery, "Brief report: Effects of pressure vest usage on engagement and problem behaviors of a young child with developmental delays," *Journal of autism and developmental disorders*, vol. 39, no. 8, pp. 1218–1221, 2009.

[21] T. Grandin, "Calming effects of deep touch pressure in patients with autistic disorder, college students, and animals," *Journal of child and adolescent psychopharmacology*, vol. 2, no. 1, pp. 63–72, 1992.

[22] A. R. Deris, E. M. Hagelman, K. Schilling, and C. F. DiCarlo, "Using a weighted or pressure vest for a child with autistic spectrum disorder," *ERIC Document Reproduction Service No. ED490780*, 2006.

[23] N. L. VandenBerg, "The use of a weighted vest to increase on-task behavior in children with attention difficulties," *American Journal of Occupational Therapy*, vol. 55, no. 6, pp. 621–628, 2001.

[24] M. Zimmer, L. Desch, L. D. Rosen, M. L. Bailey, D. Becker, T. P. Culbert, H. McClafferty, O. J. Z. Sahler, S. Vohra, G. S. Liptak *et al.*, "Sensory integration therapies for children with developmental and behavioral disorders," *Pediatrics*, vol. 129, no. 6, pp. 1186–1189, 2012.

[25] *Q Sensor, http://www.affectiva.com/*.

[26] W. Boucsein, *Electrodermal Activity*. Springer New York Dordrecht Heidelberg London, 1992.

[27] J. L. Andreassi, *Psychophysiology: Human behavior & physiological response*. Psychology Press, 2000.

[28] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," *Handbook of psychophysiology*, vol. 159, 2007.

[29] M. R. Snodgrass, J. G. Kim, K. Karahalios, M. Borden, and J. W. Halle, "The effects of a pressure vest on task engagement, challenging behavior, and a physiological measure of stress for a child with intellectual disability," University of Illinois at Urbana-Champaign, www.ideals.illinois.edu, Tech. Rep., 8 2015.

[30] C. H. Kennedy, *Single-case designs for educational research*. Prentice Hall, 2005.

[31] C. D. Wickens and C. M. Carswell, "The proximity compatibility principle: its psychological foundation and relevance to display design," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 3, pp. 473–494, 1995.

[32] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer, "Wrangler: Interactive visual specification of data transformation scripts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 3363–3372.

[33] I. Guyon, *Feature extraction: foundations and applications*. Springer Science & Business Media, 2006, vol. 207.

[34] *Xcode, https://developer.apple.com/xcode/*.

[35] *Core-plot, https://code.google.com/p/core-plot/*.

[36] J. Seo and B. Shneiderman, "A rank-by-feature framework for interactive exploration of multidimensional data," *Information Visualization*, vol. 4, no. 2, pp. 96–113, 2005.

[37] C. A. Brewer, "Color use guidelines for data representation," in *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 1999, pp. 55–60.