Interactive Relevance Search and Modeling: Support for Expert-Driven Analysis of Multimodal Data

Chreston Miller
Center for HCI, Virginia Tech
2202 Kraft Drive,
Blacksburg,Va. 24060, USA
chmille3@vt.edu

Francis Quek
Center for HCI, Virginia Tech
2202 Kraft Drive,
Blacksburg,Va. 24060, USA
quek@vt.edu

Louis-Philippe Morency
Institute for Creative
Technologies, USC
Los Angeles, CA 90094, USA
morency@ict.usc.edu

ABSTRACT

In this paper we present the findings of three longitudinal case studies in which a new method for conducting multimodal analysis of human behavior is tested. The focus of this new method is to engage a researcher integrally in the analysis process and allow them to guide the identification and discovery of relevant behavior instances within multimodal data. The case studies resulted in the creation of two analysis strategies: Single-Focus Hypothesis Testing and Multi-Focus Hypothesis Testing. Each were shown to be beneficial to multimodal analysis through supporting either a single focused deep analysis or analysis across multiple angles in unison. These strategies exemplified how challenging questions can be answered for multimodal datasets. The new method is described and the case studies' findings are presented detailing how the new method supports multimodal analysis and opens the door for a new breed of analysis methods. Two of the three case studies resulted in publishable results for the respective participants.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*relation systems, temporal logic*

General Terms

Algorithms, Experimentation

Keywords

Structural Model Learning, Temporal Behavior Models, Model Evolution, Human-Machine Cooperation, Temporal Event Data, Model Discovery

1. INTRODUCTION

There is a multitude of annotated behavior corpora (manual and automatic annotations) available as research expands in multimodal analysis of human behavior [5, 13, 17, 25]. Many of these corpora and supporting visualization tools store, organize, and display multimodal data based on the structural nature of behavior. By structure we mean discrete events that hold ordered relations in time that may vary between occurrences. For example, the visualization tools MacVisSTA [29], ANVIL [15], and EXMarALDa [33] display multimodal data as interval events with support for continuous signal data. The creation of such corpora and the visualization of their content has received much attention in recent years [5, 14, 17, 34]. Data visualization enables a researcher (e.g., a domain expert with a deep understanding of the science relating to the data) to gain insight into her data and so indirectly supports analysis. However, aiding researchers in the *actual analysis process* of such multimodal behavior corpora has gained less attention.

Our main motivation is to *include* the expert in the analysis. As noted in [10] in the related field of cyber security, domain expertise is a critical element that should be incorporated in a solution. Normally multimodal corpora analysis either separates the expert from her data (e.g., statistical analysis or machine learning) or only engages the expert 'after-the-fact' in a tedious task of scrubbing through annotated/raw data. It is important to include the expert as an integral part to gain the advantage of their knowledge and expertise. This may appear common sense, but is surprisingly overlooked.

There are several nuances of human behavior that make their analysis challenging and thus require a certain level of expert involvement. *First*, human behavior is variant. The idea represented by a behavioral interaction, e.g., a greeting between two individuals, may be exhibited many different ways in the data making identification difficult. *Second*, every observed behavior has the potential to be relevant to an expert depending on his/her analysis goal(s). Hence, there is no concept of "noise" but rather one of relevance. *Third*, a behavior's value to an expert may not be based on frequency or statistical significance but on subjective relevance. *Lastly*, for experts, there is no training data to build classifiers as the behavior sought may vary greatly or the behavior(s) of interest may not be known yet. They leverage their knowledge to identify and discover what is relevant. These enumerated nuances have been observed in multimodal behavior analysis [6, 18, 19, 27, 28]

In response to these challenges we present Interactive Relevance Search and Modeling (IRSM), a new analysis method that includes the expert by supporting them in interactive exploration and analysis of their data. Experts have background knowledge, experience, and expertise that allow them

to identify meaningful *behavior phenomena*. A *behavior phenomenon* represents a specific behavior of interest, such as a greeting between two individuals (A and B) in a social setting of a particular culture. A *pattern* represents a particular formulation of a *behavior phenomenon* in the data, e.g., A <walks up to> B [within 1 second] A <shakes hand> B and A <says "Hello">. The technical definition of a *pattern* is from [21]. We engage the expert in an interactive data-driven discovery process to evolve a *pattern* in compliance with a desired formulation and discover occurrences of *behavior phenomena*. This results in *pattern(s)* representing *behavior phenomena* as they exists in the data and reflecting the expert's knowledge.

Through three longitudinal case studies, we observed how IRSM produces an analysis strategy of hypothesis testing with two variations. The strategy produced shows how challenging questions can be answered for multimodal datasets. The focus of this study is to produce in-depth insights on real-world case studies. Hence, the longitudinal nature of our study with a small participant pool.

The rest of the paper is outlined as followed. In Section 2 we describe related work. Section 3 provides an overview of IRSM. The approach for evaluation of IRSM is discussed in Section 4. Then in Section 5, the details of our case studies are presented. This includes the demographics of our participants and descriptions of their datasets. The methodology of our case studies is presented in Section 6. Following this, we discuss participant feedback (Section 7) and the two analysis strategies developed (Section 8). Then in Section 9 we discuss the impact of our case studies and present other resulting observations. Section 10 then concludes the paper with future work and closing remarks.

2. RELATED WORK

Our data domain is multimodal. There has been a strong trend toward creation and analysis of multimodal corpora. This is no surprise as the authors of [28] argue that deeper understanding of multi-modality is beneficial to the analysis of human behavior. Many multimodal corpora have been created in response to this observation which predominantly consist of sequences of descriptive events. The VACE/AFIT [6] multimodal meeting corpus is a detailed recording of multiple sessions of Airforce officers partaking in war gaming scenarios in a meeting setting. The Semaine corpus [17] is a collection of emotionally colored conversations. The Rapport and Face-to-Face corpora [13, 25] are sets of speaker-listener interactions. One of the largest corpora to date is the AMI corpus [5] which contains 100 hours of recorded meetings. Mörchen created a series of datasets of varying degrees of modalities [24]. These corpora and datasets are key exemplars of a growing research community interested in such data.

With the increasing number of multimodal datasets, tools are needed to visualize the data for analysis. These tools have been developed to visualize multi-channel annotations coupled with varying degrees of multi-channel support of audio and video. Well known examples of these tools are MacVisSTA [29], ANVIL [15], Theme [1], EXMARaLDA [31, 9], ELAN [38], C-BAS[1], Transformer [2], and VCode [14]. The AMI corpus uses a different approach through use of the Nite XML toolkit which provides extensive support for complex annotation representation and supportive interface. The Nite XML toolkit visualization is centered around transcription text of a corpus being annotated and is linked to supportive media, e.g., audio or video.

Besides tool support, there is the need to support the expert in analysis of their data. This has become the focus of a large research area: visual analytics, especially for intelligence analysis. There are many tools developed specifically for intelligence analysis that are beyond this paper's scope, however, the goal of these tools is in line with IRSM. They are designed to engage the expert in the analysis process allowing application of their knowledge. Example work can be seen in cyber analytics, visual analytics, and intelligence analysis [4, 10, 12, 35, 37]. Such ideas are related to Relevance Feedback (RF) [30] which iteratively applies the user's input to intermediate results, and feeds their input back into automatic processing algorithms to produce the next set of results. IRSM differs from simple RF in that it allows the expert to apply structural learning principles (as opposed to parametric) to identify how *behavior phenomena* actually occur in the data (*patterns*).

Contributions: Other well-known data discovery approaches exists, such as T-patterns [16] (used in Theme [1]), FEM [26], and intelligence analyst (IA) software, e.g., Jigsaw [37]. All are actively used. However, IRSM is different with regards to several aspects:

- **Semi-intervals**: IRSM uniquely leverages semi-intervals to provide a more flexible pattern representation.
- **Infrequent patterns**: While T-patterns allows the discovery of patterns based on statistical methods and FEM identifies patterns based on frequency thresholds, IRSM accounts for when a pattern is not based on either of these, e.g., infrequent or anomalies. It simply searches for the existence of *patterns*.
- **Multi-visualization**: IRSM can view different combinations of patterns. The user can define/evolve multiple *patterns* and explore/analyze the data based on them. One case study participant familiar with IA software noted this ability is something she had not seen before.

3. IRSM DESCRIPTION

This section contains an overview of Interactive Relevance Search and Modeling (IRSM). In many multimodal corpora, behaviors are encoded as annotated event intervals with temporal order being implicitly or explicitly defined. We approached the problem by looking at *patterns* of annotated events within multimodal data. As stated earlier, a *pattern* represents a particular formulation of a *behavior phenomenon* in the data. An example *pattern* is a greeting among two individuals with the possible formulation (from earlier): A <walks up to> B [within 1 second] A <shakes hand> B and A <says "Hello">. Here one could potentially 'evolve' the *pattern* by successively adding/removing relationships with other events, and/or pruning relational connections. However, evolving this pattern without guidance is a large search space even for a small *pattern* [23].

Our solution is founded on creating a formalism of a *pattern* based on structure, timing, and ordered relationships. We operate on a pattern at the semi-interval level (start or end of an interval). This representation was first introduced in [11] and successfully used for unsupervised pattern mining in [24]. Semi-intervals allow for the representation of partial

[1]Developed at Arizona State, http://www.cmi.arizona.edu/, but the url for C-BAS is broken.

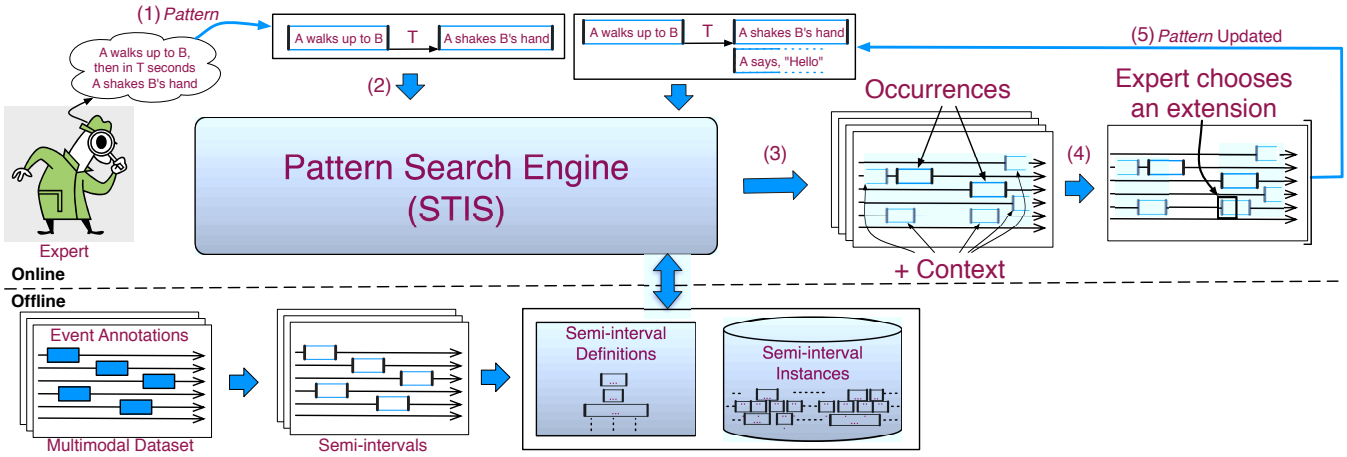

Figure 1: Interactive Relevance Search and Modeling (IRSM) overview. Offline, from a multimodal dataset create a semi-interval set organized as *definition* and *instance* look-up tables. Online, an expert provides an event sequence that is converted into a *pattern*. This *pattern* is given to Structural and Temporal Inference Search (STIS) [21] which searches to identify *pattern* occurrences. A set of occurrences is returned with the context of each occurrence. Based on the context information, the expert chooses an extension to the *pattern* and restarts the procedure with the newly updated *pattern*.

or incomplete knowledge of temporal relationships as operations are defined on parts of an interval and not the whole. *Patterns* are evolved at the semi-interval level, which we call a *1-step* change, representing the smallest change that can occur. IRSM was motivated by observations of experts viewing/analyzing data with the focus of their analysis relying on temporal relationships between events, their order, and the conceptual meaning of event order [6, 18, 19, 27, 28].

IRSM is data-driven to help constrain the generation of alternatives and produce a convergence to a desired *pattern*. We engage the expert in an interactive data-driven discovery process to evolve a *pattern* to a desired formulation that matches relevant *behavior phenomena* occurrences in the data. An overview of our method can be seen in Figure 1. Offline, given a set of event annotations (e.g., from a multimodal data-set), create a semi-interval set organized in a database of definitions and instances, i.e., look-up tables.

The online process is based on prior work that has been functionally tested and published ([23, 21]) but not human tested until the case studies presented in this paper. The 5 steps of Figure 1 describe an iterative and interactive process based on [23] in which *pattern* evolution is performed to identify particular *patterns* within media streams. The overall idea is to allow an expert to provide a starting point (seed) and evolve the seed to match relevant occurrences. The expert provides a seed event sequence (1) representing a hypothesis that is converted into a *pattern*. Then, (2) leverage the search and identification abilities of Structural and Temporal Inference Search (STIS) [21]. STIS allows us to accurately identify *pattern* occurrences and context (3). A graphical representation of the *pattern* is provided in which the user can view the context of each occurrence along with an overview of where the occurrences occur in the annotated data. The context provides potential extensions to the *pattern* supporting multimodal analysis of events within context (initially investigated in [22]). The expert also has the option to view the original source data (e.g., video) from which their annotations were created allowing a visual coupling between annotations, *pattern* occurrences, and the original data to verify the relevance of occurrences and place them in the original context. The expert then chooses a relevant extension (4), the *pattern* is updated (5), and the process iterates. At any time, the expert can load different *patterns* (one at a time) to compare their occurrences, contexts and evolve each. This allows exploring and analyzing the data using multiple *patterns*.

In (1) and (5), the expert also is provided the ability to define/adjust a *pattern* using relational logic similar to the relationship principles of Allen [3], Freksa [11], and regular expressions in a graphical query system. The *pattern's* structure can be adjusted according to the expert's guidance. This also includes using variables and wildcards in defining parts of a *pattern*, e.g., X <looks at> Y or X <*> Y (i.e., X performs some action toward Y). Variable binding is applied to discover values bound to *, X, and Y, respectively.

In summary, IRSM provides the ability to focus a search in multimodal data and evolve search parameters (evolve the *pattern*) and identify how *behavior phenomena* actually exist in the data. IRSM was designed to aid experts in multimodal analysis, however, the design also took into accunt that the user may not be an expert. Expertise is beneficial in the quality of analysis and is not tied to IRSM's functionality. IRSM simply provides a view of the data, search capabilities into the data, and *pattern* evolution for analysis.

4. EVALUATION

The accuracy of the respective pieces of IRSM have already been quantitatively tested. STIS was extensively tested against two Frequent Episode Mining (FEM) algorithms [26] in [21] where STIS was shown to perform comparable or better. FEM algorithms operate on interval and point data in order to identify event sequences of varying lengths (episodes) given certain timing restrictions between events and identify these episodes given a threshold (frequency or statistically based). The *pattern* evolution process of [23] (steps 1-5 in Figure 1) was extensively tested against a FEM algorithm and shown to perform comparable or better.

Our goal is to evaluate IRSM with real-world longitudinal case studies. The strength is to show that IRSM can address the wide differences in use and research requirements of researchers, i.e., apply IRSM to projects from real researchers and not a streamlined lab environment. In order to show

Table 1: Use-Case Participant Demographics

Participant	Gender	Research Experience	Previously Conducted Data Analysis	Previously Used Data Analysis/Visualization Software
P1	Male	2 years	Study of self-report data through transcription and coding	Statistical packages (e.g., R and JMP) plus Excel
P2	Male	1.5 years	Statistical and visual, Spotfire	JMP, SAS, and Spotfire
P3	Female	2.5+ years	Text analytics, geospatial, quantitative analysis, multimedia analysis, social network	Jigsaw, IN-SPIRE, Canopy, Palantir, Force-SPIRE, Analyst's Workspace, Excel, JMP, MySQL, Tableau/Eureka, Spotfire, Light_SPIRE

Table 2: Case Study Datasets' Contents Overview

	Semi-intervals	Unique Semi-intervals	Channel Min	Channel Mean	Channel Max
P1	2218	252(max)	7	14.22	23
P2	2784	6	3	3.83	4
P3	8545	25(max)	10	12.43	14

this, a real-world situation is required. Given the potentially infrequent nature of *behavior phenomena*, a more controlled study may not be representative of the analysis IRSM is meant to support. Hence, three real-world case studies were conducted for the evaluation of IRSM. These case studies represent a large endeavor given their longitudinal nature (details in next section).

5. CASE STUDY DETAILS

A set of three longitudinal case studies were conducted where the first author worked closely with three researchers interested in analyzing their own multimodal datasets. These datasets consisted of unimodal and multimodal temporal event data describing human behavior and interaction. In this section, we will describe in turn the researcher demographics and the details of their datasets.

5.1 Demographics

Three researchers from the Center for Human-Computer Interaction at Virginia Tech were independently recruited and had no prior knowledge of IRSM or its capabilities. We recruited them by offering a beneficial new way of analyzing their data that differed from other approaches. The demographics of the researchers can be seen in Table 1. Each participant had conducted research for at least 1.5 years using some form of data analysis prior to working with us. These prior analyses were conducted using standard analysis techniques and software packages.

5.2 Datasets' Descriptions

In this section we describe the datasets of each researcher who participated in our case studies. For simplicity, we will refer to them as P1, P2, and P3. Each participant's data needed some formatting before our system could use it. We developed a separate program that would take as input the participant's data in various forms and output the representation (SQLite database) that IRSM uses.

Data Characteristics: The data of each participant are multi-channel events represented by time points and/or intervals (or a mixture), i.e., multimodal data [22, 32, 36]. The events were annotated from their media sources either automatically and/or manually. Each channel represents events of a certain type (i.e., event type). This is illustrated in Figure 2. Table 2 provides an overview of the participants' dataset contents. The table summarizes the total number of semi-interval events across all sessions, the maximum number of unique semi-interval types per session (i.e., alphabet of semi-intervals), and the minimum, mean, and maximum number of channels across all sessions.

Participant 1's data: P1 was studying collaborative behavior in a small group setting. P1's data consisted of 23 sessions with three participants in each given the task to collaboratively build a story from pictures to describe the design of a new dining hall. Each participant had their own laptop with a shared and private space for viewing and placing pictures. The participants took turns contributing to the shared space. Each session was video recorded and transcribed for contributing features to the story. The data of each session consists of a sequence of events depicting when each participant (A, B, and C) contributed a feature. P1's analysis focus was on identifying interruptions/out-of-turn instances that exhibited collaborative behavior. **He hypothesized:** *that by looking at patterns of contributions that did not follow the simple turn-taking of the group, he could find evidence of collaboration among the participants.* P1 had previously performed quantitative analyses of feature contributions using statistical packages (see Table 1).

Participant 2's data: P2 was studying a new multi-scale interaction technique for large, high-resolution displays. The interaction technique consisted of using 1, 2, or 3 fingers on a trackpad to control the speed of the cursor, e.g., 1 finger is normal speed, 2 is faster, and 3 is fastest. There were 8 sessions each consisting of three trials where participants used a combination of 1, 2, or 3 fingers (according to personal choice) to reach targets that appear on the display. Once a target is reached, a new one appears elsewhere on the display. Each trial consisted of 17 targets. Event logging was used to record the different finger modes used. The events recorded from logging were used to create time sequential intervals representing the finger mode used at a given time for a given target. P2's analysis focus was identifying finger mode trends/behaviors among participants (*patterns*) that explain good/poor performance. **He hypothesized:** *that participants with good performance had more different finger mode behaviors than participants who did not perform as well.* P2 had previously performed quantitative statistical analysis in terms of participant speed, accuracy, error, and number of clutches (i.e., raising hand from trackpad). He also performed analysis through visualization of finger mode traces using Spotfire (`http://spotfire.tibco.com/`).

Participant 3's data: P3 was studying cooperative use of a large, high-resolution display for performing an intelligence analysis task. There were 7 sessions with 2 people per session sharing a large, high-resolution display, each with their own mouse. All sessions were video recorded where annotations were hand coded for apology events, possessive speech events, location discussion events, significant speech events, and events for re-finding either by computer or physically on the display. A mouse log was also created for each pair of participants, however, P3 chose to only process the mouse logs and create events for three of her sessions as she was unsure of their usefulness. P3's analysis focus was whether the display employed would be instrumental in facilitating common ground among each pair of participants. **She hypothesized**: *that the display would serve as a medium for common ground* [7, 8]. P3 had previously performed analyses of her data consisting of quantitative measures that included solution correctness for the intelligence analysis task and an analysis of mouse clicks to identify different interaction levels in sections of the display space. P3 also performed qualitative analyses through semi-structured interviews of her participants, manual video coding to identify situations of interest, and viewing periodic screenshots taken during each session to observe the use of display space.

Figure 2: Example of multimodal data (semi-intervals highlighted as vertical bold lines) with speech, gaze and gesture channels.

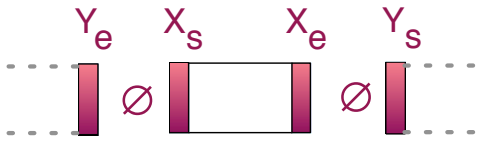

Figure 3: Master *pattern* created by P1. This *pattern* represents when Person X is the sole interruptor of Person Y ('s' and 'e' represent start and end semi-intervals, respectively, and ∅ represent absence of other semi-intervals).

6. METHODOLOGY

Our case studies' focus was how IRSM supports exploratory analysis. Each participant had his/her own goal and approached it through open-ended analysis. There were no predefined tasks as each case study was self-guided. This is a challenging scenario to support, but, the participants were very successful in addressing their respective analysis goals.

Two types of sessions (training and independent) were conducted in our case studies. The screen of the computer used was captured and event logging performed. After each session, a semi-structured interview was conducted to record the participant's experience. We started working with our participants 2-4 months before any sessions were held. This time was used to learn about each participants data and to transform it into the proper format.

IRSM was fully functional but a moderator was required as an operator. We were interested in testing IRSM (the method) and not the specific UI, hence, the moderator became part of the UI as the system operator. The moderator could take commands and requests and perform them, and, in doing so, fulfilled the functionality that allowed the participants to utilize IRSM. We created a *training script* based on earlier pilot studies so the participants could learn to use IRSM independent of the moderator's input. This lead to a set of training sessions and then independent sessions.

Training Sessions (TS): Three training sessions were conducted for each participant. The first consisted of going over the detailed training script designed to familiarize each participant with IRSM and its capabilities. Since this first TS was purely for training, no semi-structured interview was performed. Also, no screen capture was performed during the TSs. Each participant was also provided with a *feature list* for reference that summarized IRSM's capabilities. In the second and third TSs, each participant analyzed his/her data with minimal input and help from the moderator. The participants were allowed (and encouraged) to ask the moderator any questions they had.

Independent Sessions (IS): After training, the participants performed independent sessions where the moderator provided help only when it was deemed absolutely necessary. The sole purpose of the moderator was to run the system (be part of the UI) and take notes. Four ISs were run for each participant. The one exception was P2 who was satisfied with his results by the end of his second IS, and hence, only two ISs were run for P2. The TSs and ISs were conducted over a period of four weeks. Each TS and IS ranged from 30 minutes to 1 hour.

7. PARTICIPANT FEEDBACK

Due to space limitation, we briefly discuss the feedback from our participants gathered during the semi-structured interviews. More details are available in [20]. However, in the next section, we highlight the major result of our case studies. In these interviews we asked each participant about their experience using IRSM. We also asked them several 5-point Likert scale questions about the usefulness of IRSM's different features (Mean 4.76), how IRSM helped in discovery (Mean 4.5), and the relevance of identified *patterns* (Mean 4.75). Generally speaking, our participants favorably compared IRSM to previously used analysis approaches. They commented multiple times on how IRSM provided capabilities not available with other approaches and how IRSM's supportive view and exploration of the data helped them better understand their data. The main constructive criticism was the necessity to add different features and capabilities to improve IRSM's effectiveness, e.g., minor interface adjustments, search multiple datasets simultaneously, or perform more automation in identifying trends or anomalies; all of which have been addressed or are currently in development.

8. STRATEGIES DEVELOPED

The analysis strategies developed by our participants followed a common thread: hypothesis testing. Each would specify initial *pattern(s)* based on their initial hypothesis, use that to learn about their data, then re-specify the *pattern(s)* based on knowledge gained. Two variations of hypothesis testing were observed: Single-focus and Multi-focus.

8.1 Single-Focus Strategy

We call the first variation Single-Focus Hypothesis Testing. The steps for Single-Focus Hypothesis Testing are illustrated in Algorithm 1. Single-Focus Hypothesis Testing was used by **Participant 1** to discover how interruption/out-of-turn behavior existed in his data. In this variation, the focus is on identifying the structure and existence of one specific *pattern*. The identification of this *pattern* will provide evidence for the pertinent hypothesis. One starts by specifying an initial *seed pattern* based on an initial hypothesis about the data. Then, explore the data to see what matches are made to the *seed pattern*. The knowledge gained through exploration is used to update the *seed pattern* to improve its matching potential. The focus is on identifying relevant matches to this single *pattern*.

P1's hypothesis was that evidence of collaboration could be found through identifying *patterns* of contributions that did not follow simple turn-taking. The Single-Focus Hypothesis Testing strategy of IRSM helped P1 focus his analysis to mold a single *pattern* to identify such interrupting/out-of-turn occurrences. He began by creating an interjection model that represented when person X interrupts person Y within a certain timeframe. P1 learning more about his data and how this behavior can exist resulted in a "master *pattern*" (Figure 3 - red rectangles represent semi-intervals). This *pattern* represents when X interrupts Y (notice complete interval of X) where X is the sole interrupting event.

Algorithm 1 Single-Focus Hypothesis Testing Strategy

- Formulate an initial hypothesis to investigate.
- Specify an initial *seed pattern* based on initial hypothesis.
- Perform the following in a non-specified order:

repeat

  - Look at identified matches and investigate relevance (i.e., whether matches are related to the analysis goal, mis-codings, etc...).
  - Based on knowledge gained through exploration, respecify *seed pattern* and start exploring again.
  - When a relevant match is found, record the match with a descriptive label.

until Satisfied with *pattern*

- The result is a "master *pattern*" where focus is on exploration and investigation of matches to this *pattern*.
- The "master *pattern*" may need to be updated during exploration/identification in which repetition of some (or all) of the above steps will be conducted.

P1 was able to find the evidence he desired for this kind of collaboration, however, he noted more analysis is needed. Two excerpts from P1's final independent session summarize his impression of IRSM:

> " Anytime you take another look at your data you are going to learn something more so it's a nice opportunity to do that. A lot of my analysis of these data so far has been mostly quantitative and this [our approach] is relatively quantitative as well but I am able to translate some of my qualitative interests into these *patterns* we are searching for and that's pretty cool "
>
> "It has been helpful in finding evidence of collaboration for sure...that was my goal."

8.2 Multi-Focus Strategy

We call the second variation Multi-Focus Hypothesis Testing. The steps for Multi-Focus Hypothesis Testing are illustrated in Algorithm 2. Multi-Focus Hypothesis Testing was used by **Participant 2** and **3** to discover support for their respective hypotheses. In this variation, the focus is on identifying the structure and existence of a set of *patterns*. The identification of this set provides evidence for the pertinent hypotheses. One starts by specifying initial *seed patterns* based on an initial hypothesis about the data. Then, explore the data to see what matches are made to the *patterns*. The knowledge gained through exploration is used to update the *patterns* to improve their matching potential and create other *patterns* that can aid in identifying the desired *behavior phenomena*. This allows the "parallel" processing of multiple *patterns* during analysis (i.e., contributing multi-visualization discussed in Section 2). During this process a new hypothesis may be formulated and become the new focus dependent on what is discovered (arguably this could also occur in Single-Focus Hypothesis Testing). The focus is on identifying relevant matches to a set of sculpted *patterns*. Once a set of *patterns* is achieved, count and tabulate the recurrence of each *pattern*. The main difference in this variation is that the focus is on identifying the formulation and existence of a set of *patterns* instead of one *pattern*.

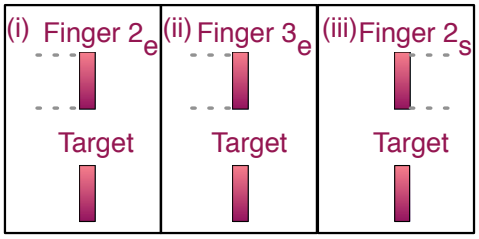

Figure 4: Example *patterns* used to finalize P2's analysis. (i) and (ii) represent when a participant reaches a target with either finger mode 2 or 3, respectively, i.e., finger mode interval ends at a target. (iii) represents the participant starting towards a new target (after reaching the previous one) using finger mode 2.

Algorithm 2 Multi-Focus Hypothesis Testing Strategy

- Formulate an initial hypothesis to investigate.
- Specify multiple initial *seed patterns* based on initial hypothesis.
- Explore the data based on the initial *seed patterns*.
- Based on knowledge gained through exploration, respecify *seed patterns* and continue exploring.
- During exploration, the participant may develop another hypothesis and then focus in that direction.
- Multiple *patterns* are specified and iterated over for each hypothesis.
- Re-specification/exploration of all *patterns* until a representative set of *patterns* are molded.
- Iterate through the representative set of *patterns* and count their recurrence in the data.

Participant 2's original hypothesis was focused on finding finger mode *patterns* that explained good performance. Using Multi-Focus Hypothesis Testing, P2 explored patterns to support this hypothesis. During his exploration he stumbled upon a new hypothesis that became his new focus. Example *patterns* for this new (and final) hypothesis are illustrated in Figure 4. This new hypothesis was looking for a correlation between finger mode duration and performance and reflected an interest in how participants approached a target (Figure 4i and ii) and how they began a new target once the previous one was reached (Figure 4iii). Multi-Focus Hypothesis Testing allowed P2 to focus his analysis on a set of molded *patterns* and count their recurrence. Unfortunately, his findings did not support his new hypothesis. Two excerpts from P2's final independent session summarize his impression of IRSM:

> "So all these sessions were pretty much part of one big exploration which was getting numerical values of my interest [based on defined *patterns* that were meaningful to him] and once all those numerical values were gathered, I tested my hypothesis and completed my hypotheses testing..."
>
> "...this process helped me in two things. First, kind of helped me to look at these values, these numbers in a different way in the sense that this tool broadened my horizons in the sense that I could look at more things that I couldn't normally see. Second good thing about this tool is

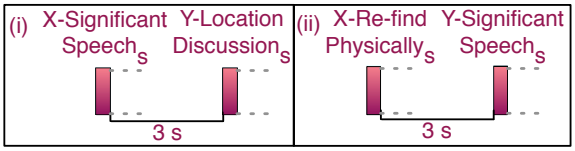

Figure 5: Example P3 *patterns*. (i) represents Person X starting a significant speech event within 3 seconds of Person Y starting a location discussion event. (ii) represents Person X performing physical re-finding on the display within 3 seconds of Person Y starting a significant speech event.

> that it allowed me to gather information that I needed for my hypothesis testing"

Participant 3's hypothesis was the display would serve as a medium for common ground. Using IRSM, and the Multi-Focus Hypothesis Testing strategy, P3 focused on learning about what was coded in her data, specifically events related to interactions with the display. Example *patterns* are illustrated in Figure 5. She focused on a specific set of *patterns* representing different interactions with the display. These included *patterns* representing significant speech following/preceding a location discussion of a document on the display (Figure 5i) or a physical re-finding on the display (Figure 5ii). The *patterns* P3 identified and counted at the end of her analysis explained why different groups performed better than others. This provided evidence for her hypothesis, but she comments that more analysis is needed. An excerpt from P3's final independent session summarizes her impression of IRSM:

> "...a lot of it was learning and exploring, not only about the tool's functionality but also my data. ...my *patterns* got a lot more specific as the time went on and I was able to direct...my approach. It also, this tool helped me find out what actually was going on in the data instead of me just guessing *patterns* in the beginning. So I'd say I'm a lot more comfortable with it now and I'm actually trusting the results that it gives me."

9. RESULTS AND DISCUSSION

Overall, our case studies demonstrated the promise and effectiveness of IRSM for multimodal behavior analysis. Each participant successfully utilized IRSM to facilitate their analysis goals. Notably, P1 and P3's analysis resulted in publishable results (currently unpublished). There were two important overarching results to our case studies. The *first* was that IRSM supports a researcher in multimodal analysis to approach their data with open-ended questions and seek answers. Our participants noted several times how the analysis approach possible with IRSM was different from anything they had ever done before and how beneficial IRSM was to their analysis. They were able to approach their data with open-ended questions and identify evidence to answer their questions. P1 noted that "...pretty quickly [he] was able to look for some evidence of collaboration". P2 commented that he "...could have done the same thing with a plain CSV file but it wouldn't be so fast, it would have taken some time to understand what is going on". Lastly, P3 said she had "...not been able to tackle this problem [evidence of common ground] before". Each participant had challenging questions of a conceptual nature and sought answers from a mixture of qualitative and quantitative data. As observed by P1, they were able to translate their qualitative interests into *patterns* and seek and find answers to their open-ended questions.

The *second* important result was how IRSM supports a researcher in multimodal analysis to focus their analysis at different levels. A researcher is able to focus on seeking a single *pattern*, going deep in the analysis with the *pattern*. We saw this accomplished through use of Single-Focus Hypothesis Testing. Or, the researcher is able to focus on a set of *patterns* and gather evidence from multiple angles. We saw this accomplished through use of Multi-Focus Hypothesis Testing. Note that Single-Focus Hypothesis Testing can be utilized for each *pattern* used in Multi-Focus Hypothesis Testing. Both strategies allowed the participants to explore their data at different levels and satisfy their analysis goals.

Other Observations: There were other results observed during the case studies. These are summarized here.

We observed that how a *pattern* was updated differed from the expected updating strategy. A *pattern* was intended to be updated *directly* from the context of *pattern* occurrences. However, *pattern* updating was mainly performed through re-specification of participants' initial *pattern(s)* iteratively until the desired *pattern* formulation(s) were achieved. Updates were informed through exploring the data and learning about its contents. Each iteration was used as a probe into the data to inform the creation of the next iteration. The view into the data furnished by IRSM was the major factor in informing *pattern* updating (evolution). This view can be described as visually providing the context of *patterns* and a visualization overview of the multimodal data channels common in many multimodal analysis tools.

Overall, our participants discovered the value of how IRSM allowed them to view and search their data. A positive artifact of IRSM was the identification of mis-codings and lack of annotations in datasets, showing another advantage of IRSM. This can be leveraged to identify areas of datasets where corrections need to be made or identify areas where finer detailed annotations are needed.

10. CONCLUSIONS

We presented three longitudinal case studies purposed to investigate a new analysis method. Our results provided evidence that IRSM facilitates multimodal behavior analysis and presents a new way of searching through temporal event data and performing analysis. We have developed an approach that aids in refining a *pattern* to match how occurrences actually exist in the data. This resulted in the creation of two analysis strategies: Single-Focus Hypothesis Testing and Multi-Focus Hypothesis Testing. Each were shown to be beneficial to multimodal analysis through supporting either deep analysis of a single *pattern* or analysis across a set of *patterns* to capture multiple angles in unison. Notably, two case studies resulted in publishable results.

The positive results demonstrated how IRSM merits further study in the areas of open-ended analysis scenarios, focused analysis scenarios, more diverse datasets, and researchers with more diverse backgrounds. We believe the results presented in this paper will support the creation of a new breed of analysis methods.

Overall, we see the results of our work as impacting the multimedia search, multimodal analysis, and behavior analysis communities through presenting a newly developed anal-

ysis method and results of a longitudinal, hands-on experience with researchers analyzing multimodal data. We hope this spurs further study of strategies and techniques for searching and identifying relevant *patterns* in multimodal data, and more generally, temporal event data.

11. ACKNOWLEDGMENTS

This research was partially funded by FODAVA grant CCF-0937133, NSF IIS-1053039, NSF IIS-111801, and the U.S. Army Research, Development, and Engineering Command (RDECOM). The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

12. REFERENCES

[1] *http://www.noldus.com/human-behavior-research/products/theme*, Checked: April, 2012.

[2] *http://www.oliverehmer.de/transformer/*, Checked: Dec, 2010.

[3] J. F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, 1983.

[4] C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *CHI '10*, pages 55–64. ACM, 2010.

[5] J. Carletta et al. The ami meeting corpus: A pre-announcement. In *MLMI*, volume 3869 of *LNCS*, pages 28–39. Springer Berlin / Heidelberg, 2006.

[6] L. Chen et al. Vace multimodal meeting corpus. *MLMI '06*, pages 40–51.

[7] M. Chuah and S. Roth. Visualizing common ground. In *Information Visualization '03*, pages 365–372.

[8] H. H. Clark and S. E. Brennan. Grounding in communication. In *Perspectives on socially shared cognition*, pages 127–149. APA, 1991.

[9] EXMARaLDA. *http://www.exmaralda.org*, Checked: May, 2012.

[10] G. Fink et al. Visualizing cyber security: Usable workspaces. In *VizSec '09*, pages 45–56, 2009.

[11] C. Freksa. Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54(1-2):199 – 227, 1992.

[12] C. Gorg et al. Visual analytics support for intelligence analysis. *Computer*, 99(PrePrints):1, 2013.

[13] J. Gratch et al. Creating rapport with virtual agents. In *Intelligent Virtual Agents*, volume 4722 of *LNCS*, pages 125–138. Springer Berlin / Heidelberg, 2007.

[14] J. Hagedorn, J. Hailpern, and K. G. Karahalios. Vcode and vdata: illustrating a new framework for supporting the video annotation workflow. In *AVI '08*, pages 317–321. ACM.

[15] M. Kipp. Anvil - a generic annotation tool for multimodal dialogue. In *Eurospeech*, 2001.

[16] M. Magnusson. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods*, 32:93–110, 2000.

[17] G. McKeown et al. The semaine corpus of emotionally coloured character interactions. In *ICME '10*, pages 1079 –1084.

[18] D. McNeill. Gesture, gaze, and ground. In *MLMI'06*, volume 3869 of *LNCS*, pages 1–14.

[19] D. McNeill et al. Mind-merging. In *Expressing oneself / expressing one's self: Communication, language, cognition, and identity*, 2007.

[20] C. Miller. *Structural Model Discovery in Temporal Event Data Streams*. PhD thesis, Virginia Tech, 2013.

[21] C. Miller, L. Morency, and F. Quek. Structural and temporal inference search (STIS): Pattern identification in multimodal data. In *ICMI*, 2012.

[22] C. Miller and F. Quek. Toward multimodal situated analysis. In *ICMI '11*.

[23] C. Miller and F. Quek. Interactive data-driven discovery of temporal behavior models from events in media streams. In *ACM MM*, 2012.

[24] F. Mörchen and D. Fradkin. Robust mining of time intervals with semi-interval partial order patterns. In *SIAM Conference on Data Mining (SDM)*, 2010.

[25] L.-P. Morency, I. de Kok, and J. Gratch. Context-based recognition during human interactions: automatic feature selection and encoding dictionary. In *ICMI '08*, pages 181–188. ACM.

[26] D. Patnaik, P. S. Sastry, and K. P. Unnikrishnan. Inferring neuronal network connectivity from spike data: A temporal data mining approach. *Scientific Programming*, 16(1):49–77, January 2007.

[27] F. Quek et al. Gestural spatialization in natural discourse segmentation. In *Spoken Language Processing*, 2002.

[28] F. Quek, T. Rose, and D. McNeill. Multimodal meeting analysis. In *IA*, 2005.

[29] R. T. Rose, F. Quek, and Y. Shi. Macvissta: a system for multimodal analysis. In *ICMI '04*, pages 259–264.

[30] Y. Rui et al. Relevance feedback: a power tool for interactive content-based image retrieval. *TCSVT*, 8(5):644–655, Sep 1998.

[31] T. Schmidt. The transcription system exmaralda: An application of the annotation graph formalism as the basis of a database of multilingual spoken discourse. In *Proceedings of the IRCS Workshop On Linguistic Databases*, pages 219–227, 2001.

[32] T. Schmidt et al. An exchange format for multimodal annotations. In *Multimodal corpora*, pages 207–221. Springer-Verlag, Berlin, Heidelberg, 2009.

[33] T. Schmidt and K. Wörner. Exmaralda – creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19, 2009.

[34] B. Schuller et al. Avec 2012: the continuous audio/visual emotion challenge - an introduction. In *ICMI '12*, pages 361–362. ACM.

[35] A. Singh et al. Supporting the cyber analytic process using visual history on large displays. In *VizSec '11*, pages 3:1–3:8. ACM, 2011.

[36] Y. Song, L.-P. Morency, and R. Davis. Multimodal human behavior analysis: learning correlation and interaction across modalities. In *ICMI '12*, pages 27–30. ACM, 2012.

[37] J. Stasko, C. Görg, and Z. Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[38] P. Wittenburg et al. Elan: a professional framework for multimodality research. In *LREC*, 2006.