Search Strategies for Pattern Identification in Multimodal Data: Three Case Studies

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

Francis Quek
Dept. of Viz., Texas A&M
3137 TAMU, College Station,
Texas 77843, USA
quek@tamu.edu

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

ABSTRACT

The analysis of multimodal data benefits from meaningful search and retrieval. This paper investigates strategies of searching multimodal data for event patterns. Through three longitudinal case studies, we observed researchers exploring and identifying event patterns in multimodal data. The events were extracted from different multimedia signal sources ranging from annotated video transcripts to interaction logs. Each researcher's data has varying temporal characteristics (e.g., sparse, dense, or clustered) that posed several challenges for identifying relevant patterns. We identify unique search strategies and better understand the aspects that contributed to each.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval—*Information Filtering, Search Process*

General Terms

Design, Experimentations

Keywords

Temporal Event Data, Search Strategies, Multimodal Search

1. INTRODUCTION

The analysis of multimodal event data benefits from meaningful search and retrieval. Such data is characterized by ordered, temporal events. A temporal event is an event segmented from a signal that has an associated meaning and an occurrence in time with respect to the signal. We have observed in the area of multimodal behavior analysis how identified relevant behavior occurrences (or patterns) have specific temporal and ordered characteristics [5, 14, 15, 27]. The timing and order of the events (and what the events represent) is very meaningful. The context they are found in also further informs the meaning. Other multimodal analyses also share this viewpoint [4, 9, 13, 30]. Hence, we are interested in understanding what strategies can be beneficial in searching such multimodal data as this has gained less attention. One successful tool, Interactive Relevance Search and Modeling (IRSM) [20], was developed in recent years to support searching multimodal datasets. Studying how researchers employ search tools, such as IRSM, can provide insights into beneficial search strategies.

In this paper, our goal is to analyze the search strategies developed by researchers performing multimodal data analysis. We have observed that it is not a simple case of black or white as there are a number of factors that can determine a good strategy. Is the data sparse, dense, clustered, or a mixture? Is the data interval or point data or both? What is the search goal or desired pattern structure? Through three longitudinal case studies leveraging IRSM, we observed different strategies influenced not only by data characteristics but also by the individual goal of the user. We **hypothesize** at the start of our case studies that *our participants will solely employ pattern searches based on temporal constraints related to order and timing of events in creating search strategies*. We realize the meaning of events will also influence the strategies developed. Since meaning is subjective, we restrict our hypothesis to a more concrete statement. This subjective influence is observed during our case studies. There are other related search mechanisms for multimodal/multimedia data (e.g., [35, 36]) in which images, video, text, and meta data are considered. However, our focus is specifically on temporal events found within multimodal data. A unique aspect of our case studies is they are real researchers motivated to analyze their own data and not participants in a streamlined lab environment.

In Section 2 we review related work. Section 3 describes the details of our case studies. Section 4 introduces the building blocks available to our participants in which the search strategies were created. These building blocks are discussed in Section 5 and 6. After which, Section 7 describes the unique search strategies created. Discussion is provided in Section 8 in which overarching results are discussed and Section 9 concludes the paper.

2. RELATED WORK

For clarification, we view searching as identification of event pattern occurrences within multimodal data. We adopt the definition of a pattern as described in [17]. We view an event pattern (or simply a pattern) as a sequence of events that has some associated meaning to the user performing the

ICMR'14, Apr 01-04 2014, Glasgow, United Kingdom

http://dx.doi.org/10.1145/2578726.2578761 ...$15.00.

Table 1: Case Study 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

search. Such a pattern may have associated timing information (or other descriptive information - discussed later) representing another level of meaning to the user. Specifically, we are interested in pattern identification within multimodal data which are segmented into sequences of events describing the actions captured by the multimodal data, such as the interactions among humans. This segmentation creates a challenging data space to search characterized by non-numerical, temporal, descriptive data, e.g., Person A walks up to Person B at time T. The segmentation and this view of multimodal data is very common [5, 13, 15, 18, 27, 30].

The rest of this section discusses the related work of temporal relations with respect to ordered events. Our research is founded on creating a formalism of a pattern based on structure, timing, and ordered relationships. The ordered relationships that are inherent to temporal event data have been an active area of research. Allen in [2] formulated thirteen relationship principles to express all the possible ordering relationships between two interval events. Much work has been conducted to detect, extract, and represent such temporal information, e.g., [7, 10, 21, 22, 31, 33].

After Allen, Freksa revisited interval relationships at the semi-interval level (an interval's start and end) [8]. Semi-intervals allows a flexible representation where partial or incomplete knowledge can be handled as operations are on parts of an interval and not the whole. All of Allen's relationships can also be represented by Freksa's formalism. Interestingly enough, very little work has focused on using semi-intervals. The most notable was completed by Mörchen and Fradkin in [24] where they explored semi-intervals for use in unsupervised pattern mining. Other work [17, 18, 19] has explored semi-interval relationship processing.

Situated within the data mining domain, the symbolic temporal pattern mining (STPM) approach focuses on discovering "interesting" patterns among *symbolic* time series data [11, 23]. STPM is related to the need to represent time and timing among events. One such approach is T-patterns developed by Magnusson [12] in which a sequence of events will occur within certain time windows of each other, e.g., $A_1 \xrightarrow{T_1} A_2 \xrightarrow{T_2} A_3$ for time intervals T_1 and T_2. T-patterns are used as the basis of pattern representation and identification in Theme [1] where each T_i, for $i \geq 1$, is set through various statistical methods. Several other research endeavors have pursued T-Patterns [3, 34]. Related are frequent episode mining (FEM) algorithms [25, 28] that operate on interval and point data in order to identify event sequences of varying lengths (episodes) given certain timing restrictions between events. FEM algorithms identify these episodes given a threshold (frequency or statistically based).

3. CASE STUDY DETAILS

Three longitudinal case studies were conducted 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. The tool to support their analysis, e.g., support them searching their multimodal data, was Interactive Relevance Search and Modeling (IRSM) [20]. For the purpose of this paper, the details of the tool are not necessary but the reader may reference [20] for more detail. It is suffice for the reader to know that IRSM allowed the researchers to search through their multimodal data with great flexibility. The details of the features that provided this flexibility are found in later sections. In this section, a description of the researchers' demographics and datasets is provided.

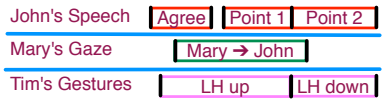

Figure 1: Example of multimodal data (semi-intervals highlighted as vertical bold lines) with speech, gaze and gesture channels.

3.1 Demographics

Three researchers from the Center for Human-Computer Interaction at Virginia Tech were independently recruited. The demographics of the researchers can be seen in Table 1. Each participant had conducted research for at least 1.5 years and familiar with some form of data analysis prior to the case studies. These prior analyses were conducted using standard analysis techniques and software packages.

3.2 Datasets' Descriptions

In this section we describe our participants' datasets and for simplicity we refer to them as P1, P2, and P3.

Data Characteristics: The data of each participant were multi-channel events represented by time points and/or intervals, i.e., annotated multimodal data [18, 29, 32]. This is commonly referred to as multivariate symbolic interval series with a mixture of symbolic time sequence data [23]. The events were annotated from their media sources either automatically and/or manually. We view the events at a symbolic semi-interval level (start or end of a symbolic interval). Point events are viewed as a single symbolic semi-interval. Each channel represents symbolic events of a certain type (i.e., event type). This is illustrated in Figure 1 where speech, gaze, and gesture channels of different individuals are represented. 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 series 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. P1 had previously performed quantitative analyses of feature contributions using statistical packages (see Table 1). A visualization overview of P1's data can be seen in Figure 6, top-left.

Participant 2's data: P2 was studying a new multi-scale interaction technique for large, high-resolution displays. The

Table 2: Datasets' Contents

	Semi-intervals	Unique Semi-intervals	Channel Min	Channel Mean	Channel Max
P1 Original	2218	252(max)	7	14.22	23
P1 Filtered and Clustered	1305	6	3	3	3
P2 Original and Normalized	2784	6	3	3.83	4
P3 Original	8545	25(max)	10	12.43	14
P3 Filtered	1163	25(max)	10	11.57	14

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. 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/`). A visualization overview of P2's data can be seen in Figure 6, top-middle.

Participant 3's data: P3 was studying cooperative use of a large, high-resolution display in performing an intelligence analysis task. There were 7 sessions with 2 people per session sharing the display, each with their own mouse. All sessions were video recorded with manual annotations 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. P3's analysis focus was whether the display employed would be instrumental in facilitating common ground among each pair of participants. 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, manual video coding to identify situations of interest, and viewing periodic screenshots taken during each session to observe the use of display space. A visualization overview of P3's data can be seen in Figure 6, top-right.

3.3 Methodology

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 was expected as each participant was analyzing their own unique dataset. This provided real world scenarios in which to observe the development and application of search strategies. Each participant was provided a computer with IRSM installed and access to their data. Two types of sessions were conducted where screen capture and event logging were performed. After each session, a semi-structured interview was conducted to record the participant's experience.

Training Sessions: Three training sessions were conducted for each participant. The purpose of the training sessions was to familiarize each participant with the features of the analysis environment. During these sessions, a moderator worked with the participants. The participants worked with a sub-set of their own data during these sessions.

Independent Sessions: After training, the participants performed independent sessions where the moderator provided help only when it was deemed absolutely necessary. Four independent sessions were run for each participant. The one exception was P2 who was satisfied with his results by the end of his second independent session. All sessions were conducted over a period of four weeks. Each session ranged from 30 minutes to 1 hour.

Analysis: For the analysis of our case studies, we applied a form of Cognitive Task Analysis (CTA) [6], which has also been applied in intelligence analysis tasks and sensemaking [26]. As outlined in [6], we *first* collected preliminary knowledge through early discussions with our participants about their analysis focus, goals, and data. We worked with our participants in this manner 2-4 months before any sessions were held. *Second*, we discussed knowledge representations with our participants for their analysis. We introduced the idea of temporal patterns of annotated events within multimodal data as a representation and approach for searching their data based on the pattern definition in [17]. *Third*, in order to elicit knowledge about our participants' experience and search strategies, we took careful observations during each session and utilized semi-structured interviews after each session. *Fourth*, in order to analyze and verify the data we acquired, we performed follow-up interviews with each participant to verify our results of how each of them performed. In one case (P2), this resulted in us altering our concluding results as the participant provided valuable feedback to guide our final assessment. *Fifth*, and lastly, with our results, we translated them into three search strategies (Section 7) with related discussion (Section 8).

4. BUILDING BLOCKS

In order for our participants to search their datasets, they were provided with data manipulation techniques and search features. To fully understand the search strategies created, it is necessary to explain these techniques and features before presenting the strategies observed. The next two sections (5 and 6) discuss the techniques and features and how each were used by our participants. The following section (7) then discusses how the techniques and features were applied to create unique and beneficial search strategies.

5. DATA MANIPULATION

Despite our initial hypothesis expecting sole use of temporal constraints, each participant's data exhibited various data characteristics that required data manipulation to facilitate searching. Here we discuss these data characteristics and the formatting techniques developed and applied.

5.1 Data Characteristics

Our participants' datasets had varying data characteristics. Each exhibited either *dense* events, *sparse* events, or event *clusters*. Examples of each, respectively, are illustrated in Figure 2. Dense event data is characterized by many events among one or more channels for a prolonged period of time. Oppositely, sparse event data has very few events among one or more channels. Clustered event data is

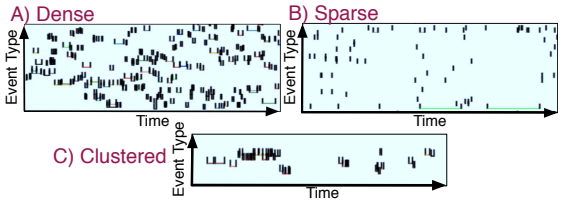

Figure 2: Different temporal characteristics.

Figure 3: An excerpt from P2's data. A) Before normalization, and B) after Normalization. For both, F1, F2, and F3 represents the different finger modes and NT a new target.

characterized by dense groups of events among one or more channels with time gaps between each group. The events of each channel also exhibited *temporal variability*. The time between events (and event length) varied with no discernible consistency between occurrences. The events also consisted of either *point data* (e.g., instant event) or *interval data* (e.g. event has a start and an end). These characteristics were not new, however, the challenge was each researcher had multiple sessions each exhibiting a combination of these characteristics. Their data was also multi-channel where channels could exhibit any combination of the described characteristics in parallel with others. These characteristics provided a challenging data-space to search effectively.

5.2 Formatting Techniques

Formatting data is not new, but, given the varying characteristics of our participants' data, carefully considered formatting was invaluable. This discussion is focused on presenting the different formatting techniques deemed beneficial given the data characteristics discussed previously.

Normalization: For our purposes, *normalization* refers to transforming all event to be the same temporal length, preserving their order, removing any time gaps between events, e.g., remove temporal variability, and produce a simple event sequence that captures the events' order. This is illustrated in Figure 3. Here we see an excerpt from P2's data in which the before and after effects of normalization are illustrated. All events were normalized to 1 time unit intervals, allowing an easy view of the ordering trends in the data. This is useful to ignore any temporal variability between event lengths and time gaps between events and focus just on the order (which may be the relevant focus). This was successfully applied to P2's data as the initial focus of P2's analysis was solely the order of certain events. P2 also used *normalization* to identify areas of his data that were of interest, then viewed those same areas in the unaltered data to gain more insight into the patterns identified.

Filtering: The second formatting technique is *filtering* out details. We realize this is not new either, however, filtering out finer details (e.g., transcription speech content and highly reoccurring events) can allow a better view of

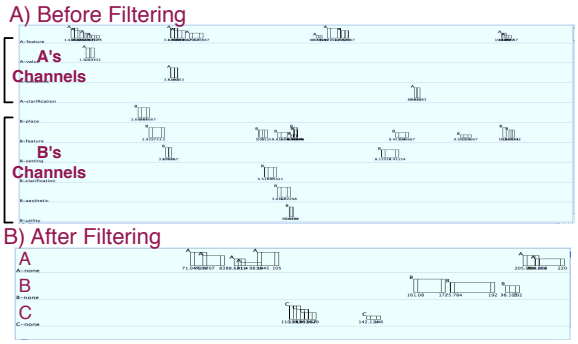

Figure 4: An excerpt from P1's data. A) Before filtering. Note that not all channels are shown as there are too many. B) After filtering where each user in P1's data has only one channel (total of three).

the structure of the data at a high level, after which, the details can be added back in to better inform specific occurrences in the data. This was applied for P1 and P3 as both of their datasets had information that distracted them from viewing the data in a meaningful way. For P1, this was filtering out all details except for when his participants took turns, i.e., turn-taking. This can be seen in Figure 4, where in Figure 4A we see the unfiltered data where each of P1's users have multiple channels. Note that user B has more channels than is depicted and user C is not even show since there is not enough space. In Figure 4B, after filtering is applied, each user has one channel that represents when they spoke. A cluster of speech events then represented a turn for each participant. For P3, this was filtering out events from the interaction mouse logs, which had a high occurrence. The mouse log data was distracting at the beginning of the analysis and P3 chose to filter out these events.

Clustering: The third formatting technique is *clustering*. This was an offline feature where events could be clustered within channels. This was useful when a channel had a sequence of events that a participant wanted to be coalesced into one event. This was solely used by P1. In his data, each turn of his participants had several events as can be seen in Figure 4B. However, P1 wanted each turn to be seen as one event. Therefore, he applied clustering for each of his participants to achieve this. Some cases were not captured since the cluster threshold (chosen to be 3 seconds) was too small, but P1 was afraid of making the clustering threshold too loose resulting in skewing the turn-taking structure. The reader should be careful to not confuse *clustering* as a formatting technique and clustered as a data description.

Time Scaling: The last formatting technique is *time scaling*. This feature was originally developed to aid in viewing datasets with long timeframes (e.g,. years). It allows one to define a base time unit (i.e., the time unit the data is stored in, such as seconds or minutes) and then define a viewing time unit. For example, if the events are stored at the minutes level, one can view the data at an hour time scale, in essence zooming out. P3 was the only participant who used this feature. Her data was stored in seconds, but she found it useful to view the data at the minute timescale. This was especially helpful since each of her datasets represented two hour-long, video captured sessions. Hence, viewing in minutes gave her an easy way to index what part of a session she was viewing, e.g., either at the beginning, near the middle or at the end.

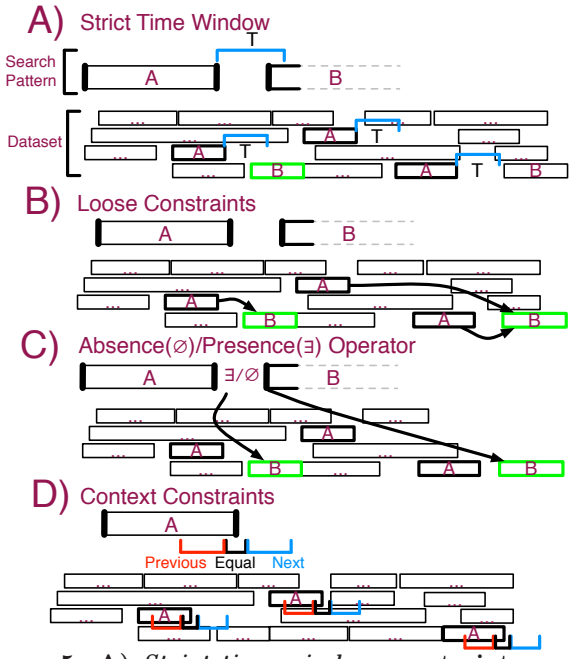

Figure 5: A) *Strict time window* constraints example. The bold, green B event is the match. B) *Loose constraint* example. C) *Absence* and *presence* operator example. D) *Context constraints* example.

6. TEMPORAL CONSTRAINTS

As seen in Figure 6, bottom row, even after applying formatting techniques, each participant's dataset still posed a challenging search space (notably P1 and P3's). The application of different temporal constraints was necessary. A temporal constraint is a constraint applied to govern the temporal relationships between events. Temporal constraints can be used in a search to specify relevant results. For example, if searching for event A followed by event B, one may only be interested in matches where B occurs within 5 seconds of A. Such a temporal constraint defines the relationship desired between A and B. This is an example of a temporal constraint class seen in [12, 25, 28].

6.1 Pattern Constraints

Pattern Constraints are constraints that are applied to matching a pattern in the data, i.e., identifying *pattern* occurrences. These aided our participants in expressing their search criteria and identify results that were relevant.

Strict Time Windows: This is the most well-known temporal constraint where the relationship between events is defined through a strict time window. An example can be seen in Figure 5A with $T = 5$ where one is looking for B starting within 5 seconds of A. This category has been successfully applied in symbolic temporal mining [25, 28] and pattern discovery in behavioral data [12]. In our case studies, there were two classes of strict time windows. One in the classic sense of sequence order, e.g., B follows A within 5 seconds, and the other in terms of equality. The *first*, a *next constraint*, allows one to define a temporal window between events. Only events falling within this temporal window are considered. This is illustreated in Figure 5A.

The *second* class, an *equal constraint*, allows one to define what it means for two events to be equal. Depending on the context and meaning of the events, the idea of two events being equal can vary, especially if there is any uncertainty in the timing of events caused by the originating sensors. Due to the nature of either the participants' data, transformations applied to the data, or both, none of our participants required adjusting the *equal constraint* from its default value of 1 video frame, i.e., 33 milliseconds.

Loose Constraints: If events have temporal variability, then relying on a strict time window may not be realistic. To address this, apply *loose constraints* in which order between events is preserved but no temporal constraints are applied. This is illustrated in Figure 5B. However, when this is done many search results may be meaningless as they contain overlap. In Figure 5B, there are three matches but two of them overlap with each other. What is sought is the match with A temporally closest to B. To address this issue, we apply the concept of non-overlapping support for the search results. An idea from data mining [25], non-overlapping patterns are a set of patterns that do not have any temporal overlap. We adopt this idea and return results that are the temporally most tight. This allowed our participants to apply loose constraints and not be presented with irrelevant, overwhelming results. P1 found this very useful as his data varied from dense, to sparse, to clustered.

Absence/Presence Operator: P1 was very interested in patterns where events were not interrupted by any other event. This resulted in an absence, $\emptyset$, and presence, $\exists$, operator, illustrated in Figure 5C. This allows one to specify whether or not other events were present (or not) between specified events of their search pattern. This was heavily used by P1 who desired to identify matches where specified events in his pattern had no other events between them.

6.2 Context Constraints

Temporal constraints were also needed in governing the context returned with identified matches. Our participants were not only interested in identifying matches to their searches but also the context in which those matches occurred. They were interested in a situated view of their matches, i.e., viewing the context of each of their matches. Hence, the development of *context constraints* which govern how much context is desired for each match. These constraints are based on the suggestion categories of [19] in which events within certain time windows *previous*, *concurrent to*, and *after* each match are also provided. This is illustrated in Figure 5D with the different time window categories. Each match of the pattern has a certain amount of its situated context returned based on how the time windows are defined. *Context constraints* were applied to each semi-interval of the participants' patterns, as seen in the figure.

7. SEARCH STRATEGIES

Given the building blocks' descriptions, we will now present how our participants used those blocks in different configurations and created search strategies. Each participant created a unique strategy given their respective data characteristics and analysis goals. For reference, an overview visualization of the participants' data is show in Figure 6. Here the original of each participant, P1, P2, and P3, are along the top row, respectively. The result of the applied formatting techniques are along the second row, respectively. The x-axis is time and the y-axis is the event types grouped by the sessions within the respective participant's datasets. Also, an overview of the data characteristics of our participants' data

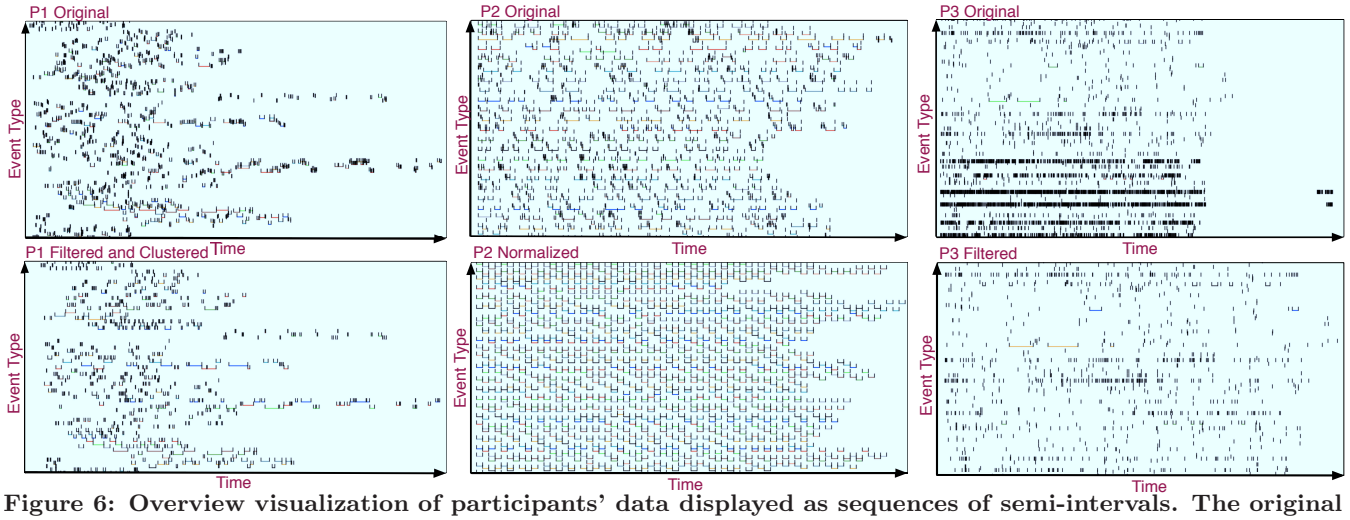

Figure 6: Overview visualization of participants' data displayed as sequences of semi-intervals. The original of each participant, P1, P2, and P3, are along the top row, respectively. The result of the applied formatting techniques are along the second row, respectively. The x-axis is time and the y-axis is the event types. Visualization created using TDMiner, `http://people.cs.vt.edu/patnaik/software`.

and the search and formatting techniques can be seen in Figure 7. The provided key succinctly explains the color coding. Note that P1 and P3 are placed side-by-side to emphasize the mutually exclusive aspect of the strategies applied. The organization of the data characteristics, the constraints, and formatting techniques is to allowing easier visual comparison between the three case studies.

The analysis performed to identify the search strategies was qualitative in nature and based on observation. Such strategies were clearly identifiable over the duration of the longitudinal case studies as each participant had a unique focus, goal, and approach to their analysis task that was clearly observable. Their unique focus, goal, and approach was recorded through the semi-structure interviews conducted after each session and observations recorded during each session. The presented search strategies and patterns employed by each participant are a representation of each participant's unique focus, goal, and approach. We will now discus these observed strategies.

Flexible Matching: P1 created the strategy of *flexible matching* in which one uses the temporal constraints that allow the most flexibility in matching occurrences in the data. As can be seen in Figure 7A, P1 had a dataset with a range of characteristics. Starting his analysis, P1 was interested in turn-taking. Hence, he applied *filtering* to filter out unnecessary elements of his data so he could view the turn-taking structure. He then applied *clustering* so each turn taken by the users in his data was seen as a single event interval. Such formatting techniques allowed P1 to have a relevant view of his data. The contents of P1's datasets after applying the formatting techniques can be seen in Table 2 and a visual overview is illustrated in Figure 6, bottom-left.

P1 originally tried *strict time windows* as a *pattern constraint*. An example pattern of this can be seen in Figure 8A (i). Here P1 has defined three semi-intervals, one after the other within 3 seconds (outside of the 33ms *equal constraint*). He also defined variables (X and Y) for matching the participants in his data (predicate logic). The 's' and 'e' represent a start or end semi-interval, respectively. Due to the temporal variability of his data, this was not very successful. Hence, he tried *loose constraints*, which provided better results. However, he was still not satisfied. He then turned to applying the *absence operator* in tandem with *loose constraints*, each carefully placed in his pattern. This resulted in the final pattern (Figure 8A (ii)) that P1 used successfully for a majority of his analysis sessions.

Transition-Point Matching: P2 created the strategy of *transition-point matching* in which one uses constraints and formatting techniques to facilitate identifying and viewing transition-points. What constitutes a transition-point is dependent on the user conducting the search. For P2, this was the transition between specific finger modes and either reaching a target or starting a new target. P2's data consisted of temporally variable interval events (Figure 7A).

P2's interest was looking at specific orderings between finger mode usage. Hence, he applied *normalization* to streamline his data so the ordering structure would be more easily apparent. Doing so allowed him to search based on boundaries (*equal constraint*), i.e., transition points between finger mode events. Other information was lost due to *normalization*, such as the timing between events, however, this lost information was not relevant to P2. Example patterns can be seen in Figure 8B. Here P2 has defined two sets of two semi-intervals designed to identify when one of his participants reached a target using either finger mode 2 or 3, respectively. Since *normalization* removed time gaps between finger events, this allowed high accuracy for identification of the boundaries of interest. The high accuracy of an *equal constraint* was noted in [19]. The contents of P2's datasets after applying *normalization* can be seen in Table 2 and a visual overview is illustrated in Figure 6, bottom-middle.

Strict Matching: P3 created the strategy of *strict matching* in which one focuses on more strict constraints to identify matches. As seen in Figure 7A, P3 had a dataset with a range of characteristics, similar to P1's. P3 was concerned about her mouse log data being distracting as it was very dense, hence, she first *filtered* out this data. She then began exploring her data with simple pattern searches to better understand what existed in her data. She experimented with using *next*, *loose*, and *context constraints*. The *loose constraints* did not operate well with her data as most of the events in the matches were temporally too far apart to be meaningful. The reason for this was later realized to be the matches P3 sought actually did not occur very of-

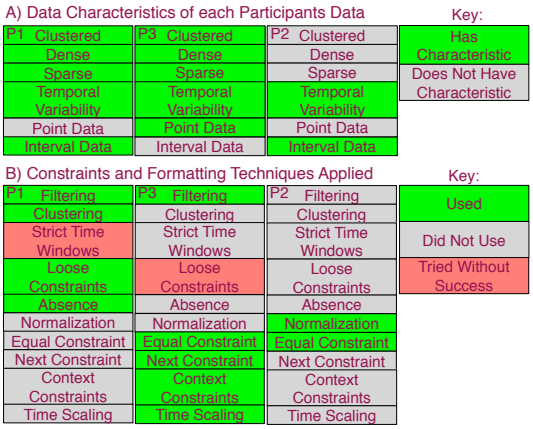

Figure 7: A) Data characteristics of each participants data. Color key on the right. B) The possible constraints and formatting techniques the participants could use. Color key on the right.

ten. She successfully adjusted the *next constraints* according to a timeframe that was meaningful to her. After which, she successfully adjusted the *context constraints* to report a meaningful context window. Through her exploration and analysis, she discovered her data was characterized further as having little tight pockets of activity, explaining why applying *next* and *context constraints* was so successful.

Examples of two of P3's patterns can be seen in Figure 8C. The first one is an example of a single semi-interval P3 used to "sniff" through the data by looking around occurrences matches of this pattern using *context constraints*. In this case, it was one of her users (hence the use of '*' as in a regular expression) discussing the location of something on the display. The second pattern is an example of a pattern used later in her analysis to identify when her participants had a significant speech event within 3 seconds of the other participant discussing the location of something on the display. The contents of P3's datasets after applying the formatting techniques discussed can be seen in Table 2 and a visual overview is illustrated in Figure 6, bottom-right.

8. DISCUSSION

We observed how each participant, each faced with different challenging data characteristics, were able to successfully search their data and identify relevant instances to their respective analysis goals. Each participants' analysis focus dictated the patterns used which then dictated the strategy created or applied. There were three overarching results.

First, the data characteristics drove how each participant approached their respective analysis goals. This is viewed as a *characteristic barrier* that each participant had to discover and address. All participants used formatting techniques as a tool to better view their data in a meaningful way. This was necessary due to the varying characteristics of each dataset. Once they were able to identify areas of interest, they viewed the unaltered data for full details. P1 and P3 used *filtering* in this manner and P2 used *normalization*.

Second, the meaning of the data and each participant's analysis goals drove how they approached searching their data. This is viewed as a *conceptual barrier* each participant had to address. This is exemplified when comparing P1 and P3. Each had very similar data in terms of data characteristics, however, the constraints and formatting techniques applied were almost mutually exclusive as can be visually

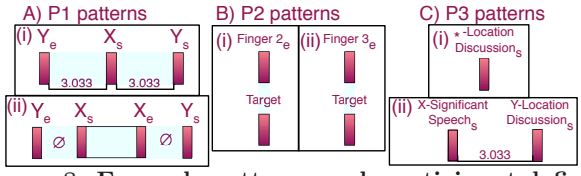

Figure 8: Example patterns each participant defined as a search query.

seen in Figure 7B. The reason behind such a difference is their individual analysis goals and what their data represented, especially to each individually. How one approaches data is dependent on the characteristics of the data, what one is looking for, and the meaning of the data content. It is also interesting the different outcomes when trying *loose constraints*. This was likely caused by how the data characteristics differed in the places where relevant matches were found in their respective datasets.

Third, and last, were features that were important to researchers searching multimodal data. Our participants found it important to have a graphical visualization that enabled them to view and understand the structure and relationships of events within their datasets. Most multimodal analysis tools provide some form of support of this kind of visualization. Specifically, IRSM was designed to support the researcher this way. The strength behind how IRSM approached this was found in interactively defining and adjusting temporal patterns that represented a relevant focus for each participant and connecting this visualization and search results to the overview visualization of the data. More details of how this was done can be found in [16]. Our participants also found it important and beneficial to be able to define their searches based on temporal relationships of events that were relevant to them. This allowed them to apply their knowledge and expertise in guiding the search.

As seen from these results, our original hypothesis was partially correct. Our participants needed formatting strategies along with temporal constraints to support them searching their respective multimodal datasets. Then given this search support, each participant adjusted their search strategy according to the characteristics of their data and their respective analysis goals to create unique search strategies.

Contributions: We presented real-world situations of multimodal search and have shown how researchers actually search through multimodal data. Specifically the data they searched had multiple characteristics in varying combinations throughout each dataset, a very challenging scenario. This has revealed insights on how to further support multimodal analysis and especially what is important to researchers. Our participants were most interested in viewing their data so as to understand its structure better and the ability to define meaningful searches based on how events interact, e.g., events' temporal structure and relationships.

9. CONCLUSION AND FUTURE WORK

We were able to present the search strategies employed, discovered, and evolved during three longitudinal case studies for participants analyzing and searching multimodal data. Formatting techniques *along* with temporal constraints were required to aid in searching the multimodal datasets. Through these case studies, we observed three unique search strategies that were created for different situations.

The results of our case studies can inform future automatic content retrieval systems as researchers must be aware

of the *characteristic* and *conceptual barriers* when designing such systems. Inclusion of aspects that are important to researchers searching multimodal data is necessary for achieving more accurate and relevant results. Careful attention to detail must be made about various constraints and formatting techniques and how they can affect search. The concepts explored and discovered in these case studies can be viewed as building blocks for future systems that wish to operate on such data.

10. 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.

11. REFERENCES

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

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

[3] O. Brdiczka, N. M. Su, and J. B. Begole. Temporal task footprinting: identifying routine tasks by their temporal patterns. In *IUI '10*, pages 281–284. ACM, 2010.

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

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

[6] R. E. Clark et al. Cognitive task analysis. In *Handbook of research on educational communications and technology*. Lawrence Erlbaum Associates, 2008.

[7] P. Cohen. Fluent learning: Elucidating the structure of episodes. *AIDA*, pages 268–277, 2001.

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

[9] 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.

[10] P. Kam and A. Fu. Discovering temporal patterns for interval-based events. *Data Warehousing and Knowledge Discovery*, pages 317–326, 2000.

[11] S. Laxman and P. Sastry. A survey of temporal data mining. *Sadhana*, 31(2):173–198, 04 2006.

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

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

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

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

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

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

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

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

[20] C. Miller, F. Quek, and L.-P. Morency. Interactive relevance search and modeling: Support for expert-driven analysis of multimodal data. In *ICMI '13*. ACM, 2013.

[21] C. Mooney and J. Roddick. Mining relationships between interacting episodes. In *SDM'04*, SIAM, 2004.

[22] F. Mörchen. Algorithms for time series knowledge mining. In *KDD '06*, pages 668–673.

[23] F. Mörchen. Unsupervised pattern mining from symbolic temporal data. *SIGKDD Explor. Newsl.*, 9(1):41–55, 2007.

[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] 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.

[26] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis.

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

[28] P. S. Sastry and K. P. Unnikrishnan. Conditional probability based significance tests for sequential patterns in multi-neuronal spike trains. 2008.

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

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

[31] E. Schwalb and L. Vila. Temporal constraints: A survey. *Constraints*, 3(2/3):129–149, 1998.

[32] 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.

[33] A. Ultsch. Unification-based temporal grammar. Technical Report 37, Philips-University Marburg, Germany, 2004.

[34] V. V. Vishnevskiy and D. P. Vetrov. The algorithm for detection of fuzzy behavioral patterns. In *Measuring behavior*, pages 166–170, 2010.

[35] E. Younessian, T. Mitamura, and A. Hauptmann. Multimodal knowledge-based analysis in multimedia event detection. In *ICMR '12*, pages 51:1–51:8. ACM, 2012.

[36] D. Zellhöfer et al. Pythiasearch: A multiple search strategy-supportive multimedia retrieval system. In *ICMR '12*, pages 59:1–59:2. ACM, 2012.