To appear in *Intelligent Multimedia Information Retrieval* (Mark Maybury, Ed.)

Interactive Visualizations for Temporal Analysis: Application to CSCW Multimedia Data

Stacie Hibino and Elke A. Rundensteiner

Electrical Engineering & Computer Science Dept.
Software Systems Research Laboratory, The University of Michigan
1301 Beal Avenue, Ann Arbor, MI 48109-2122 USA
hibino@eecs.umich.edu, rundenst@eecs.umich.edu

Abstract

Although multimedia data is commonly collected by various researchers for a variety of purposes, previous support for specifying temporal queries and analyzing such data for temporal trends has been limited. In this chapter, we present a new paradigm for temporal analysis—one where users *browse* the data in search of temporal trends and relationships by using simple mouse manipulations to incrementally pose temporal queries within an integrated MultiMedia Visual Information Seeking (MMVIS) environment. In MMVIS, our specialized temporal visual query language (TVQL) not only allows users to pose specific temporal queries based on strict relationships (e.g., when A events start at the same time as B events), but also to relax the constraints to explore *similar* relationships (e.g., when A's start within a few seconds of B's). The temporal visualization (TViz) of results highlights the strengths of temporal relationships by *clustering* them together rather than distributing the display of them over time in a timeline format. TViz is dynamically updated *as* users manipulate the TVQL query filters. In this chapter, we present a case study using our approach to temporally analyze video data collected as part of a CSCW study. This case study illustrates how our approach simplifies the process of examining temporal trends and complements traditional timelines and statistical analyses. It also indicates how we could enhance intelligent multimedia information retrieval through supporting the analysis of temporal structure of multimedia documents.

1. Introduction

The Need for Temporal Analysis. Temporal data is commonly collected by various researchers for different purposes. User interface evaluators collect video data and logfiles, doctors review cardiology diagrams, sports analysts collect team and individual statistics, etc. Previous support for analyzing such temporal data has primarily focused on the use of variations of timelines (e.g., Harrison et al. (1994)), the use of statistical methods (e.g., Markov analysis), and/or the analysis of temporal sequences (e.g., Sanderson et al. (1994)) rather than analysis of any type of temporal relationship—sequential, parallel, or overlapping. New paradigms are needed to analyze and explore *temporal relationships* between various events in these multimedia documents. In our work, we provide a solution to this temporal analysis problem, with special focus on analyzing video.

Moving Towards an *Object-Level* of Video Analysis. Current research in *bit-level* video analysis is making significant advances towards efficiently automating the identification and indexing of objects and events occurring within a video (e.g., Hauptman and Witbrock (this volume), Mani et al. (this volume), Zhang et al. (1995)). Even when events need to be subjectively coded according to interpretation, tools are available for doing so at a speed proportional to the time required for real-time playback (Weber and Poon (1994)). Now that we can efficiently abstract atomic objects and events, we can thus move on to a more complex, *object-level* of video analysis—one where we can analyze *relationships* between objects and events. In our research, we are examining this object-level of video analysis. In particular, we are exploring issues related to supporting:

- *direct queries* over a video (collection) based on specific temporal relationships (e.g., when do events of type A temporally occur during events of type B?),
- *analysis* of these relationships between such events (e.g., *how often* do A events occur during B events?), and
- *browsing* and exploring variations on these relationships (e.g., *comparing* results of different analyses, such as how often do A and B events start at the same time? versus how often does A occur any time during B?).

Chapter Overview. This chapter is divided into five additional sections. In Section 2, we describe the details of our new exploratory paradigm for temporal analysis. In Section 3, we present the case study of applying this approach to real video data. This is followed by an evaluation in Section 4. In Section 5, we discuss related work, and in Section 6, we present our conclusions.

2 Temporal Explorations with Interactive Visualizations

2.1 A New Paradigm for Temporal Analysis

In our new paradigm for temporal analysis, users can temporally *explore* data in search of *temporal relationships* and trends. While this approach builds on existing work by Ahlberg and Shneiderman (1994) in Visual Information Seeking (VIS), we go beyond the scope of the original VIS and focus on exploiting a particular dimension, namely the temporal one, for the purpose of video analysis.

Similar to VIS, users in our MultiMedia VIS (MMVIS) can browse a database of information through direct manipulation of buttons and sliders. This use of dynamic query (DQ) filters provides us with an easy-to-use *visual* paradigm for developing and posing questions. A visualization of the results is dynamically updated as users adjust a query filter. Users thus incrementally specify and refine queries and can see the direct correlation between adjusting parameter values and corresponding changes to the display of results.

More specifically, MMVIS provides an *exploratory approach to temporal analysis*, consisting of the following user process:

1. Select subsets of the data via subset query palettes.
2. Query for temporal relationships between subsets via specialized temporal query filters (TVQL).
3. Review visualization (TViz) of results for temporal trends.
4. Customize visualization for further clarification, if desired, and go to 3.
5. Go to 2 to incrementally adjust temporal query or to 1 to select new subsets.

In MMVIS, each *subset query palette* includes a multi-select listbox for each type of annotation characteristic (i.e., name, action, receiver, and category). Sample subset selection palettes are included as part of Figures 4 and 6. TVQL and TViz form the primary core of MMVIS and are summarized below. A more detailed description can be found elsewhere (Hibino and Rundensteiner (1996a, 1996b)).

Note that the underlying database of our MMVIS system stores a collection of video annotations that abstract *atomic* objects and events in the video data[1]. While we provide primitive support for creating annotations manually within MMVIS as part of our tool suite, users can import events that have been automatically indexed by bit-level video analysis systems.

2.2 Temporal Visual Query Language (TVQL)

Specifying Primitive and Neighborhood Temporal Queries. Given two events A1 ( ) and B1 ( ) with nonzero duration, Allen (1983) has defined thirteen possible primitive temporal relationships between them (see Figure 1). Note that one endpoint relationship between two events may pose a

[1] Our video annotations are formally described elsewhere along with the justification for processing video annotations rather than the raw video frames (see Hibino and Rundensteiner (1996b)).

constraint on one or more of the others. Besides analyzing these primitive relationships, it is also desirable to specify combinations of the primitives (e.g., to look at situations where events start at the same time but may end at different times, corresponding to combining the *starts*, *started by*, and *equals* primitives). Rather than forcing users to explicitly specify a number of complex disjunctions and conjunctions for combining the temporal primitives, we propose an alternative approach based on the principle of *temporal neighborhoods* (Freksa (1992)).

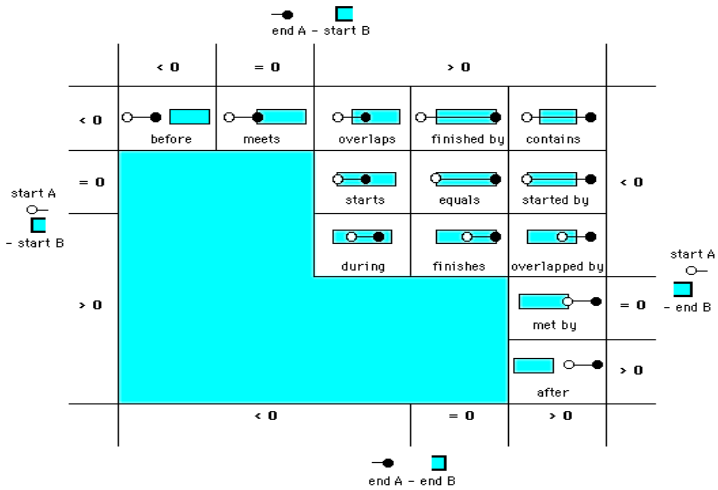

Figure 1. Relationships between temporal primitives and the four defining endpoint difference relations.

Two primitive temporal relationships between two events are defined to be (*conceptual*) *neighbors* if a continuous change (e.g., shortening, lengthening, or moving of the duration of the events) to the events can be used to transform either relation to the other [without passing through an additional primitive temporal relationship]. Thus, the "before" ( ) and "meets" ( ) relations *are* neighbors, because we can move the ending point of A from before the start of B to its start without specifying any additional primitive relationship. In contrast to specifying arbitrary combinations of the primitives, this notion of temporal neighborhoods supports users in selecting *similar* primitives (i.e., equivalent to selecting a series of adjacent cells such as a row, column, or grid from Figure 1).

TVQL Description. While a complete specification of our temporal visual query language (TVQL) can be found elsewhere (see Hibino and Rundensteiner (1996b)), we review its basic principles here as needed for the remainder of this chapter. Using a temporal query filter for each of the defining endpoint difference relationships described above (Figure 1), we can define a temporal query interface capable of specifying not only all individual temporal *primitives*, but also *temporal neighborhoods*. This allows users not only to browse for temporal relationships between two subsets, but to do so in a *temporally continuous* manner.

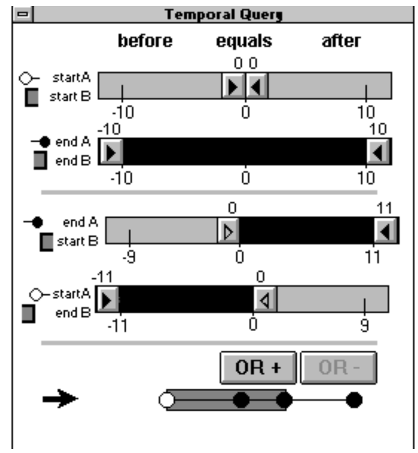

Figure 2. TVQL palette.

The TVQL palette (Figure 2) has three primary components: the temporal query filters (i.e., sliders for specifying temporal parameters), disjunctive OR+ and OR- buttons for combining discontinuous temporal primitives or neighbors (e.g., to specify a query such as (A *meets* B) OR (A is *met by* B)), and a dynamic temporal diagram that visually displays the qualitative semantics of the specified query. The temporal DQ filters are used to examine *quantitative* ranges for the endpoint relationships, such as *startA-startB* set to 0 in the top DQ filter. They are also intelligently bound to one another to prevent the specification of invalid queries. As users adjust one query filter, the other filters are automatically updated accordingly. In Figure 2, the user only has to set the filter thumbs of the top *startA-startB* query filter to 0. The second filter is unaffected, but the bottom two filters are automatically constrained. The underlying framework of the temporal endpoint relationships and their interactions are used to derive these automated constraints (Hibino and Rundensteiner (1995, 1996b)). Note that a filled or open arrow thumb of a DQ filter indicates when the endpoint of a range is included or excluded respectively.

To enhance the TVQL user interface, we have incorporated qualitative descriptive labels along the top and side and our dynamic temporal diagrams along the bottom of the palette. The labels allow users to "read" the relationship specified and the diagrams provide visual confirmation of the temporal primitive(s) specified (though not quantitative values as given by the filters). If subset A specified person P1 and subset B specified all Plan design rationales, then Figure 2 illustrates how users could ask the query "show me how often person P1 starts at the same time as a Plan starts." The descriptive labels can be used to "read" the top query filter as "start A equals start B." The relationship between the temporal ending points is unconstrained as indicated by the selection of all values in the second (i.e., *endA-endB*) query filter. This is also reflected in the temporal diagram, which indicates that the end of A (represented by a filled circle) is before, equal to, or after the end of B.

The power of TVQL is that it allows users to incrementally specify queries by sliding both within and between the specification of temporal primitives and within and between temporal neighborhoods. For example, suppose we want to see when person P1 finishes speaking up to five seconds before a Plan starts (i.e., when A events occur *before* B events by up to five seconds). In order to specify the *before* ( ) temporal primitive, we want endA to be before startB. Thus, we would adjust the third (endA-startB) DQ filter of TVQL to select a range of values less than zero and greater than or equal to -5. Now, by simply sliding the left thumb of the endA-startB filter, we could easily compare when P1 finishes speaking up to five, four, three, etc. seconds before a Plan. By setting the right thumb of the endA-startB filter to include zero, we can add the *meets* () relationships to the temporal query. By sliding the right thumb past zero, we can add the *overlaps* () temporal relationship (i.e., indicating when P1 starts a Plan after finishing another design rationale). These simple mouse manipulations thus illustrate how we can slide from specifying:

- only the *before* temporal primitive,
- to the temporal neighborhood of *before* or *meets*,
- to the temporal neighborhood of *before*, *meets* or *overlaps*.

When coupled with a visualization of results, TVQL can thus be used to temporally *browse* the data without any particular temporal query in mind. The TVQL temporal diagram is dynamically updated as users manipulate the sliders, thereby enabling them to simply slide the DQ filter thumbs back and forth until an interesting result appears, and then use the temporal diagram to identify the type of temporal query specified. Thus, the power of TVQL is that it can be used to 1) specify particular temporal queries (primitives or neighborhoods), 2) browse (i.e., slide within and between) temporal relationships of different types of events, and 3) move seamlessly between querying and browsing.

2.3 Temporal Visualization of Results (TViz)

TViz, presented in the main MMVIS window, is used to abstract, highlight, and compare the relative frequency of the temporal relationships (specified by TVQL) between selected subsets. That is, as a TVQL query is incrementally refined, MMVIS processes changes in the query and passes newly retrieved information to TViz, where the visualization of results is dynamically updated. Thus, in contrast to a text-based tabular display of results, TViz provides a visual abstraction of the answer set and how it changes as queries are refined. TViz is a variation of the visualizations used by Olson et al. (1995) to describe temporal sequences.

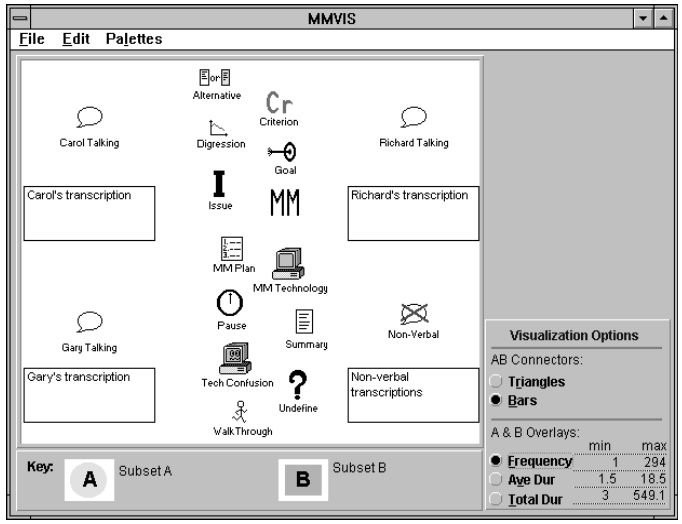

Figure 3. The main MMVIS window.

The Main MMVIS Window. The main MMVIS window (Figure 3) is divided into three areas: the primary *visualization area* containing icons representing the various types of annotations in the database, the *key* below the visualization area to indicate the color-coded data subsets, and the *visualization options* in the lower right of the window to allow users to customize their view.

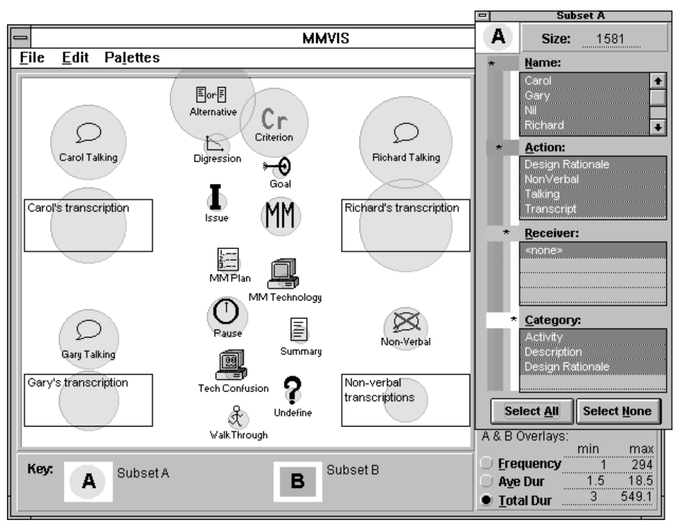

Figure 4. Visualization of selected subsets.

Visualization of the Selected Event Subsets. In our temporal visualization (TViz), selected subsets are visually highlighted with transparent overlays in the main MMVIS window—subset A is indicated with yellow circles and subset B with blue squares. Figure 4 presents an example where the user has set Subset A to all types of annotations and Subset B to none. By doing so, the user can use the visualization options to gain an overall big picture comparing the relative *frequency*, *average duration*, or *total duration* of all types of events (visually indicated by the relative sizes of the transparent circular overlays). In Figure 4, for example, we can see, that the *total duration* that Richard speaks is longer than that of Carol or Gary.

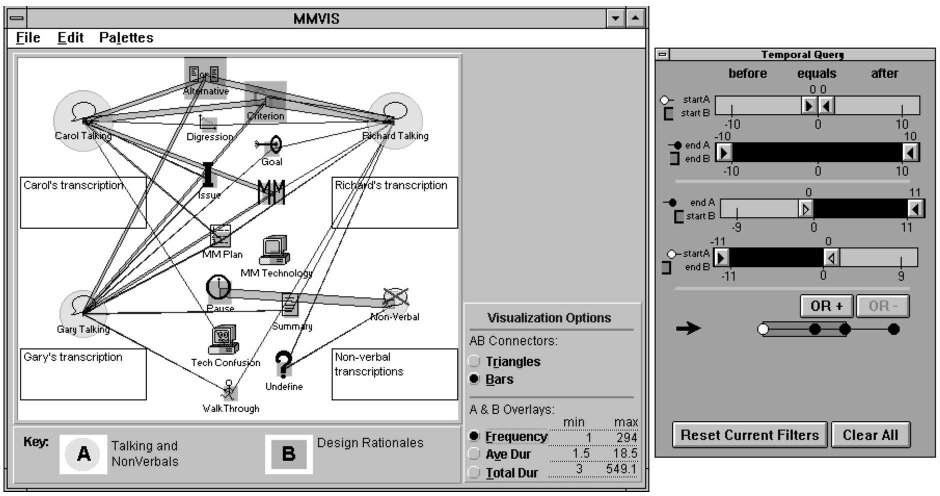

Figure 5. Sample TViz (left) resulting from temporal query (right palette).

Visualization of the Temporal Relationships. Once users have selected A and B subsets, they can then use TVQL to specify a temporal query. As they manipulate the temporal filters, they see connectors between the centers of A and B events appear and disappear, grow and shrink, thereby indicating the existence and strength of the temporal relationship currently specified. The base width of the connector denotes the relative frequency that the temporal relationship occurs. Figure 5 shows the TViz for the *all starts* temporal query, where events start at the same time but may end at the same or different times. The thick bar between NonVerbal and Pause indicates that these types of events frequently start at the same time. The temporal relationship connectors are also user-customizable, being viewable as triangles (to reinforce the direction from A to B) or bars. Figure 5 shows the view by *bars*, where the bar widths indicate the relative frequency.

3. CSCW Case Study: Temporally Exploring Real Data

3.1 Description of the Data Set

The sample data set is based on video collected during a computer-supported cooperative work (CSCW) study (Olson et al. (1995)). During the session, three subjects worked together in a simulated design meeting to draft the initial requirements for an automatic post office. Subjects worked from remote locations using a shared editor, and had two-way video links to each of the other subjects. This video setup provided a virtual conference where the subjects could both see and hear one another. The final video data is a composite of the three subjects. In this case, subject 1 ("Carol") was recorded in the upper left quadrant, subject 2 ("Richard") in the upper right, and subject 3 ("Gary") in the lower left.

In terms of temporal analysis, the original researchers worked on identifying potential trends in temporal sequences of different types of events (e.g., to see "if subjects are currently discussing an Alternative for design, are they likely to immediately follow that with Criteria for that alternative?"). Each event in the database was coded to indicate who was speaking or what non-verbal action was taking place (e.g., a pause or laughter), what was said or a description of the nonverbal, and the starting and ending times of the event. Events were also coded as one of the following thirteen design rationales (DRs): Issue, Alternative, Criterion, Meeting Management (MM), MM Plan, Summary, Digression, Goal, Walkthrough, Pause, MM Technology, Technology Confusion, and Undefined (see Olson et al. (1996)). In our case study, we used the CSCW video data to examine temporal relationships between people speaking and the DRs taking place (e.g., to see whether a Digression is always initiated by one person) and we re-coded the data to separate this information out. While the original data was purely sequential, the re-coded data introduced temporal overlaps.

3.2 Using Subset Event Visualizations for an Overview of the Data

In order to examine the interactions between people speaking and the design rationales taking place, we can set subset A to NonVerbal and Talking annotations and subset B to Design Rationales. The corresponding annotations are highlighted in the main visualization area. Using the visualization display options for "A&B Overlays," we can compare relative frequency, average duration, and total duration of different events.

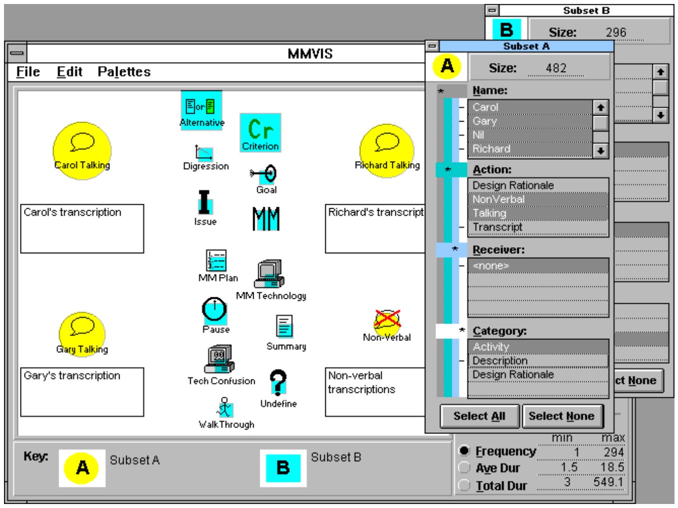

Figure 6(a) View of subsets by relative frequency.

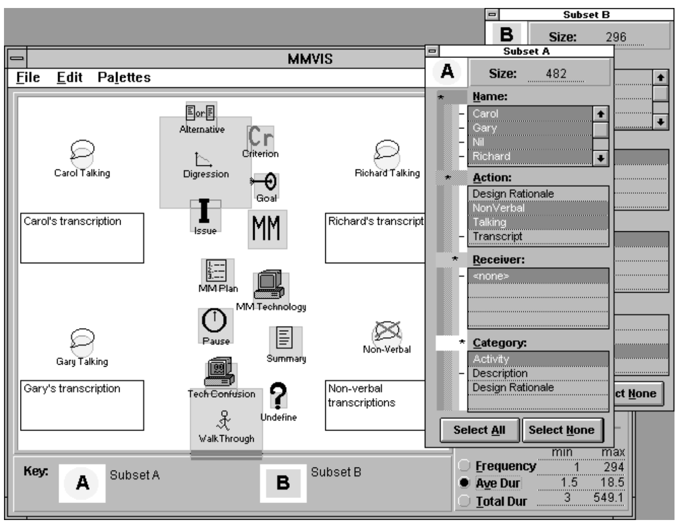

Figure 6(b) View of subsets by average duration.

Figure 6. Selecting multiple subsets. In this example, subset A (indicated by circular overlays) is used to select Talking and NonVerbal events while Subset B (indicated by square overlays) highlights Design Rationales. The visualization options allow users to customize the overview visualization (e.g., to contrast (a) relative frequency with (b) average duration).

In Figure 6a, we can compare the relative frequency of events while in Figure 6b, we can view differences in average duration. (In Section 2.3, Figure 4 showed relative total duration.) The contrasting sizes of the individual overlays within the same, and between different, visualization views provide temporal information that can be used as part of the temporal analysis (e.g., while Figure 6a shows that Alternatives and Criteria occur more frequently than Digressions, Figure 6b indicates that Digressions have a longer average duration).

3.3 Using Relationship Visualizations for Temporal Analysis

Once subsets have been formed, we can specify temporal queries and review the temporal visualizations for temporal analysis. In this section, we present a sample scenario to illustrate how a researcher might use our MMVIS environment to temporally explore the case study data.

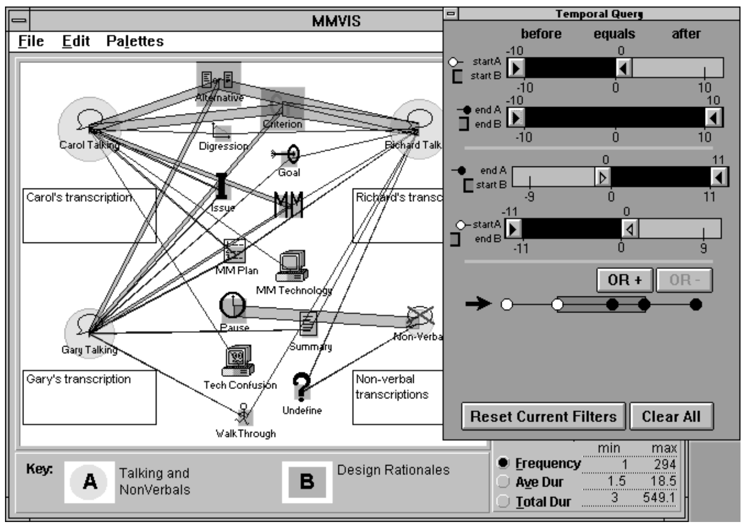

Figure 7. Examining temporal relationships. In this example, the user is querying for Talking and NonVerbal events which initiate *DRs.*

As we begin analyzing the interaction between people speaking and the DRs, we might first want to see who or what tends to start a DR. We could begin by asking a query such as when do activities (i.e., Talking and NonVerbal events) *start* at the same time as DRs (see Figure 5). However, even more interesting is to examine who or what *initiates* (i.e., starts at the same time or before) a DR. We can easily modify a *starts* TVQL query to an *initiates* one by simply dragging the left thumb of the top (startA-startB) query filter from its middle (i.e., "equals") position to the left, thereby setting "startA before or equal to startB" (Figure 7).

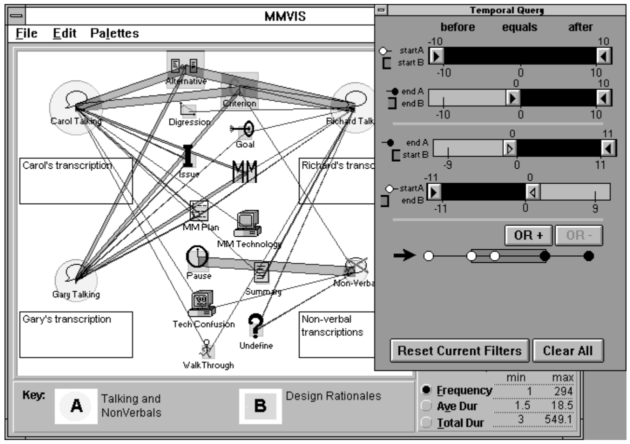

Figure 8. Visualization of the frequency in which a person Talking or a NonVerbal temporally ends *a DR.*

Now that we have seen an overview of when A events *initiate* B events, we might be interested in seeing when A events *end* B events (see Figure 8). For example, we may want to see who ends a Digression. We note, however, that when someone stops talking at the same time as a DR, they don't necessarily "cause" the DR to end (e.g., Richard may be the last one to participate in a Digression rather than the one who initiates a context switch from a Digression to some other DR).

One way to examine such *context switching* is illustrated in Figure 9. In this scenario, a context switch can occur when someone participates in one DR and finishes talking *after* that DR ends (thereby starting the next DR, since DRs occur one-by-one in sequence), or when someone or something (i.e., such as silence, a NonVerbal action) starts at the same time that the current DR ends. We can update the *ends* temporal query to the *context switching* situation in two simple mouse manipulations to the TVQL query filters (compare the top two query filters in Figure 8 to those in Figure 9), and we can also watch the visualization change *as* we manipulate the query filters. By comparing the visualizations in Figures 8 and 9, we see that there is a difference between the *ends* and *context switching* temporal relationships. These are just a few examples of temporal interactions between activities and DRs that we can study. We could continue exploring the data in this fashion, looking for other temporal trends.

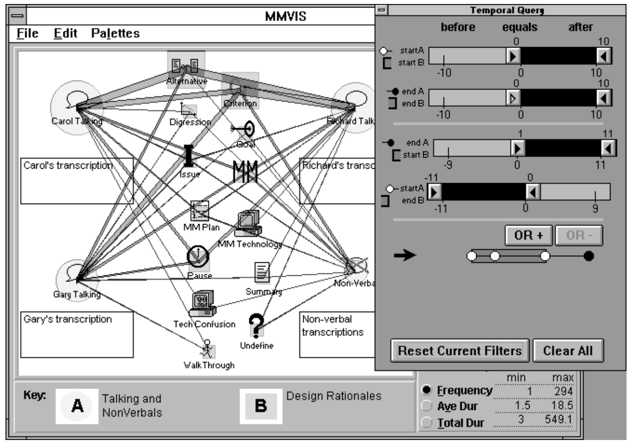

Figure 9. Visualization of the frequency in which Talking or NonVerbal events temporally start a new DR (i.e., which actions lead to a context switch*).*

3.4 Results of Temporal Analysis

In this subsection, we present a series of observations that can be obtained from reviewing visualizations in Figures 6 to 9. This provides us with some feedback on the capability of our proposed paradigm for temporal analysis.

Frequency vs. Average Duration vs. Total Duration. Figure 6 illustrates how even simple display options can provide information about the relationship between the frequency and duration of different types of events. Using this two-part figure, we can draw several conclusions.

Based on the *frequency* visualization (Figure 6a), we see among other observations that:

[Frequency counts range from 1 to 294.]

- Subjects talk with similar frequency, but Gary speaks the least frequently.
- NonVerbals occur less often than any one person talks.
- Alternatives and Criteria occur most frequently, followed by Pauses, Meeting Management, and other DRs.
- Carol and Richard appear to talk with similar but slightly higher frequency than that of Alternatives and Criteria.

Based on the *average duration* visualization (Figure 6b), we see the following:

[Average duration (aveDur) ranges from 1.5 to 18.5 seconds.]

- Average duration of NonVerbals is greater than any of the Talking events.
- Each person talks with about the same aveDur.
- Digressions have the largest aveDur, followed by Walkthroughs. Undefined and possibly Goals have the smallest aveDur. All remaining DRs have similar aveDurs which are greater than the aveDur of Undefined and less than aveDur of WalkThrough.

By comparing and contrasting parts (a) and (b) of Figure 6, we find for example:

- Digressions do not occur very often, but when they do occur, they last, on average, longer than other DRs. A similar situation is true for WalkThroughs.
- Although Alternatives and Criteria occur with the highest frequency, they do not occur with the highest aveDur.
- The total time that Gary speaks is less than Carol or Richard. This is indicated by the fact that Gary speaks with the smallest frequency but a similar aveDur. It is also verifiable by the Total Duration display option (see Figure 4).

These results lead us to some interesting observations, some of which are desirable and expected and some of which could lead to more in-depth analysis. For example, the low frequency of Digressions indicates that the subjects tended to stay on task (a desirable result). However, the larger average duration of Digressions indicates that when they do become distracted or go off-track, then more time is wasted until they get back on track. In order to confirm that the Digressions did *not* overpower the rest of the DRs in general, it is important to have the option to view the subset highlighters by total duration. This view was shown in Figure 4, and does indeed show that the total duration of Digressions is *not* the largest when compared to the other DRs.

In terms of Talking and NonVerbal events, the low frequency of the NonVerbals indicates that the subjects were talking more frequently than not talking (another probably desirable result). In addition, we see that Carol seems to speak with the largest frequency, but with slightly smaller average duration. Although it is not obvious whether her frequency and average duration is significantly different from those of the other subjects, it is a curious result. By examining Carol's transcripts, we see that she frequently uses short words and utterances of acknowledgment and encouragement, such as “uh-huh” or “ok.” It would be interesting to see if such a trend is correlated with female versus male subjects.

Frequency and average duration of Talking events could provide information about which person, if any, emerged as a leader of the simulated design meeting, even though no leader was appointed in the study. For example, one might expect a leader to generate lots of ideas and thus talk a lot. However, another sign of leadership can be the ability to listen (i.e., speak less frequently) and to facilitate and focus the direction of the meeting rather than spend a lot of time presenting ideas. Since Gary seems to speak the most infrequently, it would be interesting to see if he takes on a facilitator role. One way to answer such a question is to analyze Gary's participation in Digressions. That is, does Gary participate in Digressions? Does he initiate them? Does he frequently speak right after a Digression (i.e., does he move the discussion away from the Digression and back on track)? We can easily examine these and similar types of questions with our temporal analysis tools. Our tools thus allow researchers to embark on a much more detailed analysis of human behavior in a meeting setting, if so desired.

Relationships between Actions and Design Rationales. Figures 7 to 9 illustrate three different types of temporal relationships (corresponding to initiates, ends, and context switching) between Talking/NonVerbal actions and the thirteen DRs. Using these figures, we can identify various temporal trends, some of which are listed below:

Based on Figure 7 (*initiates* temporal relationships), we see that:

- Carol never initiates a Goal, Summary, or WalkThrough.
- Carol is the only one who initiates MM Technology and Technology Confusion.
- Richard never initiates an Issue.
- Gary never initiates a Digression.
- Carol initiates Meeting Management (MM) the most, while Richard initiates it the least frequently.
- A Pause is only initiated via a NonVerbal action.
- The strongest relationships are Talking/Alternative, Talking/Criterion, and NonVerbal/Pause pairs.

Although Carol strongly initiates Meeting Management (MM), the first four results indicate that she is probably not someone who acts as the leader of the design meeting. Gary's potential as a leader increases with the fact that he does not initiate a Digression and does initiate MM more than Richard. We expect NonVerbals to be highly correlated with a Pause, since silence is a NonVerbal event that would be categorized as a Pause in the design meeting.

Based on Figure 8 (*ends* temporal relationship), we see that:

- Carol never ends with or after a Goal.
- Only Carol ends with or after MM Technology.
- Only Richard ends with or after a Digression.
- Gary does not end with or after over half of the DRs.

Combined with the results of Figure 7, these results suggest that Carol is not very likely to participate in a Goal while Richard participates at both the start and end of a Digression (possibly an indication that he is easily distracted, though it may also indicate that while he starts a Digression, he does help in ending it). Gary's initiative nature, indicated by the fact that he initiates more DRs than he finishes, adds more weight to his potential to be the emergent leader of the group.

Based on Figure 9 (*context switching* temporal relationship), we see that:

- Richard is the only one who does *not* move the discussion off of a Digression.
- Richard appears to initiate switching from Alternatives and Criteria the most.
- NonVerbals initiate context switching from several of the DRs, including Digressions and Technology Confusion.

Gary's potential as a leader over Richard is strengthened even more by the fact that he helps direct the discussion away from a Digression while Richard does not. The utility of NonVerbal events to context switching indicates the importance of silence during meetings.

4. Evaluation and Discussion

In this section, we use the above results to compare and contrast our interactive visualization approach to timelines and statistically-based approaches.

MMVIS vs. Timelines. Figure 10 presents a timeline of events of the first four minutes of the twenty-five minute CSCW video data evaluated in our above MMVIS examples. In contrast to the timeline format, our temporal visualizations *cluster* the occurrences of events that occur over a time period of the video, rather than displaying them sequentially. This allows us to more easily highlight the relationships between different types of events. In the timeline format, not only are the occurrences of events spread out, but the temporal relationships are also distributed. This distribution effect could make it more difficult to compare lots of relationships over time as well as to isolate the non-existence of a relationship.

For example, in Figure 9, we only need to make one examination to see if there is a link between Richard Talking and Digression. Since no link exists, we can infer that Richard never provides a context switch from a Digression to another DR. Using the timeline in Figure 10, however, we need to scan through all Digressions or all instances of Richard Talking or some subset of the highlighted regions in order to derive the same conclusions.

Although the timeline could be filtered to only show the Richard Talking and Digression events that meet the context switching situation, this would not allow us to compare *all* context switching situations. In general, (unfiltered) timelines are limited in that they (1) are less compact than TViz, (2) require users to do more work by looking across rows and making their own deductions about relationships, and (3) inhibit users from directly making relative comparisons such as temporal relationships between different types of event pairs (e.g., comparison between (Gary, Digression) and (Richard, Digression) event pairs).

Note that we are not proposing our exploratory approach as a replacement to timeline-based approaches, but rather as a *complement* to these approaches. In fact, timelines are just another type of visualization—only less compact and summarized than the ones we propose.

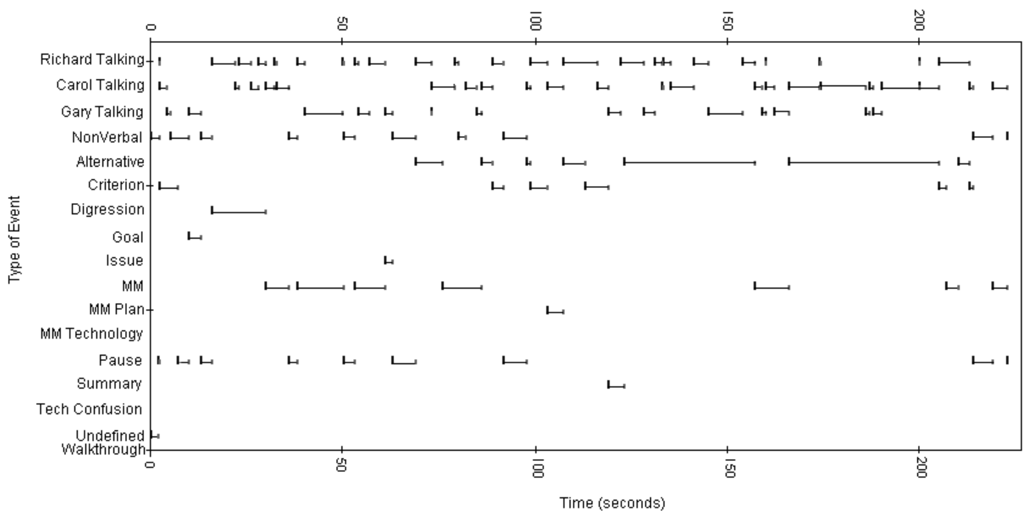

Figure 10. Timeline view of the first four minutes of the CSCW video events.

MMVIS vs. Statistics. The exploratory and interactive nature of our approach provides two powerful techniques not possible with statistical approaches: 1) users can use TVQL to *temporally explore* the data *without intentionally* stating a particular temporal relationship, and 2) users can see the *intermediate* query results as well as the final ones (since the visualization is updated *as* they manipulate the temporal filters). Users can explore temporal relationships by simply sliding the temporal query filter thumbs back and forth, in any order, and watching the visualization as it is being dynamically updated. They can continue in this fashion until an interesting visualization appears, and then use the descriptive labels and temporal diagram to identify what type of temporal relationship was specified. This approach could lead users to discover and explore temporal relationships that they might not have examined otherwise. Even when users have specific initial and final temporal queries in mind, the ability to see the intermediate results can similarly focus their analysis efforts in ways they had not initially intended.

In contrast, most statistically-based approaches are less flexible, requiring users to specify the desired temporal relationships *before* calculations are made. Statistically-based approaches can complement our approach, however, by providing a means to test for statistical significance, once the types of temporal relationships of interest have been identified.

Potential of MMVIS as a New Paradigm for IMMIR. Text-based information retrieval (T-IR) systems support users in forming a "model" document to retrieve (e.g., using a list of desired keywords and synonyms, the frequency or density of terms, the distribution of terms, etc.). In addition, many T-IR systems are further enhanced with ranking mechanisms—either ranking by the system (e.g., Callan et al. (1992)) or ranking by the user for relevance feedback (e.g., Salton and Buckley (1990)). In order to consider the various types of content- and structure-based information to form a model and to retrieve and rank similar documents, T-IR systems must be able to *analyze* documents in the collection.

While much progress is being made in IMMIR for locating objects, less work has been done to use the *structure* of multimedia (MM) documents for IMMIR. The *temporal structure* abstracted via subset selection and temporal queries within MMVIS could be directly used to form a model of the temporal structure of MM documents to retrieve. In this way, our temporal analysis approach is a right step towards more sophisticated IMMIR approaches.

For example, consider the scenario where the CSCW video used in our case study is one in a series of videos on design meetings. Suppose we wanted to know how many of these videos have events (e.g., design rationales (DRs), talking and nonverbal events) with temporal frequency similar to the case study video. We could use the subset selection palettes to select these events and view A&B Overlays by *frequency* (see Figure 6a). We could then use this visualization as a model of the temporal frequency of events, and pose a query such as "find all documents with similar temporal frequency density." Rather than displaying a text-based list of results, the system could provide visual thumbnail overviews of the "view by frequency" visualization for each of the retrieved documents. The results could be further enhanced by rank ordering the retrieved documents (e.g., documents with the same relative structure of frequency such as Alternatives and Criteria occurring more frequently than other DRs, Digressions occurring with relatively smaller frequency, etc. would be ranked fairly high). While the above discussion is based on *frequency*, the *average duration* and *total duration* views of subset selection could be similarly used in an IMMIR query.

Although recent IMMIR approaches are taking temporal relationships into consideration (e.g., "find all video clips where Dan Rather is speaking about politics" (Mérialdo and Dubois (this volume)), these approaches could be further enhanced with TVQL and its support for specifying temporal neighborhoods, fuzzifying temporal queries, etc. More importantly, the combination of TVQL and MMVIS allows us to take a more sophisticated look at the structure of temporal relationships. For example, if we used Figure 7 as a model representing who or what initiates a DR, then rather than getting a set of video clips such as Carol initiating a Digression, Richard initiating a Digression, etc., we would get a longer video document including these clips and meeting other constraints (e.g., that Gary does *not* initiate a Digression). Again, the retrieved results could be presented as rank-ordered, thumbnail overviews to visually and numerically indicate the relative match between individual documents and the query specified.

5. Related Work

While many bit-level video (BLV) analysis systems are beginning to support the specification of temporal constraints for information retrieval, such constraints are typically limited to simple sequences or intersections of temporal overlaps. For example, many researchers are focusing their efforts on BLV analysis of news broadcast video (e.g., see other chapters in this volume—Hauptman and Witbrock, Mani et al., and Mérialdo and Dubois). Many of these systems support queries such as "find all video clips where person P1 is talking about topic T1." When P1 talking is indexed separately from topic T1, this query represents the set of non-empty "intersection" video clips of P1 talking and topic T1 events. By incorporating TVQL, BLV analysis systems would have a more flexible interface for specifying temporal constraints. This would enable them to pose queries such as "find all video clips where person P1 *initiates* topic T1."

Work by Carroll et al. (1994) has also examined design rationale using video data, but their work has focused on retrieving video clips based on keyword matches rather than analyzing relationships between events. In order to meet the needs of temporal analysis, our work has focused more on retrieving information on temporal relationships rather than accessing video clips based on these relationships. While using our annotations to obtain direct access to the underlying video would be trivial to implement from a technical point of view, the difficulty is in providing such access in a meaningful manner. Consider the above query to “find all video clips where person P1 initiates topic T1.” Based on this query, we could retrieve any number of different combinations of P1 talking and topic T1 (e.g., we could retrieve clips that meet the temporal constraints but include: only P1 talking, only topic T1, or only the intersection of P1 and T1). Thus, while TVQL provides more flexibility in specifying sophisticated temporal constraints, it introduces new issues related to retrieving segments from the underlying document.

Other extensions to dynamic query filters and VIS have been explored (e.g., Fishkin and Stone (1995), Goldstein and Roth (1994)), but these extensions primarily focus on aggregation extensions to the interface. While these aggregation techniques could be incorporated into our system to enhance the formation of subsets, they do not address the temporal and relative exploratory needs of video analysis. Our TVQL and MMVIS environment represent significant extensions to VIS, tailoring the paradigm for temporal analysis.

Variations of timelines have been used for various purposes such as Gantt charts, timelines for video analysis, calendar visualizations (e.g., Mackinlay et al. (1991)), etc. Although *Timelines* (see Harrison et al. (1994)) provides support for coloring subsets of events, it does not incorporate tools for examining temporal relationships between these subsets. Instead, the users are left to scrolling through a timeline and using colors to look for temporal patterns in the video data. While parallel timelines are useful for examining the occurrence and sequence of events over time, they have limited utility for temporal analysis. Rather than placing the burden of looking for temporal trends on the user, we provide direct support for examining temporal trends within MMVIS through integrating our TVQL with our TViz. The power of MMVIS is not just that users can specify relative temporal queries, but that they can *incrementally refine* their queries and immediately see the corresponding changes to the visualization of retrieved results *as* they make refinements. A previous usability study by Ahlberg et al. (1992) on DQ filters and the VIS paradigm provides further evidence supporting the utility of this paradigm for trend analysis.

Although much work has been conducted in the area of video annotation and analysis (e.g., Davis (1993), Harrison et al. (1994), Mackay (1989), Roschelle et al. (1990), Sanderson et al. (1994)), this work has been limited in one or more of the following ways: 1) the work has focused more on novel approaches to creating annotations rather than analyzing them; 2) the temporal analysis hinges on *pre-coding* relationships rather than coding atomic information and searching or discovering temporal relationships (i.e., users can search for events, but they cannot search for *relationships* between events without previously identifying and explicitly annotating the relationships themselves); 3) visual presentations of the annotations are restricted to text- and timeline-based displays; 4) analysis is limited to temporal sequences rather than supporting all types of temporal relationships.

6. Conclusion

In this chapter, we presented a new paradigm for temporal analysis in which users can *temporally browse* data within an integrated MultiMedia Visual Information Seeking (MMVIS) environment. In MMVIS, specialized subset and temporal query filters (i.e., our Temporal Visual Query Language, TVQL) are coupled with a user-tailorable visualization of temporal relationships (TViz). TVQL provides a direct-manipulation interface for incrementally specifying and updating temporal queries and TViz is *dynamically* updated *as* users directly adjust the temporal filters, thus allowing users to browse the data in a temporally continuous manner. This interactive, exploratory technique for temporal analysis complements other timeline- and statistically-based approaches by providing a temporally clustered view of relationships and by focusing the direction of statistical analysis. We discussed the potential of MMVIS as a new paradigm for retrieving multimedia information based on temporal relationships between various events. This notion of taking temporal structure into consideration for retrieving MM documents is a novel approach to IMMIR which to our knowledge has not yet been explored.

We applied MMVIS to the temporal analysis of sample CSCW video data. This case study illustrated how our approach can be used to examine and identify temporal trends and how *different* types of temporal relationships (e.g., sequences and overlaps) can be easily explored within MMVIS. The MMVIS approach is generally applicable to any spatio-temporal data set (even if applied only to video data in this case study). Finally, our TVQL interface can enhance existing bit-level video analysis systems by providing a more flexible interface for specifying sophisticated temporal constraints.

7. Acknowledgments

This work was supported in part by a University of Michigan Rackham Fellowship, NSF NYI #94-57609, and equipment support from AT&T. Special thanks to Judy Olson for permission to use the sample data set.

References

Ahlberg, C., & Shneiderman, B. 1994. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In *CHI'94 Conference Proceedings*, 313-317. New York, NY: ACM Press.

Ahlberg, C., Williamson, C., & Shneiderman, B. 1992. Dynamic Queries for Information Exploration: An Implementation and Evaluation. In *CHI'92 Conference Proceedings*, 619-626. New York, NY: ACM Press.

Allen, J.F. 1983. Maintaining Knowledge About Temporal Intervals. *Communications of the ACM* 26(11): 832-843.

Callan, J.P., Croft, W.B., and Harding, S.M. 1992. The Inquiry Retrieval System. In *DEXA 3: Proceedings of the Third International Conference on Database and Expert Systems Applications*, 83-87. Berlin: Springer Verlag.

Carroll, J.M., Alpert, S., Karat, J., Van Deusen, M., Rosson, M.B. 1994. Raison d'Etre: Capturing Design History and Rationale in Multimedia Narratives. In *CHI'94 Conference Proceedings*, 192-197. New York, NY: ACM Press.

Davis, M. 1993. Media Streams: An Iconic Language for Video Annotation. *Telektronikk 4.93: Cyberspace* 89(4): 59-71.

Fishkin, K. and Stone, M.C. 1995. Enhanced Dynamic Queries via Movable Filters. In *CHI'95 Conference Proceedings*, 415-420. New York, NY: ACM Press.

Freksa, C. 1992. Temporal Reasoning Based on Semi-Intervals. *Artificial Intelligence* 54(1): 199-227.

Goldstein, J. & Roth, S. 1994. Using Aggregation and Dynamic Queries for Exploring Large Data Sets. In *CHI'94 Conference Proceedings*, 23-29. New York, NY: ACM Press.

Harrison, B.L., Owen, R., & Baecker, R.M. 1994. Timelines: An Interactive System for the Collection of Visualization of Temporal Data. In *Proceedings of Graphics Interface '94,* 141-148. Toronto: Canadian Information Processing Society.

Hauptman, A.G., & Witbrock, M.J. Informedia: News-on-Demand Multimedia Information Acquisition. In this volume.

Hibino, S. & Rundensteiner, E. 1996a. MMVIS: Design and Implementation of a MultiMedia Visual Information Seeking Environment. *ACM Multimedia'96 Conference Proceedings*. Forthcoming.

Hibino, S. & Rundensteiner, E. 1996b. A Visual Multimedia Query Language for Temporal Analysis of Video Data. In *Multimedia Database Systems: Design and Implementation Strategies,* eds. K. Nwosu, B. Thuraisingham, and P.B. Berra, 123-159. Norwell, MA: Kluwer Academic Publishers.

Hibino, S. & Rundensteiner, E. 1995. A Visual Query Language for Identifying Temporal Trends in Video Data. In *Proceedings of the 1995 International Workshop on Multi-Media Database Management Systems*, 74-81. Los Alamitos, CA: IEEE Society Press.

Mackay, W. E. 1989. EVA: An Experimental Video Annotator for Symbolic Analysis of Video Data. *SIGCHI Bulletin* 21(2): 68-71.

Mackinlay, J.D., Robertson, G.G., Card, S.K. 1991. The Perspective Wall: Detail and Context Smoothly Integrated. In *CHI'91 Conference Proceedings*, 173-179. New York, NY: ACM Press.

Mani, I., House, D., Green, M., and Maybury, M. Towards Content-Based Browsing of Broadcast News Video. In this volume.

Mérialdo, B. & Dubois, F. A Generic Tool for Content-Based Multimedia Browsing. In this volume.

Olson, G.M., Olson, J.S., Storrosten, M., Carter, M., Herbsleb, J., and Rueter, H. 1996. The Structure of Activity During Design Meetings. In *Design Rationale: Concepts, Techniques, and Use*, eds. T. Moran and J. Carroll. Mahwah, NJ: Lawrence Erlbaum Associates.

Olson, J., Olson, G., and Meader, D. 1995. What mix of audio and video is important for remote work. In *CHI'95 Conference Proceedings*, 362-368. New York, NY:ACM Press.

Roschelle, J., Pea, R., & Trigg, R. 1990. VIDEONOTER: A tool for exploratory analysis, Research Report, IRL90-0021, Institute for Research on Learning, Palo Alto, CA.

Salton, G. & Buckley, C. 1990. Improving retrieval performance by relevance feedback. *JASIS* 41(4): 288-297.

Sanderson, P., Scott, J., Johnston, T., Mainzer, J., Watanabe, L., & James, J. 1994. MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies* 41: 633-681.

Weber, K. & Poon, A. 1994. Marquee: A Tool for Real-Time Video Logging. In *CHI'94 Conference Proceedings*, 58-64. NY:ACM Press.

Zhang, H.J., Low, C., Smoliar, S. and Zhong, D. 1995. Video Parsing, Retrieval, and Browsing: An Integrated and Content-Based Solution. In *ACM Multimedia'95 Proceedings*, 15-24. New York, NY: ACM Press.