Temporal Mining of Artefacts Producing Meetings

Matt-Mouley Bouamrane and Saturnino Luz
Department of Computer Science
Trinity College Dublin, Ireland
{Matt.Bouamrane,Saturnino.Luz}@cs.tcd.ie

Abstract. Identifying salient parts of time-based media in multimodal meeting recordings is particularly challenging, as they often contain important information which can not be easily visualised (audio) or summarised. In an online artefact producing meeting scenario, certain self-contained information items are likely to be regularly manipulated. By capturing and timestamping participants’ space-based interactions with these items, one can define semantic objects with persisting histories. This information which is usually lost in common multimodal meeting settings offers new possibilities for meeting mining. Potential links between semantic objects, while not necessary obvious when looking at a meeting’s outcome can be uncovered by investigating temporal relationships between objects. Space-based actions also generally have strong associated semantics and are therefore appropriate for quick visual scanning. Drawing from our experience with the development of an integrated online collaborative writing environment with structured activity logging and post-meeting browsing system, we present a generic time-based artefact history model and novel time-based data mining paradigms for mining information from artefacts producing meetings. A prototype meeting browser, Meeting Miner is presented.

1 Introduction

The complexity of most projects performed in the workplace means that on a daily basis many tasks need to be carried out by teams involving people with various responsibilities and fields of expertise, sometimes residing in different places. This often means that projects are completed through an iterative process, during which they are sent back on forth between different experts. Phases of individual work are punctuated by meetings of some sort to discuss progress to date, share ideas, take decisions or allocate tasks. As computers have become ubiquitous tools for communication, possibilities for synchronous collaboration are greatly enhanced by the opportunities to capture meeting recordings and to alleviate participants of distracting or time consuming tasks such as note taking and meeting minutes production. However, as the number of stored meetings grow, so does the complexity of extracting meaningful information from the
recordings. Therefore, in order to be truly effective, a conferencing capture system needs to offer users efficient means of navigating recordings and accessing specific information.

In a specific scenario, which we will refer to as artefact focused meetings, one or many space-based items such as a textual documents, sketches, drawings, plans are either mentioned or produced during the meeting, either to support the decision-making process or in some cases, as the goal or focal point of the meeting (collaborative design). Figure 1 illustrates two examples of artefact producing meetings possible outcomes: a plan and a text document. Although the outcome is the obvious product of the collaborative meeting activities, it offers no clue about the often laborious process by which it was achieved. To illustrate this, consider the text of figure 1: it is a one sided A4 page document, yet it is the result of a collaborative writing task involving more than two hundred edits performed on a basic text skeleton. The document production process is either lost (outside the meeting participants’ individual memories) or needs to be recorded in a content rich continuous medium such as audio or video. In the latter case, the problem now consists in accessing relevant parts of the content rich media. It is a common workplace practice to have many ongoing projects at the same time, some of which might be put on hold while some condition outside the remit of the office is resolved. Consider a person who did not attend a meeting or needs to find a specific information months after the last meeting took place. While presented with the meeting outcome, one can clearly see a set of information but can be left wondering as to what are the reasons behind the choice of a specific item:

- Does this choice of material comply with fire regulations?
- Why did they decide to book a hotel outside of town?
To answer these questions, without listening to the entire recording, one needs means to access parts of the meeting recording where this information is most likely to be found: in other words efficient indexing and navigation tools. There is a growing research interest in producing applications for visual mining of multimodal meeting data in order to support users’ meeting browsing requirements. This paper is organised as follows: we first introduce related research work and existing systems developed for meeting browsing. We present a novel generic model for temporal mining of space-based artefact producing meetings. We then present the result of our own experience in developing an integrated online meeting capture and information mining architecture based on the design requirements earlier presented: the Meeting Miner.

2 Related Work

In recent years, modality translation from sequential data into the space domain: transcripts from audio through automatic speech recognition (ASR) and images from video (keyframes) has emerged as the dominant paradigm for continuous media navigation [1,2]. The MeetingBrowser [3] displays meeting transcripts time-aligned with corresponding sound or video files. The browser comprises a number of components, including a speech transcription engine and automatic summarizer. The summarizer attempts to identify salient parts of the audio and present the result to the user as a condensed script, or gist of the meeting. The SCAN (Spoken Content based Audio Navigation) [4,5] system uses acoustic and prosodic features for audio segmentation and a number of information retrieval techniques applied on ASR transcripts for speech recording indexing. SCAN user interface has three components: search, overview and transcript. The search component retrieves audio documents based on users’ queries match against the ASR transcripts of the documents contained in the database. The ten highest ranking documents are displayed along with the number of hits (number of terms of the query contained in the document transcript). The overview displays audio segments as rectangles colour-coded according to the terms from the user query and where the height is proportional to term frequency. The MeetingViewer [6] is a client application for browsing meetings recorded with the TeamSpace [7,8] online conferencing system. The TeamSpace's MeetingClient provides low-bandwidth video for participants awareness as well as supporting the use of a number of artifacts such as sharing and annotating slide presentations, creating and editing agenda and meeting action items and inserting bookmarks. In addition to participants information (joining, leaving meeting) all interactions events performed on the clients artifacts are automatically recorded and timestamped by the server. These events are subsequently used to index the meeting and are displayed on a timeline on the MeetingViewer interface to facilitate navigation. COMAP (COntent MAPper) [9,10] is a system for browsing captured online speech and text meetings. The user interface displays the textual outcome of the co-authoring task along with mosaic timeline views of participants’ speech and editing activities. An Interleave factor (IF)
metric measures levels of concurrent media activity, with intervals of greater activity deemed of greatest significance. A summary view of a recording can be generated through IF ranking. The Ferret Media Browser [11] is a client-server application for browsing recorded collocated multimodal meetings, with a combination of any available media for display and synchronised play-back. ASR transcripts, a key-word search and speech segmented according to speakers’ identity are also available. Media streams can be dynamically added to or removed from the display during the browsing task.

3 Object-based Temporal Mining

3.1 Key Concepts

The meeting scenario assumed in this paper is one where geographically dispersed users interact with a space-based document which acts as the focal point of the meeting, and can communicate through continuous media communication channels (audio and/or video). Furthermore, it is assumed that interactions with the space-based document will be computer mediated. This is important in the fact that participants’ interactions need to be automatically detected and recorded. Before we proceed, we need to define the following key concepts:

Document. The set of space-based artefacts used during the meeting. This is to a large extent the focal point of the meeting, either because the document supports the decision making process or is the meeting intended final outcome (workplan, plans). The document can be collaboratively written text, slides for a presentation, graphs, drafts and plans, for collaborative design, audio or video clips, or any combination of these.

Data Objects. Within the document, smaller data objects which can be treated and manipulated as individual semantic entities. The granularity of the semantic data objects is best defined according to the application scope. To illustrate this point, in common applications, a pixel or character would have no intrinsic semantics as opposed to a word, sentence, paragraph, a shape or image.

Primitive Operations. Primitive operations will typically modify some property of a data object. Example of primitive operations are insertion, deletion of characters in text, modifying the texture, size, of an image or shape.

Manipulation Rules. Manipulation rules need to associate an unambiguous and definite outcome when a data object manipulation affects other objects. Examples are cutting and pasting paragraphs or shapes, occlusions, etc.

Timestamp. A timestamp records information about all primitive and manipulation operations previously defined. Information recorded are: the agent who performed the operation, the nature of the operation, the time of operation and unless unfeasible, the exact content of the operation. Note that this is often partially the case in many existing applications: the nature and content of a number of past operations are stacked for undos and in collaborative applications, some form of timestamping needs to be implemented to address concurrency issues [12,13,14].
Object Log. For each object, a list of all timestamped actions (primitive and manipulation operations) which affected the object from creation throughout the meeting.

3.2 Object-based Mining

To provide access to multimodal recordings, one is faced with the challenge of structuring and integrating various orthogonal modalities (space-based vs. time-based) in an intuitive way for users. Continuous media such as audio and video, with time as inner structure, are difficult to access for lack of natural reference points, navigation is time consuming and can be confusing and summarisation is a non-trivial process. A study of users browsing and searching strategies when accessing voicemail messages, sometimes of very short duration (30s), showed that people had serious problems with local navigation of messages, and remembering messages’ content [15,16]. Many users performed time-consuming sequential listening of messages in order to find relevant information and often reported taking notes to remember messages’ content. In contrast, users displayed improved browsing performance, playing less audio when speech recognition transcripts were available as audio indexes in the user interface [17]. However, ASR currently suffers from a number of limitations. Disfluencies in spontaneous human-to-human dialogues, lack of word or sentence boundaries, poor recording conditions, crosstalk, inappropriate language models, out-of-vocabulary items and variations in speaking styles and pronunciations mean that for a certain percentage of people, some systems may have very low recognition rates [18]. In cases of an hour or longer meetings, even if the transcripts were of good quality, they may still represent quite a voluminous amount of information to scan through. Also, as spoken language is significantly different to written language, the transcripts may be difficult to decipher (due to style, repetitions, false starts, etc). Text-based information retrieval techniques can be applied to the ASR transcripts and one could argue that a key-word search can be an appropriate way of overcoming these shortcomings to quickly find specific information. However, in some cases a person looking for a specific information and the meeting participants who actually mentioned the relevant information during the meeting may use different words to refer to the same object. This would frustrate a keyword search approach and the user would need to reverse to reading the full transcripts.

More importantly, in the case where a meeting’s outcome is a space-based artefact, visual browsing may be a more appropriate way of navigating the meeting recording. Consider the outcome of figure 1a. Due to the graphical nature of the document, visual scanning is almost instant. In a collocated meeting scenario, one possible way of querying information from other participants would simply consist in pointing at a particular item and say “what about this?” Ideally, one would want to have access to all segments of the recording where the item in question is mentioned. For the reasons previously mentioned, current ASR technology can not guarantee to do this reliably. We here propose an alternative: by associating all space-based data objects with a log of all the actions which affected them during the course of the meeting, one can provide acces...
to all segments of the recording when a particular item was manipulated. We refer to this information mining paradigm as information retrieval from the data object perspective.

![Diagram of data stream and object manipulation](image)

**Fig. 2.** The data stream or timeline perspective

Common approaches to building visualisation and retrieval interfaces for browsing multimodal meetings emphasise linear access (whether sequential or random) due to the structuring role time naturally plays in multimedia data. Segmentation and indexing according to some features of the time-based media (speaker transition, shot detection) are used to define a number of media intervals. Access to specific media intervals is provided by some persistent representation of the time-based media (keyword, speaker identity, keyframes). By synchronised play-back of multiple media streams, a number of browsing systems \[6,11,19\] will ensure that space-based artefact manipulations concurrent with the current media interval will be visible to the user. In figure 2 selecting the media interval \(I_2\) will not only play the corresponding audio and video but will also display the nature of manipulations on the three objects: \(O_1\), \(O_2\) and \(O_3\) which were modified within the interval duration.

The paradigm shift we propose involves looking at inter media relationship from the object perspective. By logging all (relevant) information relating to the manipulation of a specific artefact, one can link these with all concurrent time-base media segments. Thus, access to the time-based media is now done through selecting a specific object, or specific actions performed on an object. This is illustrated by figure 3: selecting object \(O_1\) provides access to all time-based segments concurrent with a manipulation of \(O_1\): the three related intervals \(I_1\), \(I_2\) and \(I_4\). Our assumption is that in many cases, object manipulations will coincide with meeting participants focusing on the specific object. From an information retrieval perspective, this paradigm shift from time to object seems quite intuitive, shifting the emphasis from “what were people doing when they were discussing this?” to “what were they saying when they did this?”

---

1 Artefact manipulations have been represented as punctual objects to emphasise their nature as separate abstract entities. In reality, artefact manipulations are time intervals themselves (duration of manipulations)
3.3 Object Temporal Associations

One immediate property of having access to individual objects' history logs is that it enables us to discover potential associations between specific objects just by investigating the concurrency of actions performed on these objects. The following scenario illustrates how analysing temporal patterns of object manipulation uncover potential hidden information. The outcome of Figure 1a is the result of several meetings, each concentrating on resolving a specific issue. The final plan is flat: relationships between the different objects are not a-priori obvious. However, by analysing objects temporal information, we discover that in one of these meetings, the manipulation of the “Table” objects seems to be often concurrent to manipulations of the “Exit” item. Listening to audio segments where these objects are manipulated in close time proximity, the user discovers that in this particular project, the client’s preferred table layout is not compatible with fire regulations. As a result, the client’s original layout (no longer visible in the final outcome) had to be modified in order to accommodate the existing “Exit” and meet fire regulations, which explain the reasons behind the position of the “Table” objects in the final layout.
Figure 4 illustrates object temporal associations: objects $O_2$ and $O_3$ were on several occasions manipulated in close time proximity and as a result share a number of concurrent time-based media segments in their respective history log, indicating a potentially useful information link between these two separate objects. Object associations enables us to go beyond the information pattern illustrated in figure 3. Specific objects are not only linked to relevant segments of the time-based media, the wider context in which a specific object was manipulated during the meeting can be investigated within the context of other data objects manipulations. We thus define the concept of an object’s temporal neighborhood as the set of (i) time-base media segments concurrent with the object manipulation and (ii) actions performed on other objects within the previous time-based segments’ duration. We propose the following algorithm for retrieving an object temporal neighborhood:

- retrieve the set of all space-based actions performed on a specific object
- retrieve the set of all time-base segments concurrent with these actions
- retrieve all actions performed on different objects which took place within the duration of the previous set of time-base segments
- iterate through the 2 previous steps until no new actions or time-base intervals can be found

3.4 Action-based Browsing

![Fig. 5. Sequence of Actions](image)

We have so far defined space-based objects as potential information extraction and retrieval units. Another potential use of a log of space-based actions is as a navigation tool into the time-based media recording. Consider the simple sequence illustrated in figure 5: the upper part of the figure illustrates the evolution of the document during the course of the meeting while the bottom part shows the corresponding actions. One might wonder what prompted the choice of two different sets of table in the final outcome Fig. 5D. By visually observing the sequence of actions, a user may identify the exact moment when an action
of interest was performed (i.e: pasting round table), thus identifying a region of the time-based media where a potential explanation is likely to be found, as illustrated in Fig. 6. For this navigation method to be useful, only certain actions should be visible during browsing: for most applications, pixel or character-based operations are meaningless out of context. Operation filtering can be done at two stages. During the action logging stage, only actions which are potentially useful for information retrieval purposes are captured, while atomic operations are either discarded or buffered (into a more comprehensive operation). As previously mentioned, a potentially useful action entirely depends on a particular application’s scope. Information filtering can also be done at the post-meeting processing stage, where the user can dynamically choose what type of actions he is interested in (i.e: display only “paste” specific “type” of object). The definite appeal of such a navigation method is that space-based actions will generally have strong associated semantics and are appropriate for quick visual scanning, thus potentially offering a powerful indexing method into the time-based media.

Interactions are discrete, generally sparse enough so as not to overload a user with information, and tend to form natural semantic clusters over time (when a specific topic is discussed) allowing for discrimination and segmentation of topics within a meeting recording. This indexing method is also perfectly accurate in timing and content as it is not subject to recognition errors.

**Fig. 6.** Using space-based actions as a navigation tool

### 3.5 Advanced Query Model

The object history model presented in this paper offers a large choice of granularities of retrieval. If a user does not exactly know what he is looking for, a general object neighborhood retrieval, where all time-based media intervals and objects operations related to a specific object are retrieved may be appropriate, or, an action-based navigation, as detailed in the previous section. However, if the user is looking for a specific information, more advanced queries can be formulated by selecting a type of object with conjunction and disjunction of actions (primitive and manipulation) types and nature of actions attribute. An example
of advanced query is: select all objects round table, where operation is hatch and attribute is hatch pattern. The general form of an advanced query is:

retrieve object type ∧ action type ∧ action attribute.

4 Implementation: The Meeting Miner

The general space-based history model presented in this paper is the result of our experience in building an online collaborative writing environment with structured activity logging [20,21] and audio communication channel (Real Time Protocol multicast). In this environment, the chosen granularity of space-based artefact units for capturing operation logs are the paragraphs of text. These are self-contained information items with persisting histories when the segments are moved or altered. While paragraphs can be the subject of a number of physical manipulations like the ones previously described (cutting, pasting, moving, merging) the required number of primitives and manipulation rules applicable to paragraphs in a general usage scenario are limited, and therefore presented us with an ideal and manageable case study for the development of a space-based artefact history capture and management architecture. A detailed description of the timestamping model designed to manage paragraph history in case of modifications to document structure can be found in [22] and paragraph level retrieval and browsing as well as preliminary evaluation results can be found in [23]. The Meeting Miner user interface is presented on figure 7. We here describe the main components:

Document View displays the meeting final text document in the lower pane. Each paragraph in the document view is indexed for easy cross referencing. Participants' individual contributions can be highlighted according to a color code. When the paragraph-based view is selected, clicking on an individual paragraph will display the paragraph temporal neighbourhood.

Upper pane The upper pane can be one of two views, depending on how the user wishes to browse the meeting. In a paragraph-based retrieval, clicking on a specific paragraph will prompt the display of the tree-structured paragraph retrieval unit consisting of the content of editing nodes, and corresponding audio nodes, with the name of all the active participants within the duration of these temporal intervals (figure 8). An alternative view is the topic view, or contextual neighbourhood view. When this mode is selected, regions where audio contributions are likely to be related to the topic selection (a set of keywords selected in the topic panel) are highlighted Clicking on a particular interval will play the corresponding audio.

Tool Bar. Enables to set the system settings to best suit the user's preferences.
Keyword Panel. Displays all the potential keywords from the text document identified by the system. The list of keywords can be displayed in alphabetical
Fig. 7. MeetingMiner

Fig. 8. A paragraph Temporal Neighbourhood
order, frequency ranking or simply time of appearance. The user can dynamically update the list (removing words under a certain frequency or only select keywords associated with a certain type of action, etc.)

Topic Panel. The user can dynamically choose a set of keywords. A subsequent topic search will highlight audio segments associated with these keywords. The audio intervals selected by the topic search are segments in the neighbourhood of participants’ edits which contain the keywords.

Action Panel. Used in conjunction with the timeline navigator (slider) bar, it displays the nature of concurrent participants’ edits for action-based browsing.

Participants’ Panel. Displays the names of the participants. Each participant is assigned a unique colour code which highlights on the interface the ownership of the various text and audio contributions. A little audio icon is also displayed to show participants current activities (speaking, idle, etc).

Audio Panel. Provides sequential and random access to the audio file. The browser’s audio mode settings offers the user several navigation options such as skipping silences, or, if the topic mode is selected, jumping to the next topical segments. Similar functionalities were implemented in the SpeechSkimmer [24].

Timeline Navigator (slider). The navigator’s purpose is twofold: first, it offers a reference point into the audio recording. It also offers random access to the audio file. While moving the slider, participants’ concurrent actions are displayed in the Action Panel, so the user can decide to stop and listen to a specific section of the recording if he were to see an action of particular interest (as described in 3.4).

5 Conclusion and Future Work

Based on our experience with the development of an integrated online collaborative writing environment with structured activity logging and post-meeting browsing system, we have presented a generic time-based artefact history model and novel time-based data mining paradigms for mining information from artefacts producing meetings. Capturing and timestamping participants’ space-based interactions with data objects offers new possibilities for meeting mining. Investigating temporal relationships between objects uncovers potential semantic links which are not necessary obvious when looking at a meeting’s outcome. Space-based actions also generally have strong associated semantics and are therefore appropriate for quick visual scanning. The definite appeal of such a navigation method is that space-based actions will generally have strong associated semantics and are appropriate for quick visual scanning, thus potentially offering a powerful indexing method into the time-based media. Interactions are discrete, generally sparse enough so as not to overload a user with information, and tend
to form natural semantic clusters over time (when a specific topic is discussed) allowing for discrimination and segmentation of topics within a meeting recording. This indexing method is also perfectly accurate in timing and content as it is not subject to recognition errors. Future work will integrate ASR technology within the temporal mining architecture presented. A full evaluation of the Meeting Miner as a meeting mining tool will be performed.

References
