

# Navigating Multimodal Meeting Recordings with the Meeting Miner

Matt-Mouley Bouamrane and Saturnino Luz

Department of Computer Science  
Trinity College Dublin, Ireland

{Matt.Bouamrane, Saturnino.Luz}@cs.tcd.ie

**Abstract.** We present Meeting Miner, a multimodal meeting browser for navigating recordings of online text and speech collaborative meetings. Meetings are recorded through a collaborative writing environment specially designed to capture participants activities. This information, usually lost in common recordings of multimodal meetings, offers novel possibilities for indexing, navigation and information retrieval in archived meetings. Meeting Miner uses temporal information from the logs of actions captured on self-contained information items (paragraphs of text) to uncover potential information links between these semantic data units. A novel space-based action navigation scheme is presented. Keywords and topic search as well as more advanced queries can be performed by the system. We illustrate the system navigation modalities with several browsing examples.

## 1 Introduction

As computers become ubiquitous tools for communication over the Internet, more and more applications are being designed and implemented which support online remote synchronous collaboration, including environments for collaborative writing, learning and design [1, 2, 3, 4]. Computer mediated communication offers users the possibility of recording online meetings thus freeing participants from distracting and time consuming tasks such as note taking and production of minutes. 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 the recordings. Accessing specific parts of time-based media (audio/video) is particularly challenging as they often contain information which can not be easily visualised or summarised. The work presented in this paper is placed within the context of a growing research interest in applications for visual mining of multimodal meeting data in order to support users' meeting browsing requirements. The paper is organised as follows: we review existing paradigms and systems for multimodal meeting browsing, we then describe the specificity of our own meeting recording environment — a collaborative writing architecture with an audio communication channel and structured activity logging of participants interactions — and finally, present Meeting Miner and illustrate the system navigation modalities with several browsing examples.

## 2 Related Work

Multimodal meeting browsing is an open research area as it seeks to resolve issues relating to multimedia search and retrieval, visualisation and integration of multiple modalities, cooperative computing and cooperation modelling. Existing systems include the Meeting Browser [5] in which automatic speech recognition (ASR) transcripts are used for meeting navigation and summarization. The SCAN (Spoken Content based Audio Navigation) system [6] uses acoustic and prosodic features for audio segmentation and a number of information retrieval techniques applied on ASR transcripts for speech recording indexing. In MeetingViewer [7], interaction events are timestamped and subsequently used as an activity index in the user interface. COMAP (Content MAPper) and HANMER (HANd held Meeting browsER) [8] use an inter media activity metric for significant meeting events ranking. A detailed description of meeting browsing techniques and applications can be found in [9].

## 3 Meeting Recording Environment

The meetings targeted by our prototype are non-located computer mediated *speech-and-text* meetings typically involving a small number of participants. Recordings were produced with RECOLED [10, 11] (REcording COLlaborative EDitor), a collaborative writing environment, with structured activity logging and designed to be used along with an audio communication channel (RTP-Real Time Protocol multicast). In this environment, the chosen granularity of the data 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. After some pre-processing, the meeting recorder produces an archive consisting of: decoded audio files, a profile of user activity detailing the time of individual speech exchanges and an XML file containing the textual content of the shared document along with interaction metadata. Every time an editing operation is performed on the document, a corresponding timestamp is generated and associated with the paragraph in which the modification occurred

```
- <segment id="4.1">
  <timestamp actionid="17" agent="2" action="NewLine_Insert" start="215" end="215"/>
  <timestamp actionid="19" agent="2" action="Insert" start="215" end="217"/>
  <timestamp actionid="20" agent="2" action="Delete" start="220" end="222"/>
  <timestamp actionid="21" agent="2" action="Insert" start="221" end="221"/>
  <timestamp actionid="22" agent="2" action="Insert" start="222" end="226"/>
  <timestamp actionid="24" agent="1" action="Insert" start="231" end="231"/>
  <timestamp actionid="57" agent="2" action="Insert" start="486" end="495"/>
  budget of 3000 from the student union
</segment>
```

Fig. 1. Paragraph timestamps