

# Accessing Multimodal Meeting Data: Systems, Problems and Possibilities

Simon Tucker and Steve Whittaker

Department of Information Studies, University of Sheffield,  
Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK

{s.tucker,s.whittaker}@shef.ac.uk

**Abstract.** As the amount of multimodal meetings data being recorded increases, so does the need for sophisticated mechanisms for accessing this data. This process is complicated by the different informational needs of users, as well as the range of data collected from meetings. This paper examines the current state of the art in meeting browsers. We examine both systems specifically designed for browsing multimodal meetings data and those designed to browse data collected from different environments, for example broadcast news and lectures. As a result of this analysis, we highlight potential directions for future research - semantic access, filtered presentation, limited display environments, browser evaluation and user requirements capture.

## 1 Introduction

Several large-scale projects (e.g. [1,2]) have examined the collection, analysis and browsing of multimodal meeting data. Here we provide an overview of browsing tools, where we refer to any post-hoc examination of meetings data (e.g. searching a meeting transcript or reviewing a particular discourse) as *browsing*. As a result of our analysis, we are also in a position to highlight potential areas of research for future meeting browsers.

Despite being an emerging field, there are a large number of browsers described in the literature, and therefore the first stage of summarising the field was to determine a suitable browser taxonomy. The scheme used in this paper is to classify browsers according to their focus of navigation or attention. The taxonomy is described in more detail in Section 2 and is summarised in Table 1.

The structure of this papers is as follows. We begin by discussing how meeting browsers are classified and continue by describing each browser, according to its classification. A summary of all the browsers is then given, as a result of which we highlight directions for future research.

## 2 A Meeting Browser Taxonomy

Since browsing of meetings data is still an emerging field, the classification system used here is necessarily preliminary, but achieves a segregation of the range of browsers described in the literature. Browsers are classified primarily by their *focus*, and secondarily by properties unique to that focus. The focus of a browser is defined to be either the main device for navigating the data, or the primary mode of presenting the meeting data to the user.

Given this definition, and the range of data collected from meetings, three classes of browsers immediately present themselves. Firstly, there are browsers whose focus is largely *audio*, including both audio presentation [3] and navigation via audio [4]. Secondly, there are browsers whose focus is largely *video*; again, including both video presentation systems [5] and those where video is used for navigation [6]. The third class of browsers are focused on *artefacts* of the meetings. Meeting artefacts may be notes made during the meeting, slides presented, whiteboard annotations or documents examined in the meeting.

A fourth class of browser accounts for browsers whose focus is on *derived data forms*. Since analysis of meeting data is largely made on the nature and structure of conversation, this final class is largely concerned with browsing discourse. In this class are browsers whose focus is the automatic speech recognition (ASR) transcript and its properties [7], and those which focus on the temporal structure of discourse between participants [8].

This taxonomy is shown in Table 1 and each of the following sections describe each browser class in detail, in the order in which they were presented above. **We refer to audio and video indices as *perceptual* since they focus on low-level analysis of the data. Artefacts and derived indices are referred to as *semantic* since they perform a higher-level analysis of the raw data.**

**Table 1:** Overview of taxonomoy of meeting browsers and typical indexing elements used in each class

PERCEPTUAL	SEMANTIC
<i>Audio</i>	<i>Artefacts</i>
<ul style="list-style-type: none"><li>• Speaker Turns</li><li>• Pause Detection</li><li>• Emphasis</li><li>• User determined markings</li></ul>	<ul style="list-style-type: none"><li>• Presented Slides</li><li>• Agenda Items</li><li>• Whiteboard Annotations</li><li>• Notes - both personally and privately taken notes.</li><li>• Documents discussed during the meeting.</li></ul>
<ul style="list-style-type: none"><li>• <i>Video</i></li></ul>	<ul style="list-style-type: none"><li>• <i>Derived Data</i></li></ul>

**Table 1:** Overview of taxonomy of meeting browsers and typical indexing elements used in each class

PERCEPTUAL	SEMANTIC
<ul style="list-style-type: none"> <li>• Keyframes</li> <li>• Participant Behaviour</li> </ul>	<ul style="list-style-type: none"> <li>• ASR Transcript</li> <li>• Names Entities</li> <li>• Mode of Discourse</li> <li>• <b>Emotion</b></li> </ul>

## 2.1 Audio browsers

This section discusses browsers whose main focus is *audio*. We separate these browsers into two main subcategories. The first subcategory consists of audio browsers with detailed visual indices; the second category is audio browsers with limited, or no visual feedback.

Both Kimber *et al.* [9] and Hindus and Schmandt [10] describe a meeting browser whose primary means of navigation is via a visual index generated from speaker segmentation. The view presented to the listener is of participant involvement in the meeting - users are able to navigate to each speaker segment and can also navigate between neighbouring speaker segments.

Degen *et al.* [3] describe an indexed audio browser designed for visually reviewing recordings made with a personal tape recorder. The tape recorders allow users to mark salient points whilst recording, the marked recordings then being digitised for review on a computer. The computer interface affords several methods of browsing the recordings. Firstly, users can arbitrarily jump to any part of the recording, and can also navigate using the markings they made during the recording phase. The visual representation of the recording is of amplitude against time, displayed as a vector or colour plot. Users can also zoom in and out of this display and also have the ability to speed up playback (see the discussion surrounding *SpeechSkimmer* below).

A key element to these browsers is that the visual representations allow users to immediately see the structure of a meeting. This view, however, is dependent on the browsing environment allowing visual representations to be made. There are situations and devices which do not allow for this visual feedback, so that 'pure' audio browsing requires a substantially different interface.

*SpeechSkimmer* [4] is a system for interactive 'skimming' of recorded speech. Skimming is defined as system controlled playback of samples of original audio. A four level skimming system is implemented, each level compressing the speech further, whilst attempting to retain salient content. The first level is unprocessed playback, the second shortens pauses, whilst the third level plays back only speech which follows significant pauses. The final level uses an emphasis detector to select salient segments of the speech to present to the listener. On top of these skimming levels is a mechanism which allows the playback speed to be altered whilst maintaining the pitch of the speaker. In this way the playback speed can be increased without a significant loss in

perception. It should also be noted that the interface allows users to skim backwards in a recording - in this mode short segments of speech are played forwards but in reverse order.

Roy and Schmandt [11] describe a portable news reader implemented in a small, Walkman style device. The interface was designed iteratively in software, before being transferred to the hardware device. The resulting interface allowed listeners to playback a news report and to also navigate through the report using jump locations, computed from an analysis of pause lengths in the audio. In the course of designing the device it was noted that users preferred simpler, more controlled interfaces, preferring manual skims via jumping rather than having software controlled skims. The device also implements a form of speed-up similar to that described above, with users able to select from three different playback speeds.

Because of their nature, audio browsers are largely implemented in hardware devices and so can be argued to be distinct from meeting browsers making use of multimodal data. It has been seen, however, that these browsers have overcome the limitations of just audio and are able to provide means of browsing audio using computed indices and speed-up techniques. As a complement to this, the following section describes browsers whose primary focus is video.

## 2.2 Video browsers

The following class of browsers focus on *video*. Note that whilst each of these browsers have audio and video components, the main component for presentation or navigation in each case is video.

Foote *et al.* [5] describe a simple video browser with two primary modes of navigation. Firstly the user has the ability to jump arbitrarily to any section of the meeting, or to jump between index points which are precomputed from properties of the audio and video. The same indexing, when converted to a continuous 'confidence' measure can also be used to control the playback speed. For example, the playback speed could be related to gesture recognition so that portions of the meeting with significant gestures are played at different speeds, and index marks are made according to these significant gestures.

Girgensohn *et al.* [6] describe video interfaces centred around the use of *keyframes*. Keyframes are static images which have been automatically selected from continuous video according to some heuristic. In the video browsing system, keyframes are chosen according to an importance score, depending on the rarity and duration of each shot. Frames are then sized according to their importance (so that keyframes of higher importance are larger) and are placed linearly on the page. The resulting interface is then similar to a comic book or Japanese Manga drawings. This method can be used to produce a single summary of a full meeting and the user can playback salient portions of the meeting by selecting keyframes, or by choosing a point on a horizontal time line. A similar keyframe-based system was also developed at CMU [12].

A more complex video focused meeting browser is described by Lee *et al.* [13]. A novelty for this system is that it does not require a dedicated meeting room; instead, capture is performed by a single device, encompassing a camera which captures a

panoramic video of the meeting and four microphones to record audio. A real-time interface allows meeting participants to examine audio and video during the meeting, as well as making notes during the course of the meeting. The meeting is then archived and processed in preparation for browsing.

The browsing interface has a large number of navigational options. Central to the interface is the video screen, showing both the panorama and a close-up of the currently speaking participant. Users can navigate via a number of indexes, including representations of speaker transitions and visual and audio activity. There is also the opportunity to review an automatically produced transcript of the meeting, and to navigate the meeting via this transcript. A final option for navigating the meeting is a set of automatically generated keyframes. The interface also allows the user to review any notes made during the meeting and to examine any artefacts produced from the meeting.

We note that this class of browser is relatively small, mainly because video is largely supplemented with other browsing devices and is rarely solely used as a means of navigation. Furthermore, often meeting data does not contain the salient visual events that are useful for video browsing. The browsers described above, however, have shown that there is potential for making use of video as a means of browsing meeting data, although its value is not yet determined [12].

### **2.3 Artefact browsers**

The final browser classification based on data collected during the meeting is that of the *artefact* browser. We use the term artefact to describe any physical item recorded during a meeting which isn't audio or video. Browsers in this class fall into two subclasses: those which focus on slides presented and those which focus on notes taken by meeting participants. An important difference between this class of system and video or audio browsers is that artefacts are usually *searchable*, making it possible to both browse and search data. We will discuss each subclass in turn.

Brotherton et al. [14] describes a system for the visualisation of multiple media streams for the *Classroom 2000* project. The project views a classroom lecture as a multimedia authoring program, trying to extract useful information from classroom lectures to present in a suitable form to the user. The resultant interface is web-based and shows the slides used during the presentation. Slide transitions are indexed, allowing the user to jump between the segments of the lecture relating to each slide. The slides can be manually annotated both during and after the lecture and this information is also indexed. Further to the visual interface, audio segments relating to the slides can be played back.

Users can search through transcribed audio, slide text and lecturer annotations. In this way students are able to see how different topics relate to each other, as well as being able to locate specific information from a series of lectures.

Whilst the slide browsers described above have used lectures as their data source, there are slide browsers which have examined meetings data. An example of this is the TeamSpace project described in [15]. TeamSpace supports the organising, recording and reviewing of meetings; we shall, however, focus on the interface used to review archived meetings.

The interface consists of two main components. Firstly, there are two indexes - the first giving an overview of the full meeting and the second being of a detailed portion of the overview. The second main component for browsing is a tabbed pane containing annotated slides, agenda items and video displays. The slide view has an index showing each of the slides discussed in the meeting, so that users can jump to relevant portions of the meeting; furthermore, there is a larger view of the slide currently being discussed. A similar approach is taken to showing the meeting agenda, where the agenda acts as an index for the meeting and an indicator of the agenda item currently being discussed. Cutler *et al.* [16] describe a meetings browser in which the central component is captured images from a whiteboard. The interface also contains a participant and whiteboard index, allowing users to jump to particular segments of the meeting, or to review segments which relate to specific elements of the whiteboard annotations. Furthermore, two video components are included - a panorama of all the participants and a close up view of the current speaker. In addition to these components the browser also allows the user to speed up playback, and also to skip the contributions of selected participants.

The browsers described above have focused their attention on presenting *community artefacts* - those which can be altered or viewed by all meeting participants. The final set of browsers in this class examine more private artefacts; specifically, they make use of notes made by participants as a means of indexing and browsing meetings.

Whittaker *et al.* [17] outline the *Filochat* system which combines an audio recorder with a tablet for taking notes as a means of constructing a meeting record. The tablet acted as a virtual notebook and allowed users to store several pages of notes and organise them into sections. Users can then use the notes they have taken to jump to the relevant portion of the conversation. The interface also affords the ability to manually navigate the audio by jumping forwards and backwards. The system was tested on users both in the field and in lab experiments.

The use of notes to assist recall of meetings was also investigated by Moran *et al.* [18]. The data used for the study was collected from meetings chaired by a single person, in which audio and notes taken both on a shared whiteboard and by the meeting chair were recorded and timestamped. The meeting records were then used by the meeting chair to make technical decisions on the basis of what was said at the meeting. A detailed study of how the chair used the meeting record over a large period of time was made and identified not only how the meeting record was used, but also how the chair's use of the meeting record changed over time. This analysis identified, for example, that the chair would often annotate his notes with the word "ha", meaning that something interesting had occurred and that it would be useful to revisit this section of the meeting during the review process.

The browsers described above examine browsing of artefacts, specifically slides and notes taken by participants. User notes are a powerful means of indexing meetings, since they become both a user-defined index and a means of clarifying any confusing notes taken. Equally, however, a slide index allows users to clarify any confusion originating in presentations. We now discuss the final class of meeting browsers.

## 2.4 Discourse browsers

The final class of meeting browsers are focused on *derived elements* of meetings, specifically components such as ASR transcripts or participant contributions. This class of browsers is loosely segregated into those which focus on the transcript, those which focus on participant contributions and those whose focus is a combination of derived and raw data. Because they present ASR transcripts, these systems, like artefact browsers, not only allow browsing but also offer the ability to search.

Whittaker *et al.* [19] describe a system, *ScanMail*, for browsing voicemail messages using a similar interface to that used to browse email. Incoming voicemail messages are passed through a speech recognizer, and the ASR transcript produced by this recognizer forms the body of the “email”. The user has options to playback the message by clicking on the transcript at any point, and can also alter the playback speed. Furthermore users can search the transcripts from multiple voicemail messages in order to identify messages and segments of interest. *ScanMail* also extracts important entities such as phone numbers and names from the transcript. A similar system, *Scan* [20], supported browsing and search of broadcast news, including visual overviews of search results.

*Rough ‘n’ Ready* [21] is a news browser which focuses on the ASR transcript and various views derived from this transcript. The system allows the user to search for named entities such as people, locations, organizations and automatically derived topic markers. The interface also has a timeline, which allows users to view the temporal density of their search results and to navigate between the results and an automatically derived speaker index. In keeping with the *ScanMail* system, users can select any part of these indices or transcript elements to navigate the news reports.

Whilst these browsers have the transcript as the central focus of the interface, Bett *et al.* [22] describe a meeting browser in which the transcript is given as much prominence as a video component. The interface also contains a participant index, which indexes single or groups of speakers. In addition to these components the browser also allows the user to construct audio, video or text summaries, using text processing, for complete meetings or salient segments of the meetings. The summary is based on the transcript data and the audio and video streams are segmented accordingly to fit with the reduced transcript. The browser also supports search of a large meetings archive and indexing of discourse features and detected emotions.

The *Ferret* browser [8] also features the transcript alongside video and participant indexes. A key feature of the browser is that additional temporal annotations can be added or removed at will. For example, it is possible to add automatically derived agenda and interest indices whilst browsing a meeting. The interface is contained in a web browser, and so the transcript can be searched much like a web page, using the

browser facilities. As with other browsers, users can navigate through the meeting by clicking on the transcript or by using the derived indices. The index view is customisable and can be viewed at a variety of different zoom levels.

The *Jabber-2* system described by Kazman *et al.* [23] has many similarities with the Ferret browser. Central to Jabber-2 is the temporal view, showing the involvement of each participant. Further to this participant view is a set of keywords relating to the meeting currently being browsed. In addition to these indices is a stage of discourse mode recognition which constructs an overview of a meeting by plotting a graph showing the amount of involvement each participant had in each segment of the meeting. A previous study [24] described an alternate implementation of Jabber, denoted *JabPro*, which contained a video component and allowed users to search the meeting transcript to identify where in the meeting keywords occurred.

Also included in this class is a browser described by Lalanne *et al.* [7]. Here, the transcript is supplemented with audio and video controls, as well as a view of any documents currently being discussed. Furthermore, the meeting is indexed according to participants and properties of the documents and discourse occurring throughout the meeting. A key element to this interface is that every component is time synchronised, so that any changes or transitions in one component is automatically reflected in all the other components of the interface.

Since they make use of both raw and derived data, browsers in this category tend to have a more complex interface than those discussed in the previous classes. By segregating the interface into browsing and indexing components the browsers described in this class have overcome this complexity. Furthermore, this increased complexity has allowed for complex interface components, such as a search facility.

### 3 Summary

We have analysed browsers designed for reviewing multimodal data captured from meetings. It has been seen that these browsers can be distinguished using the focus of their presentation and navigation. Specifically, we segregated browsers into those that are focused on audio, on video, on non audio-visual artefacts, and on elements derived from this raw data (see Table 1). One observation is that a typical meetings browser consists of two main classes of components. Firstly, there are presentation elements - which are essentially realisations of the raw data; for example, audio, video and views of discussed documents. Secondly, there is an index component. This can include indexes of participant involvement, artefact changes such as slide changes, and higher-level properties of meetings such as the agenda.

It is interesting to note that development of meeting browsers has largely focused on making use of elements on the left hand side of Table 1, with elements on the right hand side being used as indices alone. The expected use of such browsers is *index centric random access* with users navigating and identifying points of interest using the indices and then reviewing that particular portion of the meeting using the display. There are, however, other procedures for accessing meeting data which become apparent once we focus on textual data and exploiting semantic information available from the right hand side of Table 1.

Below, we outline four potential areas for future research. The first two areas concern different modes of browsing meeting data, the third considers browsing with limited resources and finally we discuss the lack of evaluation and user requirement determination in current browsers. **Note that whilst the potential features have been applied to some of the browsers described above, we feel that the areas would benefit from further research.**

### **3.1 Search, topic tracking and summarisation**

By focusing on index based browsing, most browsers have ignored semantic techniques such as search, topic tracking and summarisation. The availability of transcript data generated from meetings means that search functionality is relatively simple to implement and it is surprising that it is not more widely used in current meeting browsers. As an example of a meeting browser with search, JabberPro [24] implemented a keyword searching algorithm allowing salient portions of the meeting to be identified by providing a suitable keyword. Jabber-2 [23] also supported topic tracking.

Furthermore, only a small number of the browsers discussed above made use of summarisation (most notably [4]). **Whilst the availability of the ASR transcript does not make the production of a summary straightforward, it should be noted that users are able to make good use of poor transcripts [CITE] and so it is possible that user would be able to make use of a weak summary. Summaries may not only be useful for certain meeting participants, for example minute takers, but also for controlling the meeting presentation (see below).** Another area to explore is entity extraction [21,19]. All these areas involve text processing; a potentially promising area of future research is to make use of such techniques for analysing and presenting meeting data.

### **3.2 Filtered Presentation**

A second area for future research is that of filtered presentation. In the current set of browsers it is assumed that users will want to manually review meetings by looking at indices and then browsing using these. However, there may be an advantage in using the derived indices to *control* the other components rather than to navigate them e.g [4,5].

Consider the use of a search component in [25]. Here the user enters the search terms, and the result of the search then becomes a new index with which the user can navigate the meeting. An alternate approach to using the search results would be to play back only sections of the report which relate to the search terms. The advantage of this approach is that the browser becomes a more active tool for navigating meetings and users are able to playback areas of interest directly, rather than having to determine these areas themselves.

### **3.3 Limited resource browsing**

Another common assumption made by browsers discussed in this paper is that they will be accessed solely through a computer workstation with a high resolution display. Whilst the audio browsers are naturally suited to less functional hardware, it could be advantageous to have access to a different type of browser in a limited environment.

Research in this direction should address several questions. Firstly, in relation to the previous section, studies should identify the needs of users in these environments. This, in turn, will address what sort of components the limited resource browsers will require. In this environment there will also be a problem of how to make the best use of screen space; it can be argued that textual representations would be advantageous here, since text is a relatively compact representation compared to the video or artefact views described above.

Furthermore, there are a large number of technical problems that the limited resource browsers should address, for example how the device can gain access to the relevant data. There could also be advantages in developing browsers for use during meetings - allowing an *in situ* review of the meeting to clarify current discussions.

### **3.4 User-driven development and evaluation**

Another key area addressed by a small number of the browsers discussed above, is that of measuring the *quality* of the browser and, related to this, how well the browser meets the user requirements for reviewing meeting data. The audio browsers are good examples of using user requirements to drive the design and evaluation of the browser. The NewsComm [11] audio browser went through 4 iterations of different interfaces in order to identify both what functionality users required and also how they should be able to access this functionality. Furthermore both Arons [4] and Whittaker *et al.* [19] describe lengthy user evaluation studies which examine not only how well their systems function but also evaluate the use of specific components of the systems.

It could be argued that the reason for this lack of evaluation of browsers is that, since the field is still emerging, the browsers are designed to examine the success of the underlying technologies and, therefore, evaluation of the browsers is a secondary concern to that of evaluation of the technology. Since the technology has now reached a sufficient level of maturity, however, it can be seen that robust evaluation of browsers must be considered for any new meeting browsers. With respect to this, it is promising to note that some effort is being made to be able to evaluate new browsers with respect to browsers previously developed - see [26].

## **4 Conclusion**

This paper has examined the state of the art of meeting browsers. We segregated the field into four classes, three derived from browsers focused on data collected from meetings and the fourth being browsers whose focus is derived data. We also identified four areas for future research: semantic access, limited resource browsing, filtered presentation and rigorous development and evaluation.

## References

- [1] M4 Project, <http://www.m4project.org/>
- [2] IM2 Project, <http://www.im2.ch/>
- [3] Degen, L., Mander, R., Salomon, G.: Working With Audio: Integrating Personal Tape Recorders And Desktop Computers. In: Proceedings of CHI '92, Monterey, CA, USA (1992) 413-418
- [4] Arons, B.: SpeechSkimmer: A System for Interactively Skimming Recorded Speech. ACM Transactions on Computer-Human Interaction (1997) 3-38
- [5] Foote, J., Boreczky, G., Wilcox, L.: An Intelligent Media Browser Using Automatic Multimodal Analysis. In: Proceedings of ACM Multimedia, Bristol, UK (1998) 375-380
- [6] Girgensohn, A., Boreczky, J., Wilcox, L.: Keyframe-based User Interfaces For Digital Video. IEEE Computer (2001) 61-67
- [7] Lalanne, D., Sire, S., Ingold, R., Behera, A., Mekhaldi, D., Rotz, D.: A Research Agenda For Assessing The Utility Of Document Annotations In Multimedia Databases Of Meeting Recordings. In: Proceedings of 3rd International Workshop on Multimedia Data And Document Engineering, Berlin, Germany (2003)
- [8] Ferret Browser, <http://mmm.idiap.ch/>
- [9] Kimber, D.G., Wilcox, L.D., Chen, F.R., Moran, T.P.: Speaker Segmentation For Browsing Recorded Audio. In: Proceedings of CHI '95 (1995)
- [10] Hindus, D., Schmandt, C.: Ubiquitous Audio: Capturing Spontaneous Collaboration. In: Proceedings of 1992 ACM Conference on Computer-Supported Cooperative Work, Toronto, Ontario, Canada (1992) 210-217
- [11] Roy, D.K., Schmandt, C.: NewsComm: A Hand-Held Interface for Interactive Access To Structured Audio. In: Proceedings of CHI '96, (1996)
- [12] Christel, M.G., Smith, M.A., Taylor, C. R., Winkler, D.B.: Evolving Video Skims Into Useful Multimedia Abstractions. In: Proceedings of CHI '98, Los Angeles, CA (1998)
- [13] Lee, D., Erol, B., Graham, J., Hull, Jonathan J., Murata, N.: Portable Meeting Recorder. In: Proceedings of ACM Multimedia, (2002) 493-502
- [14] Brotherton, J. A., Bhalodia, J. R., Abowd, G. D.: Automated Capture, Integration and Visualization of Multiple Media Streams. In: Proceedings of The IEEE International Conference on Multimedia Computing And Systems, (1998) 54-63
- [15] Geyer, W., Richter, H., Fuchs, L., Frauenhofer, T., Daijavad, S., Poltrock, S.: A Team Collaboration Space Supporting Capture And Access Of Virtual Meetings. In: Proceedings of 2001 International ACM SIGGROUP Conference On Supporting Group Work, Boulder, Colorado (2001) 188-196
- [16] Cutler, R., Rui, Y., Gupta, A., Cadiz, J.J., Tashev, I., He, L., Colburn, A., Zhang, Z., Liu, Z., Silverberg, S.: Distributed Meetings: A Meeting Capture And Broadcasting System. In: Proceedings of 10th ACM International Conference on Multimedia, Juan-les-Pins, France (2002) 503-512
- [17] Whittaker, S., Hyland, P., Wiley, M.: Filochat: Handwritten Notes Provide Access To Recorded Conversations. In: Proceedings of CHI '94, Boston, Massachusetts, USA (1994)
- [18] Moran, Thomas P., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., Melle, W., Zellweger, P.: "I'll get that off the audio": A Case study of salvaging multimedia meeting records. In: Proceedings of CHI '97, Atlanta, Georgia (1997)

- [19] Whittaker, S., Hirschberg, J., Amento, B., Stark, L., Bacchiani, M., Isenhour, P., Stead, L., Zamchick, G., Rosenberg, A.: SCANMail: A Voicemail Interface That Makes Speech Browsable, Readable and Searchable. In: Proceedings of CHI 2002, Minneapolis, Minnesota, USA (2002)
- [20] Whittaker, S., Hirschberg, J., Choi, J., Hindle, D., Pereira, F., Singhal, A.: SCAN: Designing and Evaluating User Interfaces to Support Retrieval from Speech Archives. In: Proceedings of SIGIR99 Conference On Research And Development In Information Retrieval, Berkley, USA (1999) 26-33
- [21] Colbath, S., Kubala, F., Liu, D., Srivastava, A.: Spoken Documents: Creating Searchable Archives From Continuous Audio. In: Proceedings of 33rd Hawaii International Conference On System Sciences, (2000)
- [22] Bett, M., Gross, R., Yu, H., Zhu, X., Pan, Y., Yang, J., Waibel, A.: Multimodal Meeting Tracker. In: Proceedings of RIAO, Paris, France (2000)
- [23] Kazman, R., Kominek, J.: Supporting the Retrieval Process In Multimedia Information Systems. In: Proceedings of Proceedings of the 30th Annual Hawaii International Conference On System Sciences, Hawaii (1997) 229-238
- [24] Kazman, R., Al-Halimi, R., Hunt, W., Mantei, M.: Four Paradigms for Indexing Video Conferences. *IEEE Multimedia* 3(1) (1996) 63-73
- [25] Chiu, P., Boreczky, J., Girgensohn, A., Kimber, D.: LiteMinutes: An Internet-Based System For Multimedia Meeting Minutes. In: Proceedings of 10th WWW Conference, Hong Kong (2001) 140-149
- [26] Flynn, M., Wellner, P.D.: In Search of a Good BET: A Proposal for a Browser Evaluation Test. IDIAP IDIAP-COM03-11 (2004)