

TQB: Accessing Multimodal Data Using a Transcript-based Query and Browsing Interface

Andrei Popescu-Belis and Maria Georgescu

ISSCO / TIM / ETI
University of Geneva
40, bd. du Pont-d'Arve
1211 Geneva 4, Switzerland
andrei.popescu-belis@issco.unige.ch, maria.georgescu@eti.unige.ch

Abstract

This article describes an interface for searching and browsing multimodal recordings of group meetings. We provide first an overall perspective of meeting processing and retrieval applications, and distinguish between the media/modalities that are recorded and the ones that are used for browsing. We then proceed to describe the data and the annotations that are stored in a meeting database. Two scenarios of use for the transcript-based query and browsing interface (TQB) are then outlined: search and browse vs. overview and browse. The main functionalities of TQB, namely the database backend and the multimedia rendering solutions are described. An outline of evaluation perspectives is finally provided, with a description of the user interaction features that will be monitored.

1. Introduction

The recording of interactions occurring in group meetings generates large amounts of data in several modalities. The use of this data for the study of multimodal human interaction, or for information extraction applications, requires the capacity to analyze, store, index and render various information flows. To index multimodal meeting recordings, metadata and annotations derived from the analysis of their content are especially useful. In this article, we show how a shallow approach to the annotation of language-mediated interaction can be used in a *transcript-based query and browsing interface* (TQB). This annotation-driven meeting browser provides a reusable and cost-effective technical solution, on condition that the meeting annotations have reasonable accuracy.

TQB represents the end-user's interface to a complex automatic *meeting processing and retrieval* (MPR) application. Such an application would enable people who did not attend a meeting (for instance a staff meeting or a business meeting), or people who want to review a specific part of a meeting that they have attended, to search for a particular piece of information connected to the meeting, and then to browse the various communication modalities that were recorded, in order to situate the retrieved piece of information in its context. For instance, this kind of interactive access to databases of meetings would greatly enhance access to corporate knowledge and memory, and would help the analysis of decision-making processes.

In Section 2 of this article, we review a number of constraints on the design of interfaces for MPR applications, based on the analysis of modalities, user studies, and considerations of feasibility. In Section 3 the data and annotations which can be searched and browsed with TQB are reviewed. Two scenarios of use are proposed in Section 4, while the main functionalities supporting them are outlined in Section 5, in particular the database backend and the multimedia rendering devices. An outline of evaluation perspectives appears in Section 6, listing the user interaction features that will be monitored.

2. Constraints on Interface Design

The design of TQB was determined by an analysis of the modalities recorded in our project's Smart Meeting Room (Wellner, Flynn & Guillemot, 2004), together with the objective of remaining close to the structure of the data and annotations, in order to enable developers of MPR applications to browse directly the data and annotations as they are stored in the meeting database. In addition, initial user studies provided indications on the most likely modalities that are needed to answer queries about the content of meetings (Lisowska, Popescu-Belis & Armstrong, 2004).

2.1. Media and Modalities for MPR

Of the five human senses, current human-computer interaction techniques are mainly based on sight, hearing and touch (Bernsen, 2002). The term *medium* is generally used to refer to the human sense used in the communication channel, while the term *modality* refers the form of information communicated through the channel. A medium can convey various modalities, and a modality can often be conveyed through various media. Examples of modalities (conveyed by various media) are linguistic form (spoken, written, or signed language), graphic form (pictures, graphs, diagrams) and gesture form.

In the data acquisition stage of an MPR application, the recording of modalities conditions the whole range of subsequent processing possibilities, therefore as many modalities as possible should be recorded. The limiting factors here are the input devices and the available storage space: for instance, while cameras and microphones are widespread, devices to record eye gaze, heart rate, or EEG are still far less common.

As regards the processing of meeting recordings, i.e. the extraction of the most relevant information, the highest informational content appears to be conveyed by *language*, as it occurs in conversations (speech available as audio or transcript) or in meeting documents (reports, slides). Other relevant modalities are face expressions (from video), the positions of the participants (from video) and their emotions (from speech and video).

Of course, different modalities can be used for the interface to the database of processed meeting recordings. For instance, such an interface can be GUI-based (including buttons, forms, etc.) or it can be language-based (spoken or written on a display). The capacity for human-computer dialogue is useful for sequences of query / answer pairs that refine an initial query, provided a robust mechanism for such a dialogue can be defined.

The data output by the meeting storage system in response to a query, or for use in a browser, can occur in a variety of formats – from timestamps or chunks of recordings to complex “facts” derived by an inference engine from the processed meeting data.

2.2. Usability and Feasibility

Our final choices for the TQB interface are based on a trade-off between feasibility and (potential) usability, which was estimated from initial user studies of meeting browsers. In a preliminary experiment that attempted to find out a range of requirements for meeting browsers (with four scenarios of use including the two scenarios mentioned in section 1), a number of constraints on the recorded modalities, their annotations, and the query processing capabilities were outlined (Lisowska, Popescu-Belis & Armstrong, 2004). These constraints highlight the importance of accessing the meeting recordings (audio) and their transcripts, the meeting documents, and metadata about the meeting participants, location, etc. – that is, some form of “understanding” of the human dialogues involved in meetings. In addition, the study showed the importance of elaborate query processing in order to be able to answer complex queries by combining online information derived offline from the meeting – a feature that is not yet targeted within our project.

However, not all these indications from the user study could be implemented, given feasibility constraints. TQB is therefore a graphical data-driven interface that gives access to the data and the annotations derived offline, but that does not attempt to do complex processing of user queries. Queries can only search for a specific utterance in a meeting, which is then situated context of the meeting and can be browsed.

In the present approach, the output can consist only in one (or more) of the media that were used for meeting recording, together with any of the annotations that were done through meeting processing. The rendering of transcriptions comprises visual rendering of the master transcript, and links to other modalities, in particular, the sound tracks and the meeting documents. The video tracks are also potentially accessible; however their overall informational content appears to be quite low, since participants do not change positions throughout a meeting (in most cases). TQB is thus a transcript-based interface with query/browsing coupling, and enriched multimedia rendering of the results. Other interfaces are also being developed for the multimodal meeting data generated in our project (Wellner, Flynn & Guillemot, 2004; Lisowska, Rajman & Bui, 2005).

3. Description of Multimodal Data

The TQB meeting browser is developed within a large project aimed at multimodal information management (see Acknowledgments). A number of multimodal meeting corpora were developed for use within the project by the

various partners. Most of the meetings were recorded in smart meeting rooms (essentially at IDIAP / Martigny, ICSI / Berkeley, TNO / The Netherlands, and at the University of Edinburgh), either according to pre-defined scenarios or as naturally occurring meetings. In one of the corpora, for instance, the scenarios simulate product design processes in a small company.

Most of the data is available at present through a media file server (<http://mmm.idiap.ch>) with public and private sections, and includes the following recorded modalities: sound (head, lapel, desktop mikes), video (individual and global cameras), slide show, pen and whiteboard capture, and storage of written documents (Wellner, Flynn & Guillemot, 2004).

3.1. Available Data

The main sources of multimodal meeting recordings are listed by order of availability in Table 1 below, according to the source (institution or project) that is the main contributor to the resource. Typically, each corpus is composed of sessions (group meetings), which are sometimes structured in series of up to four related meetings. All the recordings contain at least the audio tracks (often on separate individual channels) and their transcripts, and most of them have also video (one or more channels) and document capture. The available annotations are described in the next subsection.

Source	Nb.xtime	Media	Lang.	Annotations
ICSI-MR	75 x 60'	A T	EN	utterances, dialogue acts, discourse markers, episodes*
IDIAP	60 x 5'	A V T	EN	utterances, episodes
ISSCO	8 x 30'	A V T D	EN	utterances, dialogue acts, episodes, topics, ref2doc
Univ. of Fribourg	22 x 15'	A V T D	FR	utterances, ref2doc
AMI Project (ongoing)	~100 hours	A V T D	EN	dialogue acts, addressees, named entities, episodes, summaries, focus of attention*, gestures*

Table 1: Overview of available multimodal meeting data. Annotations marked with * cover only part of the corpus (A: audio, V: video, T: transcript, D: documents, ref2doc: references to documents)

The ICSI Meeting Recorder corpus (Morgan et al., 2003) contains anonymized naturally-occurring meetings, without the video. The first IDIAP corpus (McCowan et al., 2003) contains very short scripted meetings, the first recordings to be done in the IDIAP Smart Meeting Room. Recorded in the same meeting room, the ISSCO meetings

are longer, less-constrained, and scenario-based; some of them were included in the AMI corpus. The press-review meetings recorded at the University of Fribourg (Switzerland), in French, were used among other things to study the references made to documents, since they are document-centric meetings (Popescu-Belis & Lalanne, 2006). Finally, the meeting corpus produced by the European project AMI (Carletta et al., 2006) is still under production, but will constitute a major resource for multimodal processing, since it consists of nearly 100 hours of meetings recorded in state-of-the-art Smart Meeting Rooms, accompanied by a number of annotations, some of which are mentioned in Table 1.

This overview clearly shows the need for tools that allow straightforward browsing of a large amount of data annotations.

3.2. Annotations of Dialogues

A number of content-related annotations of the meetings are necessary to define search criteria and to answer queries based on them. The use of annotation-directed browsing also represents a step towards meeting summarization, with the advantage that all the information remains accessible to the user via the browser.

The shallow dialogue annotation (SDA) model described in detail elsewhere (Popescu-Belis et al., 2005) is a compromise between the informativeness of annotations and the feasibility of their automatic annotation. We believe that the contents of a meeting are mainly conveyed by language, as it is used in dialogues, slides, and documents. Therefore, we grant a major role to the time-aligned transcript of the meeting, which can be

generated by automatic speech recognition with variable accuracy (Stolcke et al., 2005), and to the structured representations of the contents of documents (Lalanne et al., 2005).

The transcript is the basis for the following content-related annotations: segmentation of individual channels into *utterances*, labeling of utterances with dialogue acts (e.g. statement, question, command, or politeness mark), segmentation of the meeting into thematic *episodes*, labeling of episodes with salient keywords, and document/speech alignment using explicit references to documents. The AMI meeting corpus mentioned above considers additional annotations that could be used for browsing (see Table 1), but only part of the AMI corpus will be annotated with all the dimensions studied in the project. The two main units of information that are considered here are thus the utterance and the thematic episode.

For development and evaluation purposes we use here a set of hand-annotated meetings, while keeping in mind that a fully-automated SDA parser would generate imperfect annotations (Popescu-Belis et al., 2005).

4. Scenarios of Use for TQB

TQB was mainly designed to provide access to the transcript of the meetings and their annotations, as well as to the meeting documents, as these aspects are considered to be the main information vector related to meeting interaction. One of the main distinctions regarding meeting browser functionalities is whether the browser has the capacity to manage multiple meetings (possibly with cross-meeting search) or not.

TQB
Transcript-based Query and Browsing Interface
[IM2.MDM](#) - [APB](#) | [MG](#) | [PhB](#)

Search for utterances

Speaker:

Utterance type/function:

Episode/topic:

Document mention:

Word or string:

Time interval (sec.): -

Instructions

- Search for utterances meeting all the criteria above, and browse full transcript and documents.
- When searching, click on the results to locate the utterances in the full transcript (top of frame).
- Click on a time mark in the transcript ([123.456]) to play the audio. Click again on it to stop, or click on another one to jump.
- References to documents appear as hyperlinks.
- Frames can be resized and scrolled as needed.
- Use *Backspace* to move back within a frame.

Search results

IB4010: episodes, keywords and full transcript aligned with audio

1. [agenda](#) [0-119]
2. [presentation, start](#) [119-172]
3. [Agnes, movies](#) [172-373]
4. [Agnes, Usual Suspects](#) [373-466]
5. [ratings](#) [466-547]
6. [Agnes, Sixth Sense](#) [547-661]
7. [Mirek, presentation](#) [661-784]
8. [Mirek, Schindler's List](#) [784-901]
9. [Mirek, Usual Suspects](#) [901-988]
10. [Mirek, Pulp Fiction](#) [988-1095]
11. [Mirek, Goodfellas](#) [1095-1179]
12. [Mirek, Silence of the Lambs](#) [1179-1327]
13. [Mirek, American Beauty](#) [1327-1450]
14. [Mirek, discussion](#) [1450-1592]
15. [Andrei, presentation, Saving Private Ryan](#) [1592-1906]
16. [Andrei, hero, Private Ryan](#) [1906-2044]
17. [Andrei, other candidates](#) [2044-2118]
18. [Denis, presentation, The Big Lebowski](#) [2118-2359]
19. [Denis, posters](#) [2359-2572]
20. [discussion, nominate](#) [2572-2702]
21. [vote, eliminate](#) [2702-2798]
22. [decision, vote](#) [2798-2875]
23. [next meeting](#) [2875-2960]

IB4010: documents

1. [Participants](#)
2. [Agenda](#)
3. [Agnes: presentation](#) [1 2 3 4 5]
4. [Mirek: presentation](#) [1 2 3 4 5 6 7 8 9]
5. [Andrei: presentation](#) [1 2]
6. [Denis: presentation](#) [1 2]
7. [Denis: posters](#) [1 2 3]

IB4010: Participants

Andrei Agnes

Denis Mirek

Episode #1 (keywords: agenda)

Andrei [1_2]: [47.090] HI, [47.424]
Denis [3_2]: [47.549] HI, [48.080]

Latest log at 2006-02-22 12:23:23 server time.

Figure 1: Initial state of the interface after selecting one meeting

In its present configuration, TQB focuses on search and browsing of one meeting at a time, with the initial possibility to select one meeting from the collection based on the list of meeting names. Once the meeting is selected, the interface is configured to reflect the possible values of the query attributes for the specific meetings. The scenario of use for information gathering from one meeting allows the following consultation modes.

4.1. Search and Browse

The user can search for the particular utterances that satisfy a set of constraints. TQB displays in the left-hand menu (see Figure 1) the annotation dimensions that are searchable for the selected meeting, with a menu of possible values for each of them, computed on-the-fly from the database tables. The user can set values for the following parameters related to utterances: speaker, episode and its keywords, dialogue act, documents referred to, time interval, and words (i.e. the utterance must contain a specific string). The results of the query, i.e. the utterances that match the constraints, are displayed in a separate frame (top center-right in Figures 1 and 2).

For instance, in the example shown in Figure 2 (next page), the user searched for all the questions asked by Mirek (one of the participants) which contain the string “Big Lebowski” (the name of a movie discussed during the meeting), regardless of the episode to which they belong or the documents they refer to. The four utterances matching these constraints are displayed (as transcripts) in the top right frame.

Either the user finds directly the information that they were looking for in one of the retrieved utterances, or, if not, they can use any of the utterances as a starting point to browse the meeting from the specific location of that utterance. By clicking on any of the retrieved utterances (the fourth one in the example in Figure 2), the transcript frame is scrolled down to the position of the utterance, and the user can start browsing the meeting, as we explain in the following section.

4.2. Overview and Browse

The user can also directly start using TQB by browsing the multimodal meeting data. The transcript and the meeting documents constitute the two master columns occupying the center of TQB. They both include, initially visible at the top of the frame, a hyperlinked table of contents of the respective frame (Figure 1): the list of episodes with their keywords for the transcript frame, and the list of document pages (e.g. slides) for the document frame. These tables offer a simple overview of the available information, which could be constructed automatically with better performances than a text-based summary. The document frame also includes as an initial “document” a snapshot view of the participants taken from the video (the video itself does not seem informative enough to be included) with the names and color codes of the participants¹. Browsing through these frames is enhanced by access to the audio (clicking on a timestamp starts/ stops/ shifts an audio player) and by document/transcript alignment: clicking on a referring expression scrolls to the respective document.

¹ Some of the recordings are anonymized, and some lack video. The setting we describe is the ideal and most frequent one.

A summary of the solutions proposed to implement these functionalities appears in the next section. The relevance of the TQB designed to actual meeting search and browsing will also be tested experimentally in the near future, through a competitive campaign based on the BET protocol (see section 6).

5. Functionalities of TQB and their Implementation

TQB was designed to be user-friendly to the novice user, and to be easily accessible through a web browser. TQB is a lightweight transcript-driven interface in the sense that: (1) the query functionality is implemented as a relational database of annotations that are mainly done on the transcript; and (2) the enhanced browsing functionalities are based on a transformation of the XML transcript and annotation files using XSLT stylesheets that control the resulting layout and that insert the links between modalities (transcript/audio, transcript/document).

5.1. Database Backend


The XML annotations corresponding to the SDA components (section 3.2) are stored in a database, while the original media files are stored on a standard file server.

The structure of the database reflects the described annotations and the textual transcription of each meeting. A record from the utterances table represents each utterance, by having associated an index number, start and stop timestamps and the utterance transcription. Then a table is associated for each annotation type, e.g. speakers table, episodes table, dialogue acts table, documents table, etc. The annotation tables are then related to the utterance table by the means of either the index field or the timestamp fields. The annotations are input into the PostgreSQL database either through a Java program or by converting them into tabular format using XSLT stylesheets, and loading them using regular SQL import functions.

The consultation of the database is realized through a client-server application, implemented using Java Server Pages accessed through an Apache Tomcat servlet (<http://tomcat.apache.org/>). The form-based interface gives access to most of the fields of the database, with the possible values filled displayed as menus constructed on-the-fly for each meeting. Based on the conjunction of all the query parameters set by the user, the SQL query to the database is dynamically generated and sent to the PostgreSQL server. The utterances returned by the database backend are displayed in another frame, with hyperlinks to their position in the meeting which were constructed on-the-fly based on utterance timing.

5.2. Multimedia Rendering of Recordings

In addition to the annotations, the data stored in various media on the file server needs some processing in order to be rendered as a coherent flow of information, which re-creates the essential content of the original meeting. Otherwise, if the only processing is the time-alignment of the media flows, then the resulting browser is little more than a classic media player.



TQB
Transcript-based Query and Browsing Interface
[IM2.MDM](#) - [AFB](#) | [MG](#) | [PhB](#)

Search for utterances

Speaker:

Utterance type/function:

Episode/topic:

Document mention:

Word or string:

Time interval (sec.): -

Instructions

- Search for utterances meeting all the criteria above, and browse full transcript and documents.
- When searching, click on the results to locate the utterances in the full transcript (top of frame).
- Click on a time mark in the transcript ([1:23.456]) to play the audio. Click again on it to stop, or click on another one to jump.
- References to documents appear as hyperlinks.
- Frames can be resized and scrolled as needed.
- Use *Backspace* to move back within a frame.

Latest log at 2006-02-22 12:23:23 server time.

Search results

(TQB found 4 utterances - displayed in chronological order)

Mirek [2_433]: [2311.28] [Why the movie is called The Big Lebowski? Why The Big -](#) [2314.43]

Mirek [2_438]: [2327.227] [And why The Big Lebowski?](#) [2329.254]

Mirek [2_494]: [2673.28] [May I nominate uh these - all these three movies, Big Lebowski, Schindler's List and the Silence of the Lambs?](#) [2679.088]

Mirek [2_545]: [2863.178] [So it's gonna to be Big Lebowski?](#) [2864.816]

Andrei [1_696]: [2861.808] **So** - [2862.288]

Mirek [2_545]: [2863.178] **So it's gonna to be Big Lebowski?** [2864.816]

Andrei [1_698]: [2863.278] **decision** - [2864.262]

Denis [3_598]: [2863.280] **Okay.** [2864.007]

Andrei [1_700]: [2866.288] **Mm-hmm.** [2866.800]

Mirek [2_547]: [2867.536] **Okay.** [2868.160]

Andrei [1_702]: [2869.136] **That's your fault, Mirek.** [2870.483]

Agnes [0_530]: [2870.557] **\$ If you don't like it it's your problem. \$** [2874.508]

Mirek [2_551]: [2873.479] **No no no no no, I I like it, I like it.** [2875.692]

Andrei [1_705]: [2874.224] **No, no, no, I mean yeah, it's - you voted for it.** [2877.391]

Episode #23 (keywords: next meeting)

Denis [3_600]: [2875.986] **So we have other @ anyway. So can we decide for uh a next meeting maybe?** [2882.240]

Andrei [1_706]: [2877.391] **But I'm quite happy with it, I haven't seen it.** [2879.540]

Agnes [0_532]: [2878.741] **Yeah.** [2879.2]

Andrei [1_708]: [2883.136] **Well, beginning of May I guess, to choose the film for May.** [2886.54]

Agnes [0_534]: [2886.24] **Yeah.** [2886.56]

Denis [3_602]: [2886.496] **Okay.** [2887.068]


Andrei [1_709]: [2886.54] **Is that -** [2887.152]

Mirek [2_553]: [2889.040] **Okay, beginning of May,** [2889.934]

Mirek [2_554]: [2889.024] **but there are there is plenty of**

Denis: posters

Denis: poster 1



CINÉ-CLUB MONTREUX
MONTREUX MOVIE CLUB

Friday 29th April
20h

The BIG Lebowski

Joel Coen

It's the early nineties, sometime around the first Ira Jeffrey Lebowski, known as "The Dude" to his friends, comes home and gets jumped and threatened by two other "The Dudes" who mistake him for a rich and powerful "The Dude" who owes them money. They pee on his rug because they mistake him for a rich and powerful "The Dude" who owes them money. When the Lebowski's meet, their worlds collide and more and more mixed up with one another. While "The Dudes" friends keep trying to help him, all they suc...

Figure 2: View of the interface after one query

As regards TQB, after the user clicks on one of the utterances retrieved from the database, or selects one of the entries of the tables of contents (visible in the central frames of Figure 1), or simply browses the transcript, a number of enhancements are available to improve the informational gain and user-friendliness of browsing. Most of these enhancements are implemented as HTML or Javascript, and constructed automatically for each transcription from the XML annotation files. Access to the audio is based on an instance of the RealPlayer embedded in the HTML transcript file.

Therefore, generating the resource files, as well as accessing them, makes use only of simple XML-based mechanisms. No particular software needs to be downloaded before using TQB: a web functional browser and an HTTP connection to the TQB server are sufficient.

6. Towards Evaluation of TQB

6.1. TQB Participation in BET Evaluations

An experimental comparative evaluation of meeting browsers named BET is under way (Wellner et al., 2005). BET (browser evaluation and test) focuses on the increase in informativeness brought by a meeting browser. Performance of the browsers will be measured by the average speed and accuracy of human subjects who use the browser to answer yes/no questions related to facts from each meeting used for the test.

To avoid the introduction of bias into the list of questions by the creators of the browsers, the questions are produced by independent "observers" who are asked

to create pairs of true/false statements derived from what were, in the observer's opinion, the most salient facts related to the meeting. The task of the subjects is then to distinguish the true from the false statement for each pair.

For instance, the query shown in Figure 2 could have been produced by a subject trying to disambiguate the following pair of observations: (1) The movie finally selected was *The Big Lebowski* vs. (2) The movie finally selected was *Schindler's List*. The transcript visible in Figure 2 and the corresponding audio indicate that the true statement is the first one.

The production and selection of observations is almost completed, at the time of writing, for two meetings from the AMI corpus that were selected for the first full-fledged BET evaluation. The raw list of observations had to be cleaned in order to discard duplicates and to avoid an ordering in which earlier questions would implicitly disclose the answer to later ones. Once the list of observations finalized, each tested subject will first get familiar to a browser on one meeting, then proceed to the actual testing on the second (unseen) meeting, trying to answer correctly the maximum number of questions in a fixed amount of time.

A web-based demo version of TQB will be made available at <http://www.issco.unige.ch/projects/im2/> after the current BET campaign, with a set of meetings that have already been publicly released such as the IDIAP meeting corpus (McCowan et al., 2003).

6.2. Monitoring User Interaction with TQB

To increase the benefit from the BET testing, it is appropriate to record as much as possible of the users'

interaction with BET, and to correlate this data with the accuracy of the subjects' answers to the BET questions. The logging mechanism has two components, which both log their messages to a file on the server that is indexed according to the IP number of the client using TQB. One component logs all the queries to the database with their timestamps, while the other components buffers a number of user actions performed in the various frames, and sends them at regular intervals to the server. These buffers are in fact hidden input fields in the bottom left frame (visible on the figures), which displays a confirmation message. The log files can thus be consulted through a web browser.

Two types of information are logged by the frame action logger: (1) the state and position of the RealPlayer and the position of the scrollbars in the frames at the time when the information is sent to the server (every 30 seconds at present); (2) the operations that are done in the frames (each of the Javascript function calls logs each call to the hidden variables); the following operations are monitored:

- calls to the embedded RealPlayer (start, stop, pause the audio recording);
- clicks on hyperlinks that mark references to documents;
- clicks on hyperlinks from the table of contents of the transcript frame, and the one from the document frame;
- clicks on the utterances obtained as results of a query ('Search results' frame).

The analysis of the logged interactions will indicate the most useful features of TQB and the most consulted media and annotations; in addition, a post-experiment questionnaire will help assessing the usability of TQB.

7. Conclusion

The search and browsing mechanism we propose provides an intuitive way to navigate through meeting data, which should prove its robustness in the upcoming BET evaluation campaign. TQB is a simple, direct and portable solution for accessing multimodal data, which has enough generality to be extended to other types of multimodal annotated recordings as well.

Acknowledgements

The work presented here is part of the Interactive Multimodal Information Management project (IM2, <http://www.im2.ch>), funded by the Swiss National Science Foundation through the NCCR program. We would like to thank Mike Flynn (IDIAP, Martigny) for his help with the design of the logging mechanism for TQB; Jean Carletta (University of Edinburgh) for help with the AMI data and annotations; and Philippe Baudrion (University of Geneva) for server installation and maintenance.

References

- Bernsen N. O. (2002). Multimodality in Language and Speech Systems - From Theory to Design Support Tool. In B. Granström, D. House et I. Karlsson (ed.), *Multimodality in Language and Speech Systems*, Dordrecht, Kluwer Academic Publishers, pp. 93-148.
- Carletta J., Ashby S., Bourban S., Flynn M., Guillemot M., Hain T., Kadlec J., Karaiskos V., Kraaij W., Kronenthal M., Lathoud G., Lincoln M., Lisowska A., McCowan I., Post W., Reidsma D. and Wellner P. (2006). The AMI Meeting Corpus: A Pre-announcement. In S. Renals et S. Bengio (ed.), *Machine Learning for Multimodal Interaction II*, Berlin, Springer-Verlag, pp. 28-39.
- Lalanne D., Ingold R., von Rotz D., Behera A., Mekhaldi D. and Popescu-Belis A. (2005). Using Static Documents as Structured and Thematic Interfaces to Multimedia Meeting Archives. In S. Bengio et H. Bourlard (ed.), *Machine Learning for Multimodal Interaction*, Berlin, Springer-Verlag, pp. 87-100.
- Lisowska A., Popescu-Belis A. and Armstrong S. (2004). User Query Analysis for the Specification and Evaluation of a Dialogue Processing and Retrieval System. *Proceedings of LREC 2004 (4th International Conference on Language Resources and Evaluation)*, Lisbon, Portugal, vol. III/VI, pp. 993-996.
- Lisowska A., Rajman M. and Bui T. H. (2005). ARCHIVUS: A System for Accessing the Content of Recorded Multimodal Meetings. In S. Bengio et H. Bourlard (ed.), *Machine Learning for Multimodal Interaction*, Berlin, Springer-Verlag, pp. 291-304.
- McCowan I., Bengio S., Gatica-Perez D., Lathoud G., Monay F., Moore D., Wellner P. and Bourlard H. (2003). Modeling Human Interaction in Meetings. *Proceedings of ICASSP 2003 (International Conference on Acoustics, Speech, and Signal Processing)*, Hong Kong, China.
- Morgan N., Baron D., Bhagat S., Carvey H., Dhillon R., Edwards J. A., Gelbart D., Janin A., Krupski A., Peskin B., Pfau T., Shriberg E., Stolcke A. and Wooters C. (2003). Meetings about Meetings: Research at ICSI on Speech in Multiparty Conversations. *Proceedings of ICASSP 2003 (International Conference on Acoustics, Speech, and Signal Processing)*, Hong Kong, China.
- Popescu-Belis A., Clark A., Georgescu M., Zufferey S. and Lalanne D. (2005). Shallow Dialogue Processing Using Machine Learning Algorithms (or not). In S. Bengio et H. Bourlard (ed.), *Machine Learning for Multimodal Interaction*, Berlin, Springer-Verlag, pp. 277-290.
- Popescu-Belis A. and Lalanne D. (2006). Detection and Resolution of References to Meeting Documents. In S. Renals et S. Bengio (ed.), *Machine Learning for Multimodal Interaction II*, Berlin, Springer-Verlag, pp. 64-75.
- Stolcke A., Anguera X., Boakye K., Cetin O., Grezl F., Janin A., Mandal A., Peskin B., Wooters C. and Zheng J. (2005). Further Progress in Meeting Recognition: The ICSI-SRI Spring 2005 Speech-to-Text Evaluation System. *Proceedings of NIST MLMI 2005 Meeting Recognition Workshop*, Edinburgh, UK.
- Wellner P., Flynn M. and Guillemot M. (2004). Browsing Recordings of Multy-party Interactions in Ambient Intelligent Environments. *Proceedings of CHI 2004 Workshop on "Lost in Ambient Intelligence"*, Vienna, Austria.
- Wellner P., Flynn M., Tucker S. and Whittaker S. (2005). A Meeting Browser Evaluation Test. *Proceedings of CHI 2005 (Conference on Human Factors in Computing Systems)*, Portland, OR, USA, pp. 2021-2024.