IDIAP RESEARCH REPORT

FINDING INFORMATION IN MULTIMEDIA RECORDS OF MEETINGS

Andrei Popescu-Belis Denis Lalanne
Hervé Bourlard

Idiap-RR-32-2011

AUGUST 2011

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

Finding Information in Multimedia Records of Meetings

Andrei Popescu-Belis
Idiap Research Institute
Rue Marconi 19, PO Box 592
CH-1920 Martigny, Switzerland
phone: +41 27 721 7729 – fax: +41 27 721 7712
andrei.popescu-belis@idiap.ch
http://www.idiap.ch/~apbelis

Denis Lalanne
Department of Computer Science
University of Fribourg
Bd de Pérolles 90
CH-1700 Fribourg, Switzerland
phone: +41 26 300 8472 – fax: +41 26 300 9726
denis.lalanne@unifr.ch
http://diuf.unifr.ch/people/lalanned

Hervé Bourlard
Idiap Research Institute and École Polytechnique Fédérale de Lausanne
Rue Marconi 19, PO Box 592
CH-1920 Martigny, Switzerland
phone: +41 27 721 7720 – fax: +41 27 721 7712
herve.bourlard@idiap.ch
http://www.idiap.ch/~bourlard

April 4, 2011

Abstract

This paper surveys the work carried out within two large consortia, AMI and IM2, on improving access to records of human meetings thanks to multimodal interfaces called meeting browsers. These tools help users navigate through multimedia records containing audio, video, documents and metadata, in order to obtain a general idea about what happened in a meeting or to find specific pieces of information. To explain the increasing importance of meeting browsers, the paper summarizes findings of user studies, discusses features of prototypes from AMI and IM2, and outlines a proposed evaluation protocol, providing reference scores for benchmarking. These achievements result from an iterative software process, alternating user studies, prototypes or products, and evaluation.

Keywords: meeting support technology, meeting browsers, user requirements, evaluation.

1 Introduction

The design of technology for recording, processing, and browsing human meetings has become a significant research field in the past two decades, as evidenced by recent surveys [13, 3, 18]. Meeting support technology draws on advances in multimodal signal processing, verbal and non-verbal communication analysis, as well as multimedia information retrieval and human-computer interaction. The growing interest in the field is driven by the ever larger number of meetings held worldwide, and the availability of new, realistic data sets. However, the field has often put applications before methodology, and thus the definition of common tasks and benchmark data has lagged behind the development of individual systems.

The purpose of this article is to put into a coherent perspective the achievements and the lessons learned about meeting browsing from the experience of two long-term, multi-disciplinary consortia: the IM2 Swiss National Center of Competence in Research (Interactive Multimodal Information Management, 2002–2013), and the European AMI Consortium (Augmented Multiparty Interaction, 2004–2010), both headed by the Idiap Research Institute (with the University of Edinburgh for AMI). These consortia made significant advances in multimodal signal processing applied to multiparty meetings, generating large databases of annotated data recorded in controlled settings, such as the AMI Corpus.

This article will show how and why *assistance in fact-finding*, supported by specific *meeting browsers*, has become a central task for meeting analysis and retrieval. Though not the only possible exploitation of meeting support technology, fact-finding in meeting recordings has emerged gradually as a relevant task, following a series of back-and-forth exchanges between users and developers. Raw audio-visual recordings of meetings are indeed of little use without tools that offer more structured methods for accessing their content than simple media players do. Meeting browsers, as thought of since the 1990s, assist humans with navigation through records that include audio, video, documents, and metadata. For instance, they can help finding the exact value of a budget figure mentioned in a meeting, checking tasks and deadlines assigned to participants, or determining whether a given topic was discussed or not.

The following section analyzes user studies from within and outside our consortia, explaining why they could not lead directly to software specifications, but had to be gradually focused towards the most promising task. Then, a range of meeting and conference browsers from our consortia are discussed. The proposed evaluation method is then put into perspective, along with speed and precision reference scores for future benchmarking. A reflection on the meeting browser development process concludes the article.

2 User requirements

Published studies of user needs for meeting support technology are comparatively less numerous than analyses of specific tools, despite the fact that capturing user needs normally initiates the software development process. Our two consortia considered as starting points some of the material previously published on these matters outside AMI/IM2

[6, 1], then proceeded to elicit additional requirements discussed below. Essentially two strategies have been used: (1) analyze the use of current technology for meeting support and infer unsatisfied needs that new technology could fulfill; or (2) ask users to describe functionalities that would likely support their involvement in meetings better than existing ones do.

2.1 Analyzing the use of existing technology

Two ethnographic studies are representative of the first strategy [6, 17] (see Table 1, first two studies). Both studies explored the types of records and cues that people use to recall information from past meetings, and were carried out in a corporate context. They considered series of project-related meetings, through interviews with a dozen people each, over several weeks or months. The first study [6] additionally surveyed 500 people, with different questions. In the first study, the users found audio-visual records useful for verifying or better understanding points in a meeting, and as an accurate overall record. In the second one [17], they emphasized the limitations of official minutes for recalling specific details, which could be overcome by private notes. In both studies, searching verbatim meeting records was considered to be a potentially challenging task, which could be facilitated by structured minutes with assigned tasks and decisions.

Two other ethnographic studies [4, 2], with respectively 10 and 100 users, confirmed and extended these insights. In order to retrieve information about a past meeting they attended, people use mainly minutes and personal notes, though they often rely on personal recollection or emailed information only. The utility of audio-visual recordings alone is quite low, because watching the recording of an entire meeting is time consuming. Recordings are still viewed as useful to check what someone has said, or as a proxy for people who missed a meeting, or to remember past topics, assigned tasks, or the date of the next meeting.

2.2 Eliciting new requirements from users

Other user studies have asked participants to imagine an intelligent search and navigation tool, and to describe the tasks that it could perform, and the queries that they would address to it. In one study [4], users suggested to include arguments for decisions in automatically-generated meeting minutes, along with lists of main topics and things to do, but also simply the agenda and the names of the participants. In another set of desiderata collected from professionals by a non-AMI/IM2 survey [1], the most frequent wish was for the list of topics discussed at a meeting.

Several sets of explicit queries were collected from developers of meeting technology as well as non-technical users [8, 15]. In one study [8], 28 participants could choose between several use cases – a manager tracking employee performance or project progress, an employee missing one project meeting or joining an ongoing project – and were asked to formulate queries to access a meeting archive, resulting in about 300 queries. Users appeared to be interested in two main types of items:

1. Items related to the interaction between participants, such as decisions, questions, discussions, or disagreement.

Study	Subjects	Method	Focus	Summary of findings
Jaimes [6]	15	interviews	practice	Importance of audio-visual records for checking or better understanding specific points in a meeting.
	519	questionnaires	practice	Importance of visual cues for recall.
Whittaker [17]	12	interviews	practice	Importance of personal notes, need for to-do summaries.
Cremers [4]	8	interviews	practice / needs	Need for summaries and to-do lists.
Bertini [2]	118	questionnaires	practice / needs	Low utility of audio-visual records, except for persons who missed a meeting or for finding specific information.
Banerjee [1]	12	interviews	practice	Importance of topic lists.
Lisowska [8]	28	elicitation of queries	needs	Heterogeneity of queries, either about the interaction or about simple items in meetings.
Wellner [15]	21	elicitation of observations of interest	needs	Importance of facts, decisions, arguments leading to them, agenda, and dates.
Lisowska [7]	91	Wizard-of-Oz	needs	Importance of training for modality choice in meeting browsing.

Table 1: Comparison of user studies for meeting browsing technology. All studies except [1, 6] were conducted in relation to the AMI or IM2 consortia.

2. Items that are conceptually related to meetings, such as dates, participants, documents, presentations, plus global and local discussion topics.

The capacity to answer some of the queries requires complex processing such as topic detection or understanding of the interaction structure, but many other queries are in fact directed towards elementary information only.

From a different perspective, a large-scale Wizard-of-Oz study [7] with 91 subjects using a partially-implemented interface has found that exposure and training had a strong impact on the choice of modalities used to access a meeting archive – either speech, written language, or mouse clicks – with no natural combination standing out. Speech was slightly preferred for interaction over other modalities, as the system appeared to recognize it accurately, thanks to a dedicated human wizard who was hidden from the users.

2.3 Requirements inferred from BET statements

The BET procedure (Browser Evaluation Test) for collecting queries and using them in evaluation [15, 9] was proposed within AMI/IM2, and is discussed later in this paper. However, the BET 'observations of interest' collected over a meeting can also be analyzed to infer user requirements for browsers. In the BET collection procedure, neutral human observers are asked to formulate pairs of "parallel" statements about a meeting, of which

one is factually true and the other one is false. Observers proceed through the following stages:

1. View a meeting recording using a simple media player.
2. Write down observations of interest about it, defined as statements describing the most salient facts *for the meeting participants*.
3. Indicate whether each observation has a local or global scope.
4. Create for each statement a similar, plausible, but false counterpart.

Three meetings from the AMI Corpus were submitted to observers, resulting in 572 pairs of statements from 21 observers. Statements with the same meaning were consolidated by the experimenters into groups, resulting in 350 pairs of true/false statements, with importance scores. Consolidated groups contain on average two statements, but when considering only the statements effectively used for evaluations, each statement was mentioned on average by five observers, which demonstrates some agreement on the most important observations. Examples among the most frequently mentioned BET pairs are (with highlighted differences):

- "The group decided to show *The Big Lebowski*" vs. "...to show *Saving Private Ryan*".
- "According to the manufacturers, the casing has to be made out of *wood*" vs. "...made out of *rubber*".
- "*Susan* says halogen light is very bad for reading in" vs. "*Agnes* says halogen ...".

There are many more statements with a local scope rather than a global one: in the non-consolidated set, 63% of the statements referred to specific moments in a meeting, 30% to short intervals, and only 7% were about the entire meeting. However, this proportion might have been biased by the simple media player used in the collection procedure. Content-wise, statements fall into five categories:

1. Decisions (8%).
2. Other facts stated by participants, including arguments leading to decisions (76%).
3. Statements related to the interaction process or to the media used by participants (11%).
4. Statements about the agenda (2%).
5. Statements about the date of the following meeting (2%).

The last two categories, although infrequent, were mentioned at least once by each observer. If the count is done on the consolidated subset of statements mentioned by at least three observers each, with 251 statements, then the proportions of statements regarding decisions, agenda and dates increase to 13%, 4% and 3% respectively, while the others decrease slightly. This shows the importance to all observers of decisions, agenda and date of next meeting, and its reflection on the resulting BET resource.

2.4 Synthesis of user studies

The user studies cited above and summarized in Table 1 show that requirements for meeting archival and browsing technology are multi-faceted. Their main dimensions, however, are now better understood than they were ten years before. Requirements can be categorized in terms of: (1) the targeted time span within a meeting or series of meetings, i.e. utterance, fragment, or entire meeting; (2) the targeted media, such as audio, video, documents, presentations, emails; (3) the complexity of the information that is searched for, either present in the media or inferred from content; and (4) query complexity or modality. Still, as user studies are often difficult to generalize, more publicly-available studies are welcome, especially as the underlying technologies evolve constantly.

3 Meeting browsers

Two main types of applications partially answer the above requirements. On the one hand, meeting summarization systems offer an abstracted view of a meeting, structured for instance around its main topics – as in the early 'Meeting Browser' from CMU/ISL [14] – or around the tasks or 'action items' that were assigned – as in the CALO browser [12]. On the other hand, other meeting browsers are intended to help users with fact-finding or verification – e.g. to check figures, decisions, assigned tasks, or document fragments – although they can also be used to sample a meeting for abstractive purposes. Recent surveys [13, 3, 18] include examples of both types, which can also be classified according to the main rendered modality [13] or the complexity of their functionalities [18]. Meeting browsers take advantage of tools that record and analyze meeting data in order to build high-level indexes based on a variety of features, such as speech transcript, turn taking, focus of attention, slide changes, or handwritten notes. These indexes are used within multimodal user interfaces, helping users to locate the information that is likely to fulfill their needs.

Meeting browsers intended for fact-finding and verification (henceforth referred to simply as meeting browsers) have been the main focus of the AMI and IM2 consortia, as they strike a good balance between several divergent targets. First, they are part of promising transversal, complete applications – from media capture, through automatic analysis, to human access to multimodal meeting data. Meeting browsers answer some of the frequently-mentioned user needs for meeting support technology, and are within reach of currently-available technology. Moreover, they have sufficient generality to be of interest to the field of multimodal processing research, raising theoretical questions about the automatic analysis of human-human and human-computer interaction, as studied also by other consortia such as CALO or CHIL. The meeting browsers developed within AMI and IM2 are best classified according to the modality that is used for locating and rendering excerpts from a meeting:

1. *Speech-centric browsers* take advantage of the audio recordings and/or their transcript, often with synchronized video, possibly accompanied by higher-level annotations such as named entities, topics, or extracted keywords. A gallery is shown in Figure 1, illustrating the range of media, components, and layout.

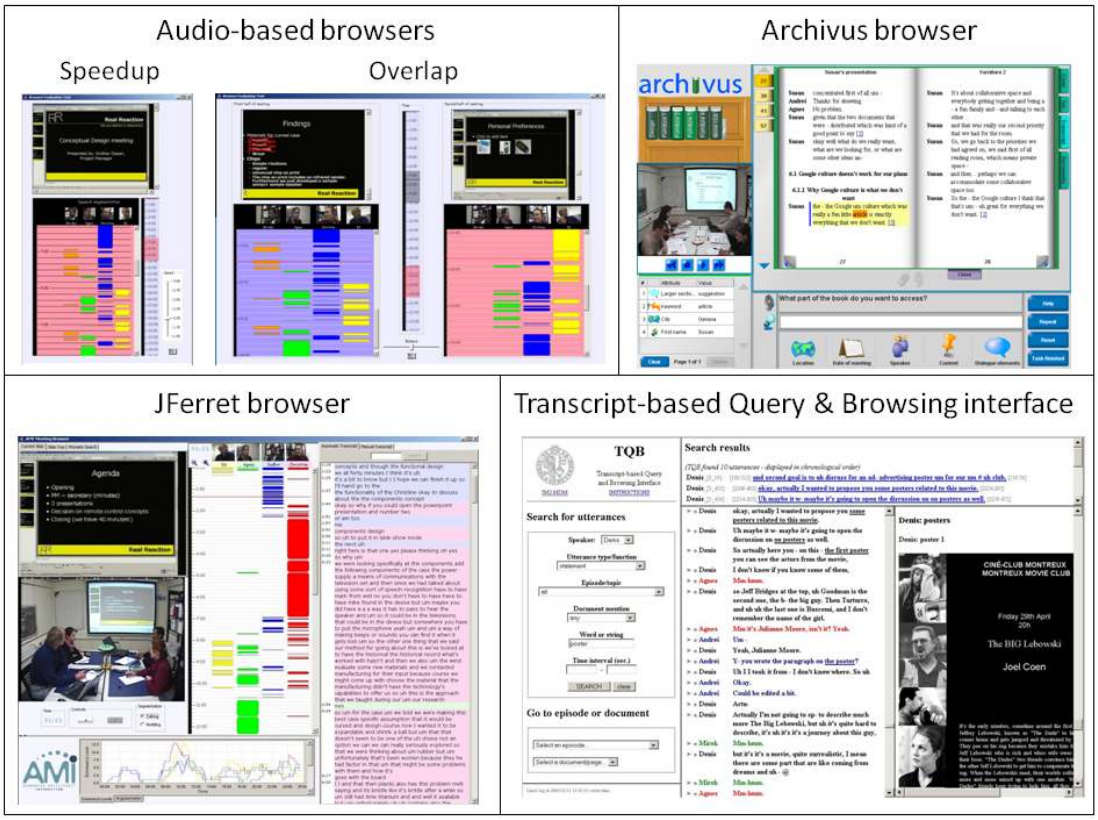

Figure 1: Five speech-centric meeting browsers from the AMI/IM2 consortia, illustrating the diversity of media and layouts. Components include audio, video, and slide players, along with speaker identification and segmentation, transcript, and various query parameters (in Archivus and TQB).

2. *Document-centric browsers* take advantage of document content and/or annotations such as slide changes, sometimes with speech/document alignment. Following this approach, two AMI/IM2-related systems were turned into commercial *conference browsers*. A gallery is shown in Figure 2.

3.1 Speech-centric browsers

Two audio-based browsers [9] provide access to audio recordings through speaker segmentation and slides, while enhancing browsing in two ways. The *Speedup browser* accelerates audio playback, while the *Overlap browser* plays two different parts of a meeting in the left vs. right channels, with the possibility to adjust manually the audio balance in order to focus on the channel that is currently most relevant.

The *JFerret browser* [15] illustrates the main capabilities of the Java-based JFerret framework for browser design, by providing access to audio, video, slides, ASR transcript (or manual one for evaluation purposes), speaker segmentation, and potentially other annotations such as dominance levels. The time-dependent components are synchronized

to a main timeline displayed with the speaker turns.

The *Transcript-based Query and Browsing interface (TQB)* [9] includes several manual annotations in order to assess their respective utility: manual transcript, dialogue acts, topic episodes and labels, and references from speech to documents.

Archivus [7] enables multimodal human-computer dialogue in a Wizard-of-Oz setting, which allows for partial implementation only, in order to gather additional user requirements. Archivus uses reference transcripts enriched with manual annotations such as speaker segmentation, topic labels, and documents, to answer queries that users express naturally with various modalities and that the system processes as sets of attribute-value constraints over one or several meetings.

3.2 Document-centric and conference browsers

JFriDoc [11] is a document-centric browser that provides time-aligned access to the documents discussed during a meeting and to the speech transcripts, with slides and audio-video streams synchronized to a timeline. The *FaericWorld system* [11] extends this approach to collections of meetings, for which similarity links between all the categories of multimedia documents are automatically calculated. Users can then query the system with full text search or directly browse through the document links, thanks to an interactive visualization.

Despite the number of research prototypes, there are still no commercially available meeting browsers intended for end-users. This is all the more surprising since several systems for holding remote meetings are commercially available, some of which even offer recording capabilities. Still, the browsers developed in AMI/IM2 have evolved towards two end-user products, but for a slightly different task, namely conference recording and browsing. The tools intended for conferences use fewer capture devices, with off-the-shelf technology, resulting in comparatively smaller amounts of data to store and process, which might explain why they were quicker to reach product stage. The two systems also answer a growing need for conference recording in flexible settings, with playback using cross-platform, user-friendly interfaces, as initiated for instance in the Classroom 2000 educational environment.

One system is commercialized through an Idiap spin-off company named Klewel (http://www.klewel.com), while the other one was developed by the University of Fribourg and the CERN in Geneva within the SMAC project (Smart Multimedia Archive for Conferences, http://smac.hefr.ch/, see [5]) and is in use at these institutions. Both systems extract a number of robust indexes, such as slide changes, text from slides, and slide/audio/video synchronization, which are helpful for browsing, and provide some support for fact-finding. The SMAC system, in addition, is able to automatically hyperlink the fragments of the scientific article that is being presented to the related audio-video sequence. Such technologies derived from research in our consortia offer these browsers an advantage over other competing systems [5].

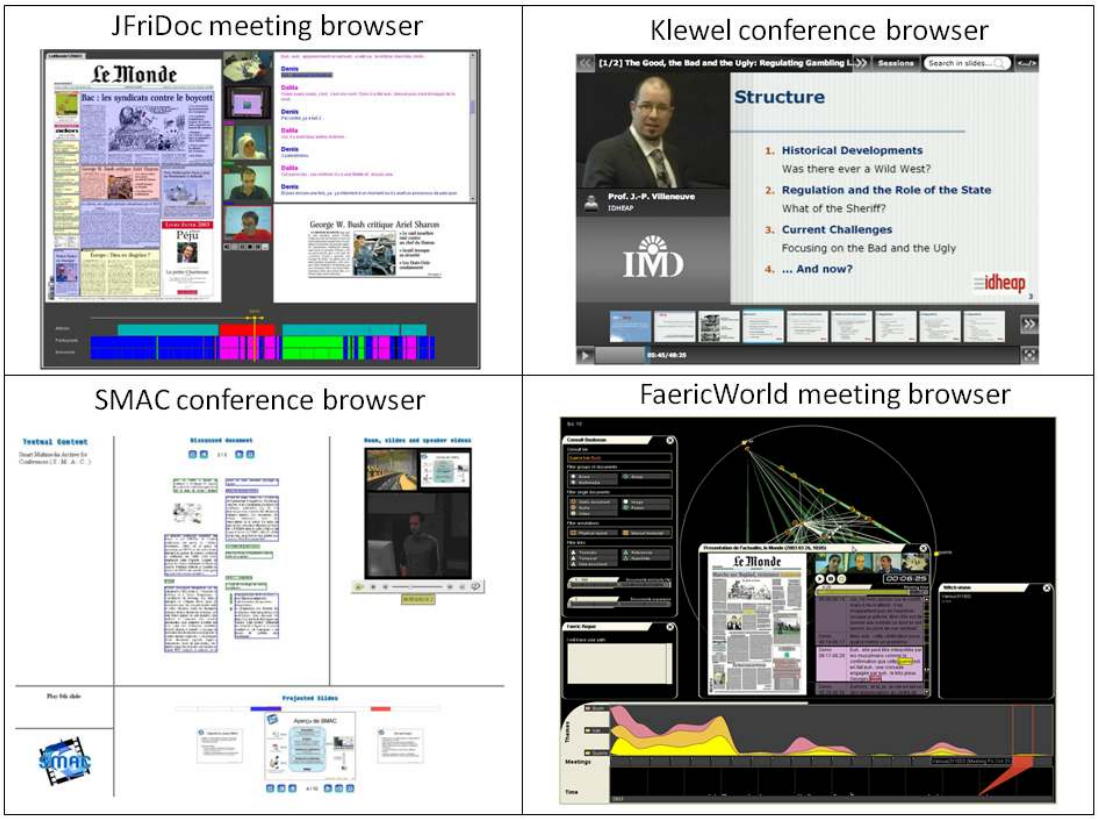

Figure 2: Document-centric meeting browsers and conference browsers from the AMI/IM2 consortia described in the text. Document/speech alignment is central to all layouts.

4 Evaluation methods and results

4.1 Current approaches

By many accounts [3, 18], the evaluation of meeting support technology is a challenging task, though it is unavoidable to demonstrate appropriateness of design, or to compare several designs, interaction paradigms, or meeting analysis components. As synthesized by Yu and Nakamura [18, p.11–12], "the criteria used to evaluate a smart meeting system include user acceptance, accuracy [of recognition mechanisms], and efficiency [...i.e.] whether a browser is useful for understanding the meeting content quickly and correctly." While accuracy of recognition is not by itself a measure of browser quality (though it influences it), the two other criteria reflect two different views of evaluation.

Several studies of individual meeting browsers have considered both approaches to evaluation. The Filochat system of the early 1990s [16] was one of the first browsers for speech recordings, which were time-aligned with personal notes. A user study demonstrated the usability of the system and helped to assess desirable and undesirable features. Laboratory tests compared three conditions (notes only, speech only, or Filochat) by measuring accuracy and speed of subjects who answered factual questions about

what they had heard. The mutual influences of processing accuracy and user behavior were studied for the CALO action item browser [12]. Finally, the AMI/IM2 JFerret meeting browser, augmented with automatically generated abstracts, was evaluated in a large experiment with 27 teams of four people holding series of meetings [10]. The results showed that JFerret outperformed two other browsers, or no browser at all as a control condition, in terms of impact on several parameters characterizing participant satisfaction and meeting success, e.g., finding an acceptable solution to a given design task.

The needs for comparing meeting browsers, at the same moment or over time, are better satisfied by efficiency-oriented evaluations rather than user studies, as they provide a more controlled environment and a standardized protocol. Efficiency can be measured over benchmark tasks that are representative of the meeting browsing activity.

4.2 The Browser Evaluation Test (BET)

The Browser Evaluation Test (BET) was introduced earlier in this paper as a procedure for collecting observations of interest about a meeting, further transformed into pairs of true/false statements [15, 9] that were analyzed to infer requirements. These statements are in fact mostly intended for evaluation, with the following protocol. Subjects who did not act as observers are asked to examine pairs of BET statements using the browser under evaluation, and to find out for each pair which is the true statement and which is the false one. Browser performance is quantified using *precision*, i.e. the number of correct answers, and *speed*, i.e. the number of pairs of statements processed per unit of time. Precision indicates *effectiveness* while speed, when averaged over comparable groups, indicates browser *efficiency*. Of course, behavior analysis, satisfaction questionnaires, and other observational techniques can be applied as well.

The acceptance of the BET as a valid test protocol must also acknowledge a number of possible biases or limitations. First, as any other evaluation method, the BET checks to what extent browsers conform to certain user requirements, but the BET elicitation method biases these requirements towards fact-finding or verification – at least by comparison to other elicitation studies, which have emphasized higher-level elements of interest such as action items, topics or decisions. While these are possibly underrepresented in the current BET set, a different set could also be elicited with an inverse bias.

Unlike many other user-oriented evaluations, the BET observers and subjects *were not* chosen among the participants to the meetings, although the observers were encouraged to make observations that would have been of interest to participants. Therefore, the BET requirements and evaluation task are targeting "null-context users", and cannot be used for comparative evaluation of browsers that offer subjective memorization devices, such as personal notes taken during a meeting, although comparisons involving third-party notes can be made. The somewhat focused spectrum of the BET is the price to pay in order to ensure reproducibility of the method, enabling comparison across browsers at different moments in time.

Still, even in such a constrained setting, comparison across BET scores must always be taken with a grain of salt, given that the precision baseline is not always 50%, time can be constrained in several ways, and the subjects' competencies and training may vary

across groups. Even within a single experiment, the variability of human performance tends to decrease statistical significance. Subjects may have different strategies, some favoring precision over speed, or vice-versa – it appeared that variability was higher for speed than for precision. The amount of manual preparation of the browsers before an experiment must be considered as well. Thus, formal comparisons are acceptable only if the same questions are used, in the same order, on comparable groups of subjects, trained in similar conditions, and having the same amount of time at their disposal. Such strict conditions for formal comparison are rarely verified, except in evaluation campaigns, which have yet to be organized for meeting browsers.

4.3 BET results and lessons learned

More than 100 subjects have evaluated several AMI/IM2 browsers using the BET, in several separate experiments. The results are summarized in Table 2 in terms of precision and speed, with 95% confidence intervals. All the experiments used the three meetings for which BET questions were produced, though the order of presentation, the definition of conditions, and other details of the experimental protocol varied across experiments. Several new questions were introduced for JFerret and Archivus. Therefore, given all the difficulties of making rigorous comparisons, the goal of this synthesis is not to point at "the best browser", but to provide an overview of the state-of-the-art in meeting browsing for fact-finding, with a range of benchmark scores obtained using a reproducible protocol.

The average discrimination time for a BET pair of statements is around 2 minutes, with a 1.5–4 minute range. Precision – generally against a 50% baseline except for open-answer conditions (JFerret and Archivus) – is in the 70–80% range, with higher values for browsers that make use of more human-processed information, i.e. TQB and Archivus. More knowledge appears thus to be helpful to increase precision, but this often means that subjects will also spend slightly more time as they manipulate more complex information. The experiments confirm that the BET questions and protocol are reliable indicators of performance, as the variance of the average answers is small enough to observe significant differences between conditions within experiments, and compares favorably to the variance observed in interactive QA evaluation experiments at TREC or iCLEF.

The main lessons learned from the BET evaluations, apart from the reliability of the BET procedure, concern the AMI/IM2 technologies that appear to be useful for meeting browsing. Transcripts are used intensively when they are of high quality, especially as users tend to perform keyword searches on them, thus pointing to the need for improved speech-to-text systems. However, annotations of the transcript such as named entities or dialogue acts seem much less helpful, at least for direct use in queries.

The documents related to a meeting are also relevant to fact-finding, if made available in the browser, especially when shown along the meeting's timeline, e.g. using automatic slide change detection and speech/document alignment. Slides can even compensate partly for the lack of transcript, as seen for audio-only browsers, which score only slightly below transcript-based ones. The video recordings were the least helpful media for fact-finding in our experiments. Personal notes were seldom available in the meetings used for testing, and were not of interest to the subjects, likely because they were not their own.

Browser	Condition	Subjects	T(s)	CI	P	CI
Audio-based	Speedup	12	99	26	0.78	0.06
browsers [9]	Overlap	15	88	23	0.73	0.08
JFerret [15]	BET set (pilot)	10	100	*43*	0.68	*0.22*
[17, p. 210]	5 gisting questions	5	<180	0	0.45	*0.34*
	5 factual questions	5	<180	0	0.76	*0.25*
TQB [9]	1st meeting	28	228	129	0.80	0.09
	2nd meeting	28	92	16	0.85	0.06
	Both meetings	28	160	66	0.82	0.06
FriDoc [11]	With speech / document links	8	113	n/a	0.76	n/a
	Without links	8	136	n/a	0.66	n/a
Archivus	T/F questions	80	127	*36*	0.87	*0.12*
[7, Ch. 6.6]	Open questions	80	n/a	n/a	0.65	*0.22*

Table 2: Comparative results of several meeting browsers evaluated in similar conditions using the BET. T is the average time in seconds needed by subjects to answer a question, and P is the average precision. Confidence intervals (CI) at 95% are absolute values; when they could not be found, standard deviations are given instead (in *italics*).

Finally, learning effects appeared to be important: even a single training session improved the subjects' performance significantly, and conditioned their preference for modalities used for browsing. While this is good news for product developers, it also introduces an additional variable that must be controlled in evaluation experiments.

5 Discussion: a software process iterating between users and developers

We have presented some of the main achievements made by the AMI and IM2 consortia, over eight years, regarding the requirements and the design of meeting browsers. The resulting picture of the software development process departs considerably from the waterfall model, according to which users have the primary role of formulating requirements for a task, and developers then attempt to design software satisfying these requirements.

In the case of meeting browsing, user requirements did not lead directly to the specification of implementable systems, chiefly because the users' needs were quite underspecified, or were beyond the reach of current technology. Moreover, researchers – who were in most cases the designers of the systems – believed it was important to include additional functionalities that seemed potentially useful and which users might have overlooked. Therefore, specifications and prototypes emerged gradually from a series of exchanges between users and developers. As in many iterative processes, evaluation results from one iteration guided to some extent the specifications for the next one.

The research and development of meeting browsers has gone through four main iterations of the software process, though not always in a strict time sequence. In the first iteration, several studies elicited user requirements and explored current uses of technology, in order to identify needs and prioritize them. In a second iteration, a specific but

not fully implemented prototype was studied in Wizard-of-Oz experiments accompanied by performance measures and analysis of user behavior. In a third iteration, several standalone prototypes were implemented, and could be compared thanks to a common evaluation framework, the BET. Finally, two end-user products for indexing and browsing conference recordings integrated significant know-how from the consortia, and are currently subject to field studies and assessment of customer satisfaction.

The analyses presented here show that, on the one hand, user requirements for meeting browsing cannot constitute a rigid, set-in-stone specification, but depend greatly on how subjects are prompted to respond, and must be gradually focused towards a specifiable and implementable task. On the other hand, trusting only technology providers to measure the usefulness of their technology was unrealistic, leading to never-ending debates in which each provider tried to prove the utility of their own approach.

During the eight years of existence of the AMI and IM2 consortia, with literally hundreds of researchers collaborating together, jumping back-and-forth from the users' to the developers' perspective has enabled us to gradually focus on the fact-finding task, which provided an applicative framework to develop innovative technologies as well as a reliable benchmark to evaluate their usefulness in a user-oriented setting. This paper has surveyed the main contributions made throughout this process, and has presented synthetically a large array of findings, with the goal of making this experience available for future undertakings in meeting support technology.

Acknowledgments

This work has been supported by the Swiss IM2 NCCR in Interactive Multimodal Information Management (SNSF, http://www.im2.ch), and by the AMI and AMIDA EU Integrated Projects (http://www.amiproject.org). The authors are grateful to their colleagues from these consortia for the long-standing and fruitful cooperation. The authors would like to thank the anonymous reviewers and the editor of IEEE Multimedia for insightful comments and suggestions.

References

[1] Satanjeev Banerjee, Carolyn Rose, and Alexander I. Rudnicky. The necessity of a meeting recording and playback system, and the benefit of topic-level annotations to meeting browsing. In *Proceedings of INTERACT 2005 (10th IFIP TC13 International Conference on Human-Computer Interaction)*, LNCS 3585, pages 643–656, Rome, 2005.

[2] Enrico Bertini and Denis Lalanne. Total Recall survey. Technical report, University of Fribourg, Department of Computer Science, August 2007.

[3] Matt-M. Bouamrane and Saturnino Luz. Meeting browsing: State-of-the-art review. *Multimedia Systems*, 12:439–457, 2007.

[4] Anita Cremers, Inge Kuijper, Peter Groenewegen, and Wilfried Post. The Project Browser: Supporting information access for a project team. In *Proceedings of HCII 2007 (12th International Conference on Human-Computer Interaction), Human Computer-Interaction, Part IV*, LNCS 4553, pages 571–580, Beijing, 2007.

[5] Jeremy Herr, Robert Lougheed, and Homer A. Neal. Lecture archiving on a larger scale at the University of Michigan and CERN. *Journal of Physics: Conference Series*, 219:082003, 2010.

[6] Alejandro Jaimes, Kengo Omura, Takeshi Nagamine, and Kazutaka Hirata. Memory cues for meeting video retrieval. In *Proceedings of CARPE 2004 (1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences)*, pages 74–85, New York, NY, 2004.

[7] Agnes Lisowska, Mireille Bétrancourt, Susan Armstrong, and Martin Rajman. Minimizing modality bias when exploring input preference for multimodal systems in new domains: the Archivus case study. In *Proceedings of CHI 2007 (ACM SIGHI Conference on Human Factors in Computing Systems)*, pages 1805–1810, San José, CA, 2007.

[8] Agnes Lisowska, Andrei Popescu-Belis, and Susan Armstrong. User query analysis for the specification and evaluation of a dialogue processing and retrieval system. In *Proceedings of LREC 2004 (4th International Conference on Language Resources and Evaluation)*, pages 993–996, Lisbon, 2004.

[9] Andrei Popescu-Belis, Philippe Baudrion, Mike Flynn, and Pierre Wellner. Towards an objective test for meeting browsers: the BET4TQB pilot experiment. In *Proceedings of MLMI 2007 (4th Workshop on Machine Learning for Multimodal Interaction)*, LNCS 4892, pages 108–119, Brno, 2008.

[10] Wilfried Post, Erwin Elling, Anita Cremers, and Wessel Kraaij. Experimental comparison of multimodal meeting browsers. In *Proceedings of HCII 2007 (12th International Conference on Human-Computer Interaction), Human Interface, Part II*, LNCS 4558, pages 118–127, Beijing, 2007.

[11] Maurizio Rigamonti, Denis Lalanne, and Rolf Ingold. FaericWorld: Browsing multimedia events through static documents and links. In *Proceedings of Interact 2007 (11th IFIP TC13 International Conference on Human-Computer Interaction), Part I*, LNCS 4662, pages 102–115, Rio de Janeiro, 2007.

[12] Gokhan Tr et al. The CALO Meeting Assistant system. *IEEE Transactions on Audio, Speech and Language Processing*, 18(6):1601–1611, 2010.

[13] Simon Tucker and Steve Whittaker. Accessing multimodal meeting data: Systems, problems and possibilities. In Samy Bengio and Hervé Bourlard, editors, *Machine Learning for Multimodal Interaction*, LNCS 3361, pages 1–11. Springer-Verlag, Berlin/Heidelberg, 2005.

[14] Alex Waibel, Michael Bett, Michael Finke, and Rainer Stiefelhagen. Meeting browser: Tracking and summarizing meetings. In Denise E. M. Penrose, editor, *Proceedings of the Broadcast News Transcription and Understanding Workshop*, pages 281–286, Lansdowne, VA, 1998.

[15] Pierre Wellner, Mike Flynn, Simon Tucker, and Steve Whittaker. A meeting browser evaluation test. In *Proceedings of CHI 2005 (ACM SIGCHI Conference on Human Factors in Computing Systems)*, pages 2021–2024, Portland, OR, 2005.

[16] Steve Whittaker, Patrick Hyland, and Myrtle Wiley. Filochat: Handwritten notes provide access to recorded conversations. In *Proceedings of CHI 1994 (ACM SIGCHI Conference on Human Factors in Computing Systems)*, pages 271–277, Boston, MA, 1994.

[17] Steve Whittaker, Simon Tucker, Kumutha Swampillai, and Rachel Laban. Design and evaluation of systems to support interaction capture and retrieval. *Personal and Ubiquitous Computing*, 12(3):197–221, 2008.

[18] Zhiwen Yu and Yuichi Nakamura. Smart meeting systems: A survey of state-of-the-art and open issues. *ACM Computing Surveys*, 42(2):8:1–16, 2010.

About the authors

Andrei Popescu-Belis is a Senior Researcher at the Idiap Research Institute. His main interests are in human language technology, multimodal data analysis, and design and evaluation of linguistic and interactive systems. He is a graduate of the École Polytechnique, Paris, and holds a PhD in computer science from LIMSI-CNRS, University of Paris XI.

Denis Lalanne is a Senior Lecturer and Researcher in the Department of Informatics of the University of Fribourg. His main interests are in human-computer interaction, usability, vizualization, and gesture analysis. He graduated from the Institut National Polytechnique, Grenoble, and received his PhD in computer science from the Swiss Federal Institute of Technology in Lausanne (EPFL).

Hervé Bourlard is the Director of the Idiap Research Institute and a Professor at EPFL. His main interests are in signal processing and statistical pattern classification, with applications to speech and language modeling, speech and speaker recognition, and multimodal processing. He received the Engineering degree and the PhD degree in applied sciences from the Faculté Polytechnique de Mons, Belgium. He is an IEEE Fellow for "contributions in the fields of statistical speech recognition and neural networks" and a Senior Member of the ACM.