Finding Information in Multimedia Meeting Records

Andrei Popescu-Belis
Idiap Research Institute, Switzerland

Denis Lalanne
University of Fribourg, Switzerland

Hervé Bourlard
Idiap Research Institute and École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Meeting browsers help users navigate multimedia records containing audio, video, documents, and metadata to determine what happened in a meeting or to find specific pieces of information.

In the past two decades, the design of technology for recording, processing, and browsing meetings has become a significant research field.[1–3] Meeting-support technology draws on advances in multimodal signal processing, verbal and nonverbal communication analysis, multimedia information retrieval, and human-computer interaction. The growing interest in the field is driven by the increasing number of meetings held worldwide and the availability of new, realistic datasets. However, the field has often put applications before methodology, and thus the definition of common tasks and benchmark data has lagged behind individual system development.

This article shows how and why fact-finding assistance, supported by specific meeting browsers, has become a central task for meeting analysis and retrieval. Raw audio-visual meeting recordings are of little use without tools that offer more structured methods for accessing their content than simple media players. Introduced in the 1990s, meeting browser prototypes help users navigate records that include audio, video, documents, and metadata. For instance, they can help users find the exact value of a budget figure mentioned in a meeting, check assigned tasks and deadlines, or determine whether a given topic was discussed. Although 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.

Specifically, we survey the achievements and lessons learned about meeting browsing from the experience of two long-term, multidisciplinary consortia: the Interactive Multimodal Information Management (IM2) Swiss National Center of Competence in Research (2002–2013) and the Augmented Multiparty Interaction (AMI) European Consortium (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,[4] generating large databases of annotated data recorded in controlled settings, such as the AMI Corpus.

User Requirements

There are fewer published studies of user needs for meeting-support technology than analyses of specific tools, despite the fact that capturing user needs normally initiates the software development process. As starting points for the IM2 and AMI Consortia, we considered some of the material previously published on these topics.[5,6] We then proceeded to elicit additional requirements. Essentially, we used two strategies. We either analyzed the use of current technology for meeting support and inferred unsatisfied needs that new technology could fulfill, or we asked users to describe new functionalities that would likely better support their involvement in meetings.

Analyzing Existing Technology Use

Two ethnographic studies[6,7] adopted the first strategy (see the first two studies in Table 1). Both explored the types of records and cues that people use to recall information from past meetings and were conducted in a corporate context. They considered a series of project-related meetings through interviews with a dozen people over several weeks or months. The first study surveyed an additional 500 people using different questions.[6] These users found audio-visual records useful for verifying or better understanding points in a meeting

1070-986X/12/$31.00 © 2012 IEEE

Published by the IEEE Computer Society

Table 1. Comparison of user studies for meeting browsing technology.

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[7]	12	Interviews	Practice	Importance of personal notes, need for to-do summaries.
Cremers[8]	8	Interviews	Practice/needs	Need for summaries and to-do lists.
Bertini[9]	118	Questionnaires	Practice/needs	Low utility of audio-visual records, except for people who missed a meeting or to find specific information.
Banerjee[5]	12	Interviews	Practice	Importance of topic lists.
Lisowska[10]	28	Elicitation of queries	Needs	Heterogeneity of queries, either about the interaction or about simple items in meetings.
Wellner[11]	21	Elicitation of observations of interest	Needs	Importance of facts, decisions, arguments leading to decisions, agenda, and dates.
Lisowska[12]	91	Wizard of Oz	Needs	Importance of training for modalitychoice in meeting browsing.

* All but the Banerjee[5] and Jaimes[6] studies were conducted by the AMI or IM2 Consortia.

and as an accurate overall record. In the second study,[7] users emphasized the limitations of official minutes for recalling specific details, which could be overcome by private notes. In both studies, users considered searching verbatim meeting records a potentially challenging task that could be facilitated through structured minutes with assigned tasks and decisions.

Two other ethnographic studies with 10 and 100 users,[8,9] respectively, confirmed and extended these insights. To retrieve information about a past meeting they attended, people mainly use minutes and personal notes, although they often rely on personal recollection or emailed information only. The utility of audio-visual recordings alone is low because watching the recording of an entire meeting is time consuming. Nevertheless, users still viewed recordings as useful for checking what someone said; as a proxy for people who missed a meeting; or as a reminder of past topics, assigned tasks, or the next meeting's date.

Eliciting New Requirements

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,[8] users suggested including arguments for decisions in automatically generated meeting minutes, along with lists of main topics and to-do lists in addition to the agenda and participant names. In another set of desiderata collected from professionals by a non-AMI/IM2 survey,[5] the most frequent wish was for the list of topics discussed at a meeting.

Several sets of explicit queries were collected from meeting technology developers and nontechnical users.[10,11] In one study,[10] 28 participants were asked to choose between several use cases—a manager tracking employee performance or project progress, an employee missing one project meeting, or an employee joining an ongoing project—and then formulate queries for a meeting archive, resulting in approximately 300 queries. Users showed interest in two main types of items:

- items related to the interaction between participants, such as decisions, questions, discussions, or disagreements and
- items that are conceptually related to meetings, such as dates, participants, documents, presentations, and global and local discussion topics.

Answering some of the queries would require complex processing such as topic detection or an understanding of the interaction structure, but others only involve elementary information.

From a different perspective, a large-scale Wizard of Oz study[12] with 91 subjects using a partially implemented interface found that exposure and training strongly impacted 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 because the system appeared to recognize it accurately, thanks to a dedicated human wizard who was hidden from the users.

Requirements Inferred from BET Statements

The AMI and IM2 Consortia proposed a browser evaluation test (BET) procedure (which we discuss in detail later) for collecting queries and using them in evaluations.[11,13] The BET ''observations of interest'' collected for a meeting can also be analyzed to infer user requirements for browsers. During the BET collection procedure, neutral observers were asked to formulate pairs of statements about a meeting, one factual and the other false. Observers followed these steps:

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. The experimenters consolidated statements with the same meaning 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 as follows (with differences noted in italics):

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

The observers created many more statements with a local scope rather than a global one; in the nonconsolidated set, 63 percent of the statements referred to specific moments in a meeting, 30 percent to short intervals, and only 7 percent 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 fell into five categories:

- decisions (8 percent);
- other facts stated by participants, including arguments leading to decisions (76 percent);
- statements related to the interaction process or the media used by participants (11 percent);
- statements about the agenda (2 percent); and
- statements about the date of the next meeting (2 percent).

The last two categories, although infrequent, were mentioned at least once by each observer. If we consider only 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 increases to 13, 4, and 3 percent, respectively, while the others decrease slightly. This shows that decisions, agendas, and the date of the next meeting were important to all observers.

Synthesis of User Studies

The user studies summarized in Table 1 show that requirements for meeting archiving and browsing technology are multifaceted. Their main dimensions, however, are now better understood than they were 10 years ago. We can categorize the focus of the requirements as follows:

- the targeted time span within a meeting or series of meetings (utterance, fragment, or entire meeting);

- the targeted media, such as audio, video, documents, presentations, or emails;
- the complexity of the information that is searched for, either in the media or inferred from the meeting content; and
- query complexity or modality.

Still, because user studies are often difficult to generalize, more publicly available studies are welcome, especially since the underlying technologies evolve constantly.

Meeting browsers exploit tools that record and analyze meeting data to build high-level indexes.

Meeting Browsers

Two main types of applications partially address the requirements we have discussed thus far. Meeting summarization systems offer an abstracted view of a meeting, structured for instance around its main topics (as in the early Meeting Browser[14] from Carnegie Mellon University's Interactive Systems Lab) or around the assigned tasks or action items (as in the Cognitive Assistant that Learns and Organizes [CALO] browser[15]). Some other meeting browsers are intended to help users with fact finding or verification (for example, to check figures, decisions, assigned tasks, or document fragments) although they can also be used to sample a meeting for abstractive purposes. Recent surveys include examples of both types,[1–3] which can also be classified according to the main rendered modality[2] or the complexity of their functionalities.[3] Meeting browsers exploit tools that record and analyze meeting data to build high-level indexes based on a variety of features, such as speech transcript, turn taking, attention focus, slide changes, and handwritten notes. These indexes are used within multimodal user interfaces to help users locate the information that will likely fulfill their needs.

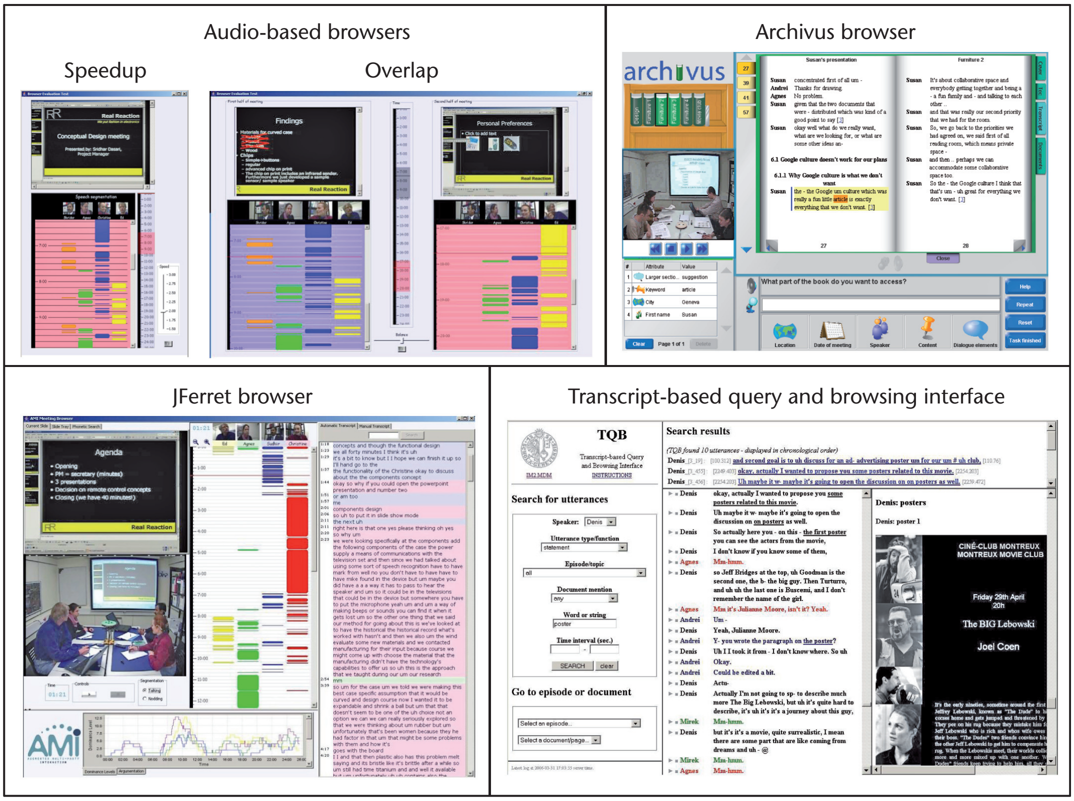
Meeting browsers intended for fact finding and verification (henceforth referred to simply as meeting browsers) have been the main focus of the AMI and IM2 Consortia because they strike a good balance between several divergent targets. First, they are part of promising transversal, complete applications—from 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 current 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 Computers in the Human-Interaction Loop (CHIL).

The meeting browsers developed within AMI and IM2 are best classified according to the modality they use to locate and render excerpts from a meeting. Figure 1 illustrates the range of media, components, and layout. *Speech-centric browsers* exploit the audio recordings and/or their transcripts, often with synchronized video, possibly accompanied by higher-level annotations such as named entities, topics, or extracted keywords. *Document-centric browsers* exploit document content and/or annotations such as slide changes, sometimes with speech and document alignment. Following this approach, two AMI/IM2-related systems were turned into commercial conference browsers.

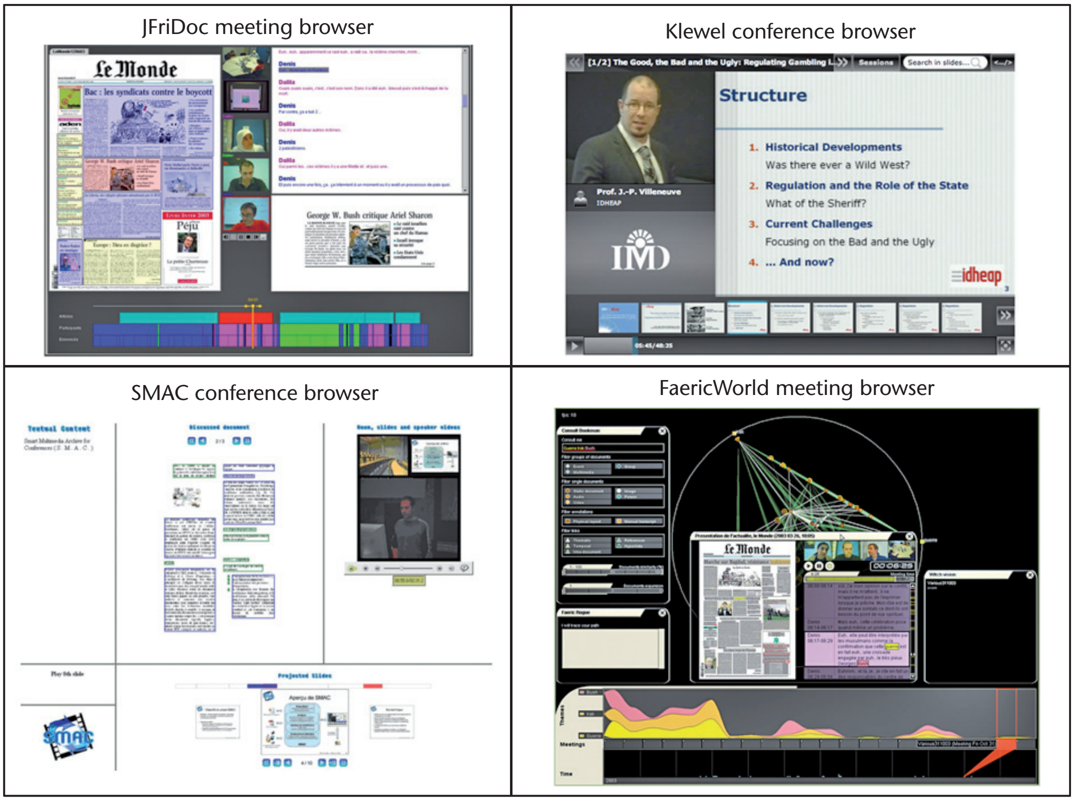
Speech-Centric Browsers

Two audio-based browsers provide access to audio recordings through speaker segmentation and slides, while enhancing browsing.[13] The Speedup browser accelerates audio playback, while the Overlap browser plays two different parts of a meeting in the left versus the right channels, allowing users to manually adjust the audio balance to focus on the most relevant channel.

The JFerret browser illustrates the main capabilities of the Java-based JFerret framework for browser design by providing access to audio, video, slides, automatic or manual transcripts, speaker segmentation, and potentially other annotations such as dominance levels.[11] The time-dependent components are synchronized to a main timeline displayed with the speaker turns.



(a)



(b)

Figure 1. Meeting browsers from the AMI and IM2 Consortia. (a) The speech-centric meeting browsers illustrate the diversity of media and layouts. Components include audio, video, and slide players, along with speaker identification and segmentation, transcripts, and various query parameters (in Archivus and TQB). (b) The document-centric meeting browsers and conference browsers exploit document content and/or annotations. Document and speech alignment is central to all layouts.

The Transcript-Based Query and Browsing (TQB) interface 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.[13] Archivus enables multimodal human-computer dialogue in a Wizard of Oz setting, which allows for partial implementation to help gather additional user requirements.[12] Archivus uses manual 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.

Document-Centric and Conference Browsers

JFriDoc 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.[16] The FaericWorld system extends this approach to collections of meetings and automatically calculates similarity links between all the categories of multimedia documents.[16] Users can then query the system with a full-text search or directly browse the document links using an interactive visualization.

Despite the number of research prototypes, no commercially available meeting browsers are available for end users. This is all the more surprising because several systems for holding remote meetings are commercially available, some of which even offer recording capabilities. Still, the browsers developed by AMI and IM2 have evolved toward 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. This might explain why they reached the product stage more quickly. 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 (www.klewel.com), while the other was developed by the University of Fribourg and CERN in Geneva within the Smart Multimedia Archive for Conferences project (SMAC,[17] http://smac.hefr.ch) and is in use at these institutions. Both systems extract several robust indexes, such as slide changes, text from slides, and slide/audio/video synchronization, which are helpful for browsing and provide some fact-finding support. In addition, the SMAC system can 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.[17]

Evaluation Methods and Results

Evaluating meeting-support technology is a challenging task,[1,3] which is necessary to demonstrate the appropriateness of design and to compare several designs, interaction paradigms, or meeting analysis components. As Zhiwen Yu and Yuichi Nakamura explained,[3] "the criteria used to evaluate a smart meeting system include user acceptance, accuracy [of recognition mechanisms], and efficiency [... that is,] whether a browser is useful for understanding the meeting content quickly and correctly." While recognition accuracy 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 evaluation approaches. The Filochat system from the early 1990s was one of the first browsers for speech recordings, which were time-aligned with personal notes.[18] A user study demonstrated the system's usability and helped to assess its features. Laboratory tests compared three conditions (notes only, speech only, or Filochat) by measuring accuracy and the speed of subjects who answered factual questions about what they had heard. The mutual influences of processing accuracy and user behavior were later studied for the CALO action item browser.[15] 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.[19] The results showed that JFerret outperformed two other browsers (as well as no browser at all as a control condition) in terms of impact on several parameters characterizing

The experiments confirmed that the BET questions and protocol are reliable indicators of performance.

participant satisfaction and meeting success, such as finding an acceptable solution to a given design task.

The need to compare meeting browsers, at the same moment or over time, is better satisfied by efficiency-oriented evaluations than user studies because 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.

Browser Evaluation Test

We introduced the BET earlier in this article as a procedure for collecting observations of interest about a meeting, further transformed into pairs of true and false statements that were analyzed to infer requirements.[11,13] These statements are in fact mostly intended for evaluation using the following protocol. The experimenters asked subjects (who had not served as observers) to examine pairs of BET statements using the browser under evaluation and to determine for each pair which statement was true. Browser performance is quantified using precision (the number of correct answers) and speed (the number of pairs of statements processed per unit of time). Precision indicates effectiveness while speed indicates browser efficiency, when averaged over comparable groups. Of course, we could apply behavior analysis, satisfaction questionnaires, and other observational techniques as well.

When using the BET as a valid test protocol, we must acknowledge possible biases and limitations. First, as with any evaluation method, the BET checks to what extent browsers conform to certain user requirements. However, the BET elicitation method biases these requirements toward fact finding or verification, at least compared to other elicitation studies that have emphasized higher-level elements of interest such as action items, topics, or decisions. Although these might be underrepresented in the current BET set, a different set could also be elicited with an inverse bias.

Unlike many other user-oriented evaluations, we did not choose the BET observers and subjects from the meetings participants, although we encouraged the observers to make observations that would have been of interest to participants. Therefore, the BET requirements and evaluation task are targeting null-context users. They cannot be used for comparative evaluation of browsers that offer subjective memorization devices, such as personal notes taken during a meeting, although it is possible to make comparisons involving third-party notes. The BET's somewhat focused spectrum is the price we pay to ensure the method's reproducibility, enabling comparisons 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 because the precision baseline is not always 50 percent, time can be constrained in several ways, and the subjects' competencies and training can vary across groups. Even within a single experiment, the variability of human performance tends to decrease statistical significance. Subjects might have different strategies, some favoring precision over speed or vice versa—variability appeared 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.

BET Results and Lessons Learned

More than 100 subjects have evaluated several AMI/IM2 browsers using the BET in separate experiments. Table 2 summarizes the results in terms of precision and speed, with 95 percent confidence intervals. All the experiments used the three meetings from the AMI Corpus for which BET questions were produced, although the order of presentation, the definition

Table 2. Comparative results of meeting browsers evaluated in similar conditions using the Browser Evaluation Test.

Browser	Condition	Subjects	Average time to answer a question (sec)	Confidence intervals*	Average precision	Confidence intervals*
Audio-based	Speedup	12	99	26	0.78	0.06
browsers[13]	Overlap	15	88	23	0.73	0.08
JFerret[11]	BET set (pilot)	10	100	*43*	0.68	*0.22*
	Five global questions	5	<180	0	0.45	*0.34*
	Five factual questions	5	<180	0	0.76	*0.25*
TQB[13]	First meeting	28	228	129	0.80	0.09
	Second meeting	28	92	16	0.85	0.06
	Both meetings	28	160	66	0.82	0.06
FriDoc[16]	With speech/document links	8	113	n/a	0.76	n/a
	Without links	8	136	n/a	0.66	n/a
Archivus	True-false questions	80	127	*36*	0.87	*0.12*
	Open questions	80	n/a	n/a	0.65	*0.22*

* Confidence intervals at 95 percent are absolute values. When they could not be found, standard deviations are given instead (in italics).

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, this synthesis aims not to determine 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 was approximately 2 minutes, with a 1.5- to 4-minute range. Precision—generally against a 50 percent baseline except for open-answer conditions (JFerret and Archivus)—was in the 70 to 80 percent range, with higher values for browsers that use more human-processed information (TQB and Archivus). Thus, more knowledge appears to help increase precision, but this often means that subjects will also spend slightly more time as they manipulate more complex information. The experiments confirmed that the BET questions and protocol are reliable indicators of performance because the variance of the average answers was small enough to observe significant differences between conditions within experiments. Also, the variance compared favorably to the values observed in interactive question answering evaluation experiments at the Text Retrieval Conferences (TREC) or the Cross-Language Evaluation Forum (CLEF).

The main lessons learned from the BET evaluations, apart from the BET procedure's reliability, concern the AMI/IM2 technologies that appear to be useful for meeting browsing. Transcripts are used intensively when they are of high quality. Users tend to perform keyword searches on the transcripts, thus pointing to the need for improved speech-to-text systems. However, transcript annotations 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 they are available in the browser, especially when shown along the meeting's timeline—for example, using automatic slide change detection and speech and document alignment. Slides can even partly compensate for a lack of transcript, as seen for audio-only browsers, which score only slightly below transcript-based browsers. 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.

Finally, learning effects appeared to be important; even a single training session improved the subjects' performance significantly and conditioned their preference for

> **Two end-user products for indexing and browsing conference recordings integrated significant know-how from the consortia.**

browsing modalities. Although this is good news for product developers, it introduces an additional variable that must be controlled in evaluation experiments.

Conclusion

We have presented some of the AMI and IM2 Consortia's main achievements in meeting browser requirements and design made between 2002 and 2010. The resulting picture of the software development process departs considerably from the waterfall model, where users have the primary role of formulating task requirements 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 underspecified or were beyond the reach of current technology. Moreover, researchers, who were in most cases the system designers, believed it was important to include additional functionalities that seemed potentially useful and that 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.

Meeting browser R&D has gone through four main iterations of the software process, although not always in a strict time sequence. In the first iteration, several studies elicited user requirements and explored current technology uses to identify and prioritize needs. In a second iteration, a specific but not fully implemented prototype was studied in Wizard of Oz experiments accompanied by performance measures and user behavior analysis. In a third iteration, several standalone prototypes were implemented and could be compared using 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 customer satisfaction assessment.

The analyses we present here show that, on the one hand, user requirements for meeting browsing cannot constitute a rigid, set-in-stone specification. They greatly depend on how subjects are prompted to respond and must be gradually focused on 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 their own approach's utility. During the eight years of the AMI and IM2 Consortia's existence, with literally hundreds of researchers collaborating, jumping back and forth from the users' to the developers' perspectives has enabled us to gradually focus on the fact-finding task, providing an application framework to develop innovative technologies and a reliable benchmark to evaluate their usefulness in a user-oriented setting. **MM**

Acknowledgments

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

References

1. M.M. Bouamrane and S. Luz, ''Meeting Browsing: State-of-the-Art Review,'' *Multimedia Systems,* vol. 12, nos. 4–5, 2007, pp. 439–457.
2. S. Tucker and S. Whittaker, ''Accessing Multimodal Meeting Data: Systems, Problems, and Possibilities,'' *Machine Learning for Multimodal Interaction,* S. Bengio and H. Bourlard, eds., LNCS 3361, Springer-Verlag, 2005, pp. 1–11.
3. Z. Yu and Y. Nakamura, ''Smart Meeting Systems: A Survey of State-of-the-Art and Open Issues,'' *ACM Computing Surveys,* vol. 42, no. 2, 2010, pp. 11–12.

4. S. Renals et al., *Multimodal Signal Processing: Human Interactions in Meetings,* Cambridge Univ. Press, 2012.
5. S. Banerjee, C. Rose, and A.I. Rudnicky, "The Necessity of a Meeting Recording and Playback System, and the Benefit of Topic-Level Annotations to Meeting Browsing," *Proc. 10th IFIP TC13 Int'l Conf. Human-Computer Interaction* (Interact 2005), LNCS 3585, Springer-Verlag, 2005, pp. 643–656.
6. A. Jaimes et al., "Memory Cues for Meeting Video Retrieval," *Proc. 1st ACM Workshop Continuous Archival and Retrieval of Personal Experiences* (CARPE 2004), ACM Press, 2004, pp. 74–85.
7. S. Whittaker et al., "Design and Evaluation of Systems to Support Interaction Capture and Retrieval," *Personal and Ubiquitous Computing,* vol. 12, no. 3, 2008, pp. 197–221.
8. A. Cremers et al., "The Project Browser: Supporting Information Access for a Project Team," *Proc. 12th Int'l Conf. Human-Computer Interaction* (HCII 2007), *Human Computer-Interaction, Part IV,* LNCS 4553, Springer-Verlag, 2007, pp. 571–580.
9. E. Bertini and D. Lalanne, *Total Recall Survey,* tech. report, Dept. of Computer Science, Univ. of Fribourg, 2007.
10. A. Lisowska, A. Popescu-Belis, and S. Armstrong, "User Query Analysis for the Specification and Evaluation of a Dialogue Processing and Retrieval System," *Proc. 4th Int'l Conf. Language Resources and Evaluation* (LREC 2004), European Language Resources Assoc., 2004, pp. 993–996.
11. P. Wellner et al., "A Meeting Browser Evaluation Test," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems* (CHI 2005), ACM Press, 2005, pp. 2021–2024.
12. A. Lisowska et al., "Minimizing Modality Bias when Exploring Input Preference for Multimodal Systems in New Domains: The Archivus Case Study," *Proc. ACM SIGHI Conf. Human Factors in Computing Systems* (CHI 2007), ACM Press, 2007, pp. 1805–1810.
13. A. Popescu-Belis et al., "Towards an Objective Test for Meeting Browsers: The BET4TQB Pilot Experiment," *Proc. 4th Workshop Machine Learning for Multimodal Interaction* (MLMI 2007), LNCS 4892, Springer-Verlag, 2008, pp. 108–119.
14. A. Waibel et al., "Meeting Browser: Tracking and Summarizing Meetings," *Proc. DARPA Broadcast News Transcription and Understanding Workshop,* D.E.M. Penrose, ed., 1998, pp. 281–286.
15. G. Tur et al., "The CALO Meeting Assistant System," *IEEE Trans. Audio, Speech, and Language Processing,* vol. 18, no. 6, 2010, pp. 1601–1611.
16. M. Rigamonti, D. Lalanne, and R. Ingold, "FaericWorld: Browsing Multimedia Events Through Static Documents and Links," *Proc. 11th IFIP TC13 Int'l Conf. Human-Computer Interaction* (Interact 2007), *Part I,* LNCS 4662, Springer-Verlag, 2007, pp. 102–115.
17. J. Herr, R. Lougheed, and H. A. Neal, "Lecture Archiving on a Larger Scale at the University of Michigan and CERN," *J. Physics: Conference Series,* vol. 219, no. 8, 2010, doi:10.1088/1742-6596/219/8/082003.
18. S. Whittaker, P. Hyland, and M. Wiley, "Filochat: Handwritten Notes Provide Access to Recorded Conversations," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems* (CHI 1994), ACM Press, 1994, pp. 271–277.
19. W. Post et al., "Experimental Comparison of Multimodal Meeting Browsers," *Proc. 12th Int'l Conf. Human-Computer Interaction* (HCII 2007), *Human Interface, Part II,* LNCS 4558, Springer-Verlag, 2007, pp. 118–127.

Andrei Popescu-Belis is a senior researcher at the Idiap Research Institute, Switzerland. His research interests include human language technology, multimodal data analysis, and design and evaluation of linguistic and interactive systems. Popescu-Belis has a PhD in computer science from LIMSI-CNRS, University of Paris XI. Contact him at andrei.popescu-belis@idiap.ch.

Denis Lalanne is a senior lecturer and researcher in the Department of Informatics at the University of Fribourg, Switzerland. His research interests include human-computer interaction, usability, visualization, and gesture analysis. Lalanne has a PhD in computer science from the Swiss Federal Institute of Technology in Lausanne (EPFL). Contact him at denis.lalanne@unifr.ch.

Hervé Bourlard is the director of the Idiap Research Institute, Switzerland, and a professor at EPFL. His research interests include signal processing and statistical pattern classification, with applications to speech and language modeling, speech and speaker recognition, and multimodal processing. Bourlard has a PhD in applied sciences from the Faculté Polytechnique de Mons, Belgium. He is an IEEE fellow and a senior member of the ACM. Contact him at herve.bourlard@idiap.ch.