

Using annotated meeting records to form an organizational knowledge repository

Derek Jacoby
University of Victoria
Dept of Computer Science
Victoria, BC, Canada
1-250-220-0467

derekja@uvic.ca

ABSTRACT

The gold standard of videoconferencing has always been to make a meeting feel as if it were in real life, providing all of the sensory and contextual cues of an in-person meeting. This does not go far enough. The current paper proposes a system to record and analyze meeting data to permit synchronous and asynchronous interaction by both local and remote participants over an organized and annotated meeting record. This capture of meeting records establishes a shared information space that is searchable and persistent across meetings. The set of captured meetings, annotated and segmented both automatically and through user annotations, can form a valuable organizational knowledge repository.

Keywords

Meeting recording, collaboration, audio recording, speaker identification, knowledge representation

1. Introduction

A great deal of modern corporate and academic culture is based around meetings. Meetings have several notable failings. They are relatively opaque in their decision making process to anyone who was not at the meeting. They depend on the fallible memories of the meeting attendees to faithfully record notes or minutes. They are also difficult to schedule because they require attendance at the same time and place. Videoconferencing provides a solution to the “same place” issue. Often, the term telepresence has been used almost as a synonym for videoconferencing. The current paper argues that the shared space that users want to be mutually present in is no longer a physical location, but rather a shared information space. This difference opens the door to rich semantic telepresence tools rather than merely high fidelity conferencing.

Videoconferencing has become a standard technology in industry and academia. Barely a day goes by in the life of many information workers without at least one remote interaction over services such as Skype or MeetingPlace. These meetings are often challenging due to technical problems and a lack of fidelity from low-resolution webcams, poor microphone performance, etc. The services are improving in terms of network reliability and transport issues and consumers are becoming more educated with regards to camera and microphone selection and placement. The goal of these improvements is primarily to faithfully replicate a high-fidelity in-person interaction. This is a worthy goal, but it does not go far enough because even in the case of the best videoconference the failings of an in-person meeting still remain. Fortunately, in the process of transmission, the meetings can be recorded. Once they are recorded they can be analyzed. One they are analyzed they can be archived in a way so as to allow

maximum retention and accessibility of the information generated during the meeting. This paper proposes a system being worked on at the University of Victoria VisID lab to record, analyze, and archive meeting records.

Passive audio/visual recording of meetings is common in certain fields. In architectural firms, for instance, planning meetings are sometimes recorded. The primary consumers of these recordings seem to be high priced lawyers who listen to countless hours of tapes to extract small details of poorly documented meetings when the parties turn around and sue each other. Otherwise the tapes are rarely used again. In other fields, it is far less common to record meetings because the interesting items to go back to are always lost in a sea of irrelevancies. This is the essential problem with recordings – they are not self-documenting. They are also linear and time consuming to listen to.

This is not to say that linear and time consuming is always bad. Indeed, if all of life were reduced to summaries it would be much less interesting. If a favorite speaker was speaking about your favorite topic, you'd probably want to listen to every word they said, and if possible, in person. But most meetings hardly fall into that category, particularly those that have already happened and are now part of history. The ideal would be to choose what you hear firsthand, what comes through recordings, and what you merely see summaries of.

Tools such as Google Wave provide richer spaces for online collaboration and introduce new methods of collaborating over meetings. In one recent class we took collaborative notes using Wave with a great deal of success. In Wave, you have a new conversation (called a “wave”) that has a number of users. Each user can say something (a “wavelet”) and it appears just like a chat session. Users can also edit each other's wavelets to produce more coherent shared documents. But in normal usage the system tends to break user contributions into fairly granular hierarchical wavelets. The current proposal intersects the Wave model of interaction by placing a similar focus on conversational turns. At the level of content analysis, this turn taking information is very important.

Although meetings are often inefficient and time-consuming, they remain the basis of much office collaboration. Tool support such as email and shared document systems relieve some need for meetings, but communication is easier over many decisions and discussions to simply meet in real-time and talk things through. The challenge to a meeting support system such as the current proposal is to not impose an undue burden on the natural flow of the meeting, but still provide analyzable information to track the

content of the meeting. This attempts to solve a major problem with most meetings currently, that the decision process is opaque to everyone except those in the meeting and the information resulting from the meeting is put through a very narrow filter of meeting minutes or fallible personal memories.

2. Scenarios

Just imagine if this direct memory of meetings were shared and accessible to everyone in an organization. Accessible in a way that allows the user to determine the level of detail and interact with the meeting records on a semantic level rather than a linear timeline. Consider the following scenarios:

a) Ben is working at an architecture firm and preparing for a meeting with the senior architects and the client over the bathrooms for a new office building. He is fairly new on the team and to get ready asks the meeting archive to pull up past discussions of plumbing fixtures. He spends an hour going through past discussions and quoting sections in his proposal.

b) On a software project at Microsoft, Jen is designing a user interface for an updated SQL user administration tool. She needs to collect the current problems and consider the design rationale behind the current design in writing her specification. If all meetings and customer calls were recorded into the system Jen could search across support databases and prior meetings for terms such as “Add user” and listen to portions of support calls and prior design meetings where the topic was discussed. Search tools to bring results down from thousands to usable numbers of relevant results, and summarization tools, will certainly be required!

Both of these scenarios require that the meetings be annotated and searchable at a very fine degree of granularity. The current research attempts to explore this future world in a more contained laboratory setting.

3. Related work

This is hardly the first proposal to record meetings. Studies of computer-facilitated meeting rooms go back to Xerox Parc over 20 years ago [Stefik, 1987.] As early as 1991 there were involved discussions of the effect of computer-facilitated meetings on organizational memory [Nunamaker, 1991.] More recent work at IBM focuses on the type of flexible work surface needed for effective retrieval of meeting content [Geyer, 2003.] Unfortunately, most attempts to annotate meeting records solely by users tend to become burdensome. In the speech recognition literature the burden on the user is lessened - automatic meeting capture has been a goal for many years – but solutions suffer from many technical issues. Nonetheless, systems are becoming ever more successful as evidenced by the recent US National Institute of Standards and Technology (NIST) push on meeting recording which has generated a number of independent meeting recognition systems such as the one reported by Andreas Stolcke in 2005. The speed of the technological development in this field means that whether the user interface community is ready or not,

the capture and display of large amounts of information derived from audio data is soon going to be necessary. A 2005 review by Tucker and Whittaker looks in detail at many of the approaches for meeting browsers. The current effort uses some elements of the automatic capture systems to bootstrap the user annotations and hopefully reduce the overall user burden. The user interface approach and focus on semantic relationships between small chunks of meetings offer further differentiation between the current work and past efforts.

4. Current Research

The physical facilities of the VisID research lab now bear some note. The main display is a rear-projected Smart DVIT (digital vision touch) screen in a wall configuration with two opposing couches in front. Another identical unit is on the other side of the couches in a tabletop configuration. Both have four HD projectors with 3840 x 2160 resolution (8.3 Mpixels), and have a size of 61” x 34” (70” diagonal). Meeting participants are provided with Hisonic wireless headset or lapel microphones. The audio signals are passed to the computer through an M-Audio NRV10 firewire mixer so that all audio channels are simultaneously recorded.

The meeting situations available in the lab consist of regular lab meetings, software design meetings, subject-specific meetings (a biology student group meets in the lab, for instance), and contrived lab study situations. The VisID group is also part of an architecture project, so the hardware in the lab is replicated in the office of an architecture firm in Vancouver, thus giving access to a set of architecture working meetings for analysis and evaluation when the prototypes reach that level of maturity.

When a meeting is recorded, it is virtually useless if it is not annotated. Finding what you are looking for in an hour of undocumented meeting recording is beyond frustrating. Current technology is not capable of completely transcribing and categorizing audio records on its own, but it can get part of the way there; at least as far as knowing who is speaking. Since conversational turns roughly correspond to speaker changes the first problem becomes one of picking out who is speaking. Initial experiments led to some success in automated speaker identification techniques, but the current focus is primarily on multichannel recordings with one channel per speaker. This multi-channel data may eventually be used to train a single channel speaker identification engine. Regardless of the technique used to separate speakers, the turn-taking implied by speaker changes provides an important initial segmentation of the audio data. You can think of all later operations taking place in chunks that roughly correspond to a conversational turn, although these chunks may be later re-sized by a user annotation.

So at this point we have an audio record segmented by speaker. That is a lot more useful than an unsegmented meeting record because you can jump around by speaker and see how long each person talked. We really haven't extracted any semantic content out of the meeting. Without semantic tagging at a fairly granular level the essential problem with audio data remains – it is just too slow to go back to. There are two complementary approaches to annotate the meeting chunks with semantic information. The first is a technical partial solution, the second is a social strategy.

First, the technical partial solution. Speech recognition has made great strides in recent years. The Windows 7 recognizer is quite usable as a single user system under intentional dictation conditions. When a user is carefully speaking to the system using a trained model the recognizer can achieve very reasonable levels of accuracy. In a meeting situation current speech recognition technologies are woefully inadequate, though. Multiple speakers mean untrained models and people talking over each other and speaking in very different patterns and speeds than the clear, even dictation needed to ensure acceptable recognition rates. Part of the goal of the speaker identification discussed above was to be able to load speaker-dependent models. The multi-channel recording also solves the issue of people talking over each other. But the real savior is the realization that we actually don't need very good recognition accuracy to enable search. The proposal is to use recognition results, which may be at 50% accuracy or less in some cases, to initially fill out a tag cloud of discussion topics that will form the seed for user annotations, as opposed to providing a direct transcription of the conversation.

Second, the social strategy. Meetings are difficult. Most people are busy and it is hard to get everyone in the same room at the same time. Videoconferencing solves the “in the same room” problem, but does nothing for the “at the same time” part of the issue. The problem is often not motivation as much as it is simply scheduling. Even when someone reluctantly misses an important meeting, they would still like to be helpful and be perceived as contributing. The social structure behind effective meeting capture strives to make the person who is unable to attend the meeting just as useful as the actual attendees, because they are the one who helps annotate and interpret the meeting data. Of course, there are many meetings that may not be interesting enough to incent users to annotate them – these meetings will rightfully disappear from the record through a lack of useful annotations.

The essential shift in thinking is similar to some of the reasoning advanced in “Total Recall” by Gordon Bell – storage is cheap, almost free, and the system benefits of total storage are overwhelming if the retrieval issue can be solved. But the retrieval problem is a big one in the case of the lifeblogging examples that Bell is concerned with. Nobody is likely to wade through a recording of every minute of someone else's life. Fortunately, in the meeting case we have a clear incentive for someone else to sit through parts of it and annotate. As all meetings are captured as a matter of course, and annotated as a matter of best practices, they become accessible as a resource to the organization as a whole.

The current focus is on obtaining appropriate multi-channel meeting data to work with and getting some user-friendly recording controls worked out. Appropriate database storage and other system plumbing are needed as well. But the interesting research problems really come in when this groundwork is all laid and working properly. The essential problem here is a user interaction problem. How do you make adding annotations easy enough that users will do it, and make access to the data easy enough that users will use it.

5. Future Directions

The current conceptualization of the user interface relies heavily on tag clouds and user annotations. The work surface itself is a fairly generic drawing surface. Added to this, of course, are audio (and eventually video) playback controls and a playlist. Search results are shown either in a list form (to enable easy sorting and filtering) or on the drawing surface to allow placement by tag cloud similarity. There is a natural tension between time-based presentations of the meeting data and semantically grouped presentations. Some mixing of the two presentations could even be interesting. An appealing use of the drawing surface is to be able to lay out very large maps of semantically linked conversations. So at a wide zoom you get an overall picture of meeting topics and can browse down into the data, as you zoom down you start to see more detailed groups and individual annotated audio chunks.

A tag cloud is merely a representation of applicable search terms. It is a good, but not perfect, visualization aid in this case. In a tag cloud, terms that are repeated will appear more prominently than terms that only appear once. More important terms can also be emphasized.

User annotations consist of edits to the tag cloud for a given chunk of audio data, grouping of chunks of meeting data into higher level topics, text notes about a chunk, hyperlinks to web data sources, and links to related meetings or meeting segments in the archive system.

Back to the scenarios presented in the introduction; as Ben or Jen wade through the results of their searches they listen to parts of some meetings, look at tag clouds of other meetings, and look at other people's annotations and summaries. They also remove inappropriate tag cloud items and provide additional meta-data through their interactions with the meeting results.

As users group their own related sets of meetings and concepts you end up with mindmaps of conceptual understandings that are naturally supported by instance-level records of many of the meetings, and the detailed interactions that led to that understanding. From a philosophical perspective, this shared information space should lead to much more transparent and informed decision making. At the very least it provides the opportunity to support a contentious decision with instance data about shared opinions. Most importantly, user annotations should arise naturally out of the use of the system rather than requiring a lot of explicit effort – those meetings and recordings that are not of sufficient interest to drive annotations will naturally begin to fall out of search results.

There are certainly many problems in the way of this vision. Technology problems, visualization and user interface problems, and social problems. The social problems around expectations of privacy are one major issue. Can we get to a point where all relevant meetings are recorded? In a work environment it seems possible if the value proposition is there and the storage costs are low enough. Setting the system up so that users have the social incentive to participate in the system is another challenge.

Telepresence as a term often reflects back to a time when interactions were largely synchronous and co-located. Today, the space that users would like to be remotely present in is primarily an information space. Tools that can help bridge the gap between a linear, unsearchable meeting recording and a conceptual information space allow users separated by time and space to interact over shared data in rich ways that build on each other as a valuable organizational knowledge repository.

6. ACKNOWLEDGMENTS

Thanks to Dr. Melanie Tory and the rest of the VisID research lab for their collaboration and support.

7. REFERENCES

- [1] Bell, G., & Gemmell, J. (2009). *Total Recall: How the E-Memory Revolution Will Change Everything*. Dutton Adult.
- [2] Nunamaker, J. F., Dennis, A. R., Valacich, J. S., Vogel, D., & George, J. F. (1991). Electronic meeting systems. *Commun. ACM*, 34(7), 40-61. doi: 10.1145/105783.105793.
- [3] Stefik, M., Foster, G., Bobrow, D. G., Kahn, K., Lanning, S., & Suchman, L. (1987). Beyond the chalkboard: computer support for collaboration and problem solving in meetings. *Commun. ACM*, 30(1), 32-47. doi: 10.1145/7885.7887.
- [4] Stolcke, A., Anguera, X., Boakye, K., Çetin, Ö., Grézl, F., Janin, A., et al. (2006). Further Progress in Meeting Recognition: The ICSI-SRI Spring 2005 Speech-to-Text Evaluation System. In *Machine Learning for Multimodal Interaction* (pp. 463-475). Retrieved November 18, 2009, from http://dx.doi.org/10.1007/11677482_39.
- [5] Tucker, S., & Whittaker, S. (2005). Accessing Multimodal Meeting Data: Systems, Problems and Possibilities. In *Machine Learning for Multimodal Interaction* (pp. 1-11). Retrieved November 17, 2009, from <http://www.springerlink.com/content/w7kebx3bgpf3gydq>.