

THE AMBIENT SPOTLIGHT: PERSONAL MEETING CAPTURE WITH A MICROPHONE ARRAY

Jonathan Kilgour, Jean Carletta, Steve Renals

School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK

ABSTRACT

We present the Ambient Spotlight system for personal meeting capture based on a portable USB microphone array and a laptop. The system combined distant speech recognition and content linking with personal productivity tools, and enables recognised meeting recordings to be integrated with desktop search, calendar, and email.

1. OVERVIEW

The Ambient Spotlight is a personal, laptop-based application which combines meeting recording using a microphone array, distant speech recognition, topic segmentation, and information retrieval to provide a way to automatically capture and structure recorded meetings, and to integrate their access with standard productivity applications such as calendaring and email. This work builds on recent work on new approaches to the capture, recognition and interpretation of multiparty meetings, performed within projects such as AMI/AMIDA [1] and CHIL [2].

There have been a number of previous demonstrations based on meeting recognition, such as multimodal browsers which synchronize automatic transcription and other annotations to a common time-line. Most closely related to the Ambient Spotlight is the AMIDA Content Linking Device [3], which uses realtime meeting speech recognition to generate the current context of a meeting, using recently recognised speech from the meeting. We have referred to this as “implicit search” or “online search without query”, since the content linking approach automatically constructs text queries (to the web, or to an application specific document collection – such as other recognised meetings) based on the recent output of the speech recogniser. The motivation for this work stemmed from the observation that meeting participants frequently type in web search queries during a meeting sometimes repeating verbatim a phrase that has just been spoken.

These previous demos have tended to focus on instrumented spaces or smart meeting rooms, in which the capture equipment is permanently installed in the meeting room. In contrast, the Ambient Spotlight is much lighter weight and

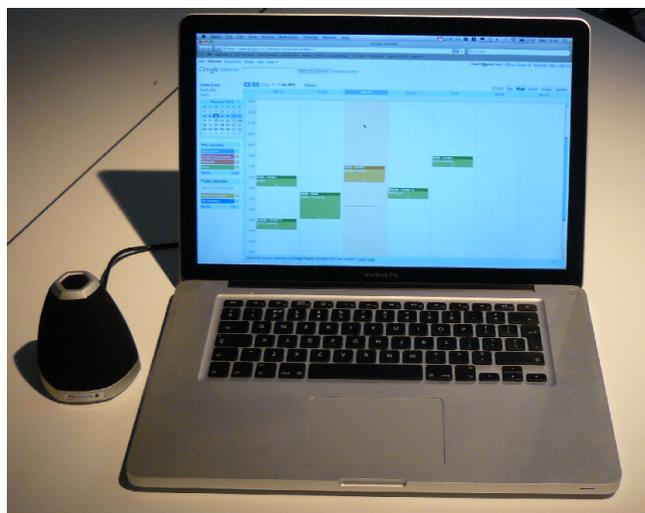


Fig. 1. Meeting capture using a Microcone with a laptop

portable; it uses the dev/audio Microcone, a 7-element microphone array in a conical geometry, which connects to a laptop using a USB interface (figure 1).

Our objective was the development of a system that can be used as part of a user’s personal productivity toolset, without requiring the need to use and master complex new software. The Ambient Spotlight uses a calendar as the main interface (upper left of figure 2), and is closely integrated with desktop search and email clients. The main idea of the Ambient Spotlight is to retrieve automatically those documents that are relevant to a particular part of the meeting. In a personal information management context, relevant documents are likely to be private documents on the users’ laptop (e.g., email, pdf documents, meeting recordings), and the information retrieval aspect of the document builds on pre-existing desktop search. We describe retrieval queries as ambient; There is no need for a user to enter explicit terms, the queries are constructed from the automatic speech recognition transcript. In practice, we beamform the meeting audio and transcribe it automatically, segmenting into chunks with a duration of 20s. A set of potentially relevant documents is retrieved for each chunk. This results in far too many potentially relevant results to present to a user, so we perform topic segmentation on the entire meet-

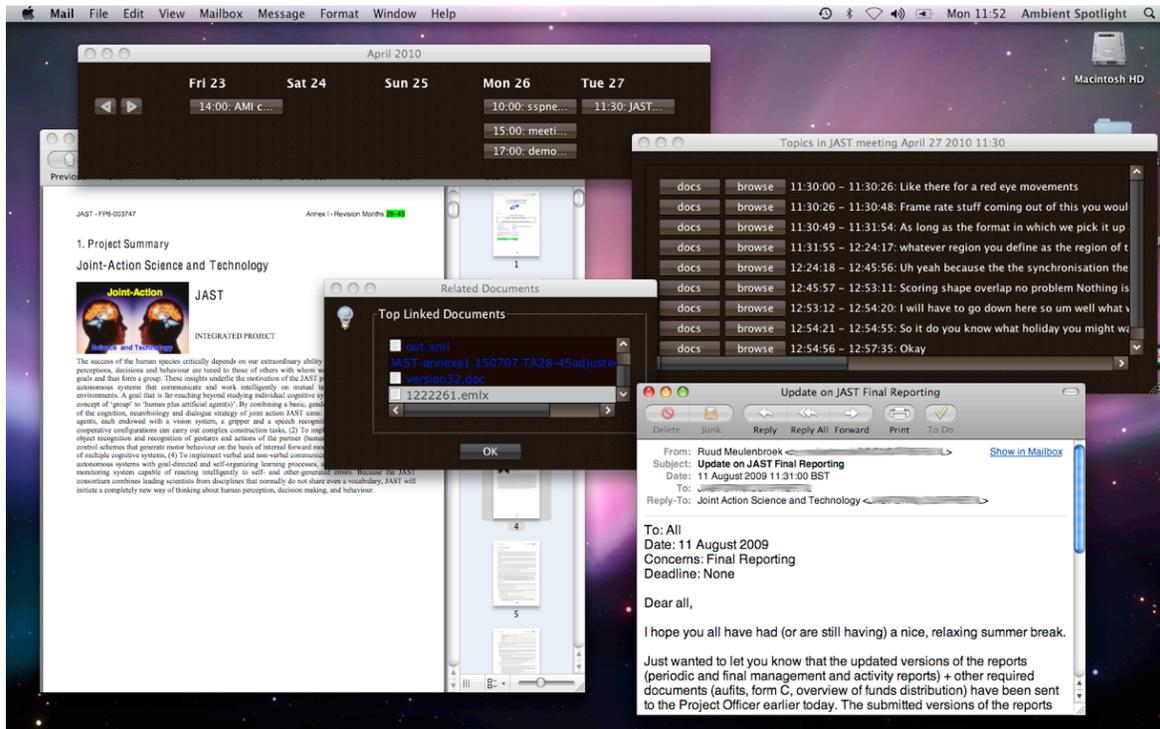


Fig. 2. Desktop interface of the Ambient Spotlight, including calendar interface (upper left), automatically linked document list (centre), topic segments (upper right), and examples of linked content (background left, lower right).

ing, and aggregate and prune the query results, resulting in a list of potentially relevant documents for each segment.

Most of the component technologies for this demonstration have been presented previously. The novelty of the demonstration lies in the application of distant speech recognition technologies in a personal setting, and their novel integration with productivity software that is used on a daily basis. This demonstration has recently been presented at conferences with a focus on multimedia [4, 5]; it has not previously been presented at a speech or audio processing workshop.

2. COMPONENTS

The Ambient Spotlight integrates a number of component technologies:

Meeting capture using a conical 7-element USB microphone array.

Speech recognition: using the AMI-ASR distant speech recognition system [6], which outputs speaker segments and speech transcripts.

Content linking: using standard desktop search (Spotlight (mdfind) on a Mac or Google Desktop).

Topic segmentation: based on the ASR transcripts [7] (upper right of figure 2).

Meeting browsing: in order to view linked multimodal material such as previously recorded meetings.

The demonstration software builds on the Google Calendar and Google Data APIs. Realtime transfer of multimodal annotations (e.g. speech transcripts or topic segments) is carried out using an annotation server referred to as the AMIDA Hub [3].

3. REFERENCES

- [1] S. Renals, "Recognition and understanding of meetings," in *Proc. NAACL/HLT*, 2010.
- [2] A. Waibel and R. Stiefelhagen, *Computers in the Human Interaction Loop*, Springer, 2009.
- [3] A. Popescu-Belis, P. Poller, J. Kilgour, E. Boertjes, J. Carletta, S. Castonovo, M. Fapso, M. Flynn, A. Nanchen, T. Wilson, J. de Wit, and M. Yazdani, "A multimedia retrieval system using speech input," in *Proc. ACM ICMI-MLMI*, 2009, pp. 223–224.
- [4] J. Kilgour, J. Carletta, and S. Renals, "The Ambient Spotlight: Queryless desktop search from meeting speech," in *Proc. SSCS - ACM Multimedia Workshop on Searching Spontaneous Conversational Speech*, 2010.
- [5] J. Kilgour, J. Carletta, and S. Renals, "The Ambient Spotlight: Personal Multimodal Search Without Query, year = 2010," in *Proc. ICMI-MLMI*.
- [6] P. Garner, J. Dines, T. Hain, A. El Hannani, M. Karafiat, D. Korchagin, M. Lincoln, V. Wan, and L. Zhang, "Real-time ASR from meetings," in *Proc. Interspeech*, 2009.
- [7] P.-Y. Hsueh, J. D. Moore, and S. Renals, "Automatic segmentation of multiparty dialogue," in *Proc. EACL06*, 2006.