The Ambient Spotlight

Personal Multimodal Search Without Query

Jonathan Kilgour, Jean Carletta, Steve Renals

Centre for Speech Technology Research / School of Informatics
University of Edinburgh
Edinburgh EH8 9AB, UK
{jonathan,jeanc,srenals}@inf.ed.ac.uk

ABSTRACT

The Ambient Spotlight is a prototype system based on personal meeting capture using a laptop and a portable microphone array. The system automatically recognises and structures the meeting content using automatic speech recognition, topic segmentation and extractive summarisation. The recognised speech in the meeting is used to construct queries to automatically link meeting segments to other relevant material, both multimodal and textual. The interface to the system is constructed around a standard calendar interface, and it is integrated with the laptop's standard indexing, search and retrieval.

Category and subject descriptors: H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems
General terms: Design, Human Factors
Keywords: speech, meetings, content linking, search-without-query

Demo video: `http://homepages.inf.ed.ac.uk/jonathan/unadvertised/ambient_spot_qt.mov`

1. INTRODUCTION

There has been intensive work on the development of new approaches to capture, recognise and interpret multiparty meetings, for example the AMI/AMIDA projects [5] and the CHIL project [6]. These projects developed new approaches to meeting capture using multiple cameras and microphones, a variety of multimodal recognition technologies (including speech recognition, speaker diarisation, and visual focus of attention), and approaches to automatically interpret meeting content including topic segmentation, summarisation, dialogue act recognition and the identification of subjective content.

This set of component technologies has led to a number of application prototypes and demonstration systems. A number of meeting browsing demonstration systems have been developed—and demonstrated at previous ICMI and MLMI conferences—which provide interfaces to browse media files, transcripts, segmentations, and other annotations synchronised to a common time line. More recently, we developed the AMIDA Content Linking Device, which was demonstrated at ICMI-MLMI 2009 [4]. Content linking uses realtime conversational speech recognition to automatically transcribe a meeting in progress, then constructing search queries from the recently detected words in order to retrieve relevant multimodal and text documents. Such a "search-without-query" approach may be viewed as a way to automatically provide context (in the form of relevant documents and media files) to an ongoing multiparty conversation, without requiring explicit search.

The above prototype systems have generally made use of complex hardware setups, requiring specifically instrumented spaces. This makes it difficult to imagine such systems being integrated as part of a user's personal productivity toolset. The demonstration we present here, called *The Ambient Spotlight*, is specifically designed for personal use [3]. It is laptop-based, using a plug-and-play USB microphone array, and has been designed to become a natural part of the work environment rather than an extra application to learn. It is closely integrated with the standard desktop search, calendar, and email tools.

The concept of the Ambient Spotlight is to bring attention to related documents when the user is reviewing meetings. In this context documents can be anything on the users' laptop including captured meetings, presentations, emails, podcasts and PDF documents. Queries are made in an ambient fashion in the sense that no query terms are entered by the user, an approach similar to the AMIDA Content Linking Device. The captured meeting audio is beamformed and passed through speech recognition software. The resulting transcript is segmented into 20 second chunks, with each chunk being analysed to produce a query used to search for documents on the user's laptop. These results are too low-level to present directly to the user so they are aggregated to provide the most popular results over an entire meeting, or—more usefully—over automatically derived topic segments from the meeting. The Ambient Spotlight uses information from a calendar application as a way of structuring the initial navigation of captured meetings.

Most of the component technologies for this demonstration have been presented previously. The focus of this demonstration is their novel integration and customisation to create the Ambient Spotlight. The components of the Ambient Spotlight include meeting capture, speech recognition, automatic query generation, topic segmentation, and the aggregation of results.

2. TECHNOLOGIES

Meeting capture: Although existing room-based systems can be used to capture meetings for the system, we have focused on meetings captured using a lightweight personal setup based on an Apple MacBook Pro and a Dev/Audio Microcone (figure 1). The Microcone is a conical USB microphone array, which includes seven

ICMI-MLMI'10 November 8-12, 2010, Beijing China
Copyright 2010 ACM 978-1-4503-0414-6/10/11 ...$10.00.

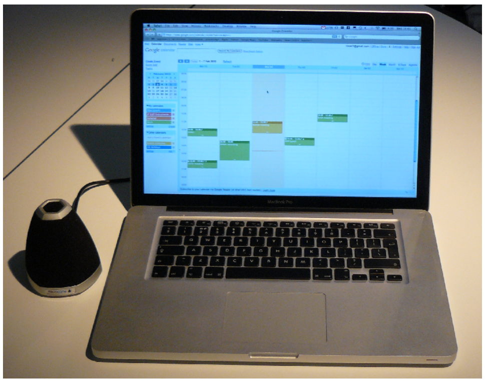


Figure 1: Meeting capture using a Microcone with a laptop

microphones, pre-amps and A/D conversion in the cone. Currently meeting capture has to be started and stopped by hand, although future versions could be controlled automatically based on the calendar, or could be always recording—although the latter option raises both privacy and disk space challenges.

Speech recognition: We employ the AMI-ASR speech recognition system developed in the AMI and AMIDA projects [1]. This ASR system was developed specifically for multiparty conversational speech, captured using multiple distant microphones. The system can run in real-time on the laptop or on a dedicated ASR server machine, or can run on a local compute cluster or as a webservice in the cloud (`webasr.com`). The ASR component outputs speaker segments and speech transcripts, which are fed into *The Hub*. The Hub was developed as part of the AMIDA Content Linking Device, and allows annotation data to be transferred between software modules in real-time, storing all data to be available for future query.

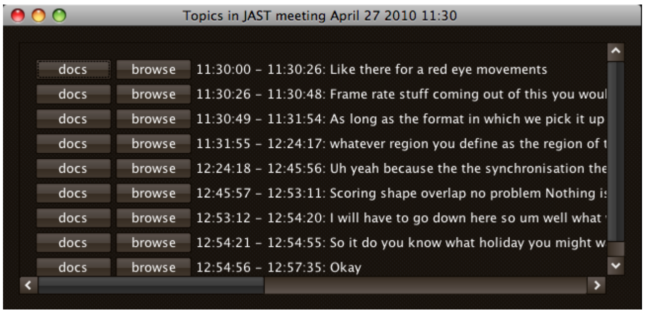
Content linking: The automatic speech transcription is read from the Hub, and for each segment of 20 seconds, the transcribed words are turned into textual queries to search over the documents on the laptop. If keywords are available (either from calendar entries or specifically associated documents) then these may be used to weight the query. We carry out the search using the Mac Spotlight tool (via the `mdfind` command); other desktop search tools could be used, e.g. Google Desktop.

Topic segmentation: It is useful to segment a meeting into topics, based on the speech transcript. This allows automatically linked content to be associated with a specific topic within a meeting, rather than the whole meeting. This can enable both more specific linked content as well as an easier to navigate interface. We have employed a topic segmenter based on [2] that uses only lexical features for both segmentation and labelling. An example of topics generated in this way is shown in figure 2.

Meeting browsing: As well as text documents, it is also useful to link multimodal content, such as other meetings (or meeting segments), provided that they have also been automatically recognised. To view linked multimodal content we use a meeting browser to enable playback, navigation, search, and summarisation of synchronised media files.

3. INTERFACE AND INTEGRATION

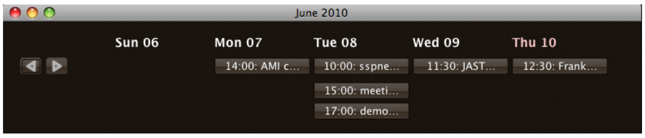
The Ambient Spotlight interface is centred on a calendar display, shown in Figure 3, which is obtained using Google Calendar and the Google Data APIs. The initial view of the calendar shows the current date furthest to the right with older meetings to the left. This reinforces the idea that this is a tool for reviewing past meetings.



Topics in JAST meeting April 27 2010 11:30

docs browse 11:30:00 – 11:30:26: Like there for a red eye movements
docs browse 11:30:26 – 11:30:48: Frame rate stuff coming out of this you woul
docs browse 11:30:49 – 11:31:54: As long as the format in which we pick it up
docs browse 11:31:55 – 12:24:17: whatever region you define as the region of t
docs browse 12:24:18 – 12:45:56: Uh yeah because the the synchronisation the
docs browse 12:45:57 – 12:53:11: Scoring shape overlap no problem Nothing i:
docs browse 12:53:12 – 12:54:20: I will have to go down here so um well what
docs browse 12:54:21 – 12:54:55: So it do you know what holiday you might w
docs browse 12:54:56 – 12:57:35: Okay

Figure 2: The Topic Display



June 2010

Sun 06	Mon 07	Tue 08	Wed 09	Thu 10
	14:00: AMI c...	10:00: sspne...	11:30: JAST...	12:30: Frank...
		15:00: meeti...		
		17:00: demo...		

Figure 3: The Ambient Spotlight Calendar Display

When a meeting is clicked in the calendar window, a query is sent to the Hub to determine if the meeting has been captured and recognised. If meeting has not been captured, then the user can choose to start meeting recording. Otherwise, the topic display (Figure 2) is brought up, showing the output of the topic segmentation process for that meeting with the timings and label of each topic to the right of a pair of buttons, `browse` and `docs`. Clicking on the `browse` button pops up a meeting browser for the meeting and make it jump to the appropriate point. Clicking on the `docs` button pops up a display of the most relevant documents related to that meeting segment.

ACKNOWLEDGMENTS

This work was supported by a University of Edinburgh IDEA Lab POPP project.

4. REFERENCES

[1] 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.

[2] P.-Y. Hsueh, J. D. Moore, and S. Renals. Automatic segmentation of multiparty dialogue. In *Proc. EACL06*, 2006.

[3] 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.

[4] A. Popescu-Belis, P. Poller, J. Kilgour, E. Boertjes, J. Carletta, S. Castronovo, 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*, pages 223–224, 2009.

[5] S. Renals. Recognition and understanding of meetings. In *Proc. NAACL/HLT*, 2010.

[6] A. Waibel and R. Stiefelhagen. *Computers in the Human Interaction Loop*. Springer, 2009.