The CALO Meeting Assistant

L. Lynn Voss
Engineering and Systems Division
SRI International
Menlo Park, CA 94025
loren.voss@sri.com

Patrick Ehlen
CSLI
Stanford University
Stanford, CA 94305
ehlen@stanford.edu

and

The DARPA[†] CALO Meeting Assistant Project Team*

Abstract

The CALO Meeting Assistant is an integrated, multimodal meeting assistant technology that captures speech, gestures, and multimodal data from multiparty interactions during meetings, and uses machine learning and robust discourse processing to provide a rich, browsable record of a meeting.

1 Introduction

Technologies that assist in making meetings more productive have a long history. The latest chapter in that history involves projects that integrate recent advances in speech, natural language understanding, vision, and multimodal interaction technologies in an effort to produce tools that can perceive what happens at a meeting, extract salient events and interactions, and produce a record of the meeting that people can later consult or analyze.

Research projects such as the ICSI Meeting Project (Janin et al 2004) have sought to produce automated and segmented transcripts from natural, multiparty speech as it occurs in meetings. Others like the ISL Smart Meeting Room Task (Waibel et al 2003), and the M4 and AMI projects (Nijholt, op den Akker, & Heylen 2005) employ instrumented meeting rooms to collect multiple streams of behavior data and analyze the interactions of meeting participants to produce a rich and flexible record of their meeting activities, while also providing a supportive environment for collaboration.

The CALO Meeting Assistant is similar to the latter in that it collects multiple streams of information about the behaviors of people in meetings, and assimilates speech, movement, and note-taking behavior to create a rich representation of the meeting that can be analyzed and reviewed at many levels. But in addition, a primary aim of our meeting assistant is to integrate its observations with those of a larger system of agents, which can assess the meeting data it collects in the context of the ongoing projects and workflow in the work lives of each of the meeting participants. Thus, it aims to reach beyond an intelligent room that understands only the activities of people in meetings, and attempts to understand the overarching concerns of those people and to interpret their behavior from the perspective of what those meetings might mean to them.

That overarching system of agents is being developed under the DARPA CALO (Cognitive Assistant that Learns and Organizes) program, which seeks to produce machine learning technology in the form of personalized agents that support high-level knowledge workers in carrying out their professional activities. The CALO system handles a broad range of interrelated decision-making tasks that are traditionally resistant to automation, partly by interacting with, being advised by, and learning from its users. It can take initiative on completing routine tasks, and on assisting when the unexpected happens.

CALO is designed from the ground up as a cognitive system. Whereas conventional, hand-coded software excels at a narrow set of capabilities in a

[†] This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. NBCHD030010. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA or the Department of Interior-National Business Center.

[*] The DARPA CALO MA Project is a collaborative effort among researchers at Adapx, CMU, Georgia Tech, MIT, SRI, Stanford University, UC Berkeley, and UC Santa Cruz.

particular domain, cognitive systems maintain explicit, declarative models of their capabilities, ongoing activities, and operating environments. These models enable CALO to extend and improve its capabilities through learning and adaptation. Cognitive systems are better equipped to cope with unexpected developments, learn to improve over time, and adapt to the contexts and requirements of different situations. CALO also uses natural interfaces that enable simple, effective interactions with humans and other cognitive systems.

The CALO Meeting Assistance Project is developing capabilities to enable CALO to participate in group discussions and meetings. Unlike instrumented "intelligent room" meeting projects, our system is designed for users in an office environment with access to the Internet, a laptop, and some small, off-the-shelf peripheral devices (such as headsets, webcams, and digital writing devices) to capture speech, gestures, and handwriting. It aims to be unobtrusive, leveraging cross-training, unsupervised learning, and lightweight supervision captured from normal user interaction (e.g., users reviewing and editing notes, or adding detected action items to a to-do list).

These data are transparently processed at a central server location and redistributed, so the meeting assistant interacts seamlessly with other CALO desktop functionalities, using a common ontology.

2 What it does

The CALO Meeting Assistant helps its owners by capturing and interpreting meeting conversations and activities, and, as appropriate, retrieving relevant information. Information gleaned from a meeting can be incorporated in the respective owner's CALO knowledge stores to, for example, track commitments and remember references to projects, people, places, and dates. An archive of each meeting provides a searchable record for users, as well as a history of training data for CALO's learning components. Areas of learning include:

Speech processing: Automatic transcriptions are produced from conversational speech among multiple speakers, while adapting to speaker and background noise, recognizing prosodic cues, learning new vocabulary, and constructing person, role, and topic-specific language models.

Visual recognition: Faces, gaze direction, gestures, and activities are detected, and detection is improved through lightly-supervised learning and unsupervised cross-training.

Discourse understanding: Dialog moves are recognized, topics are segmented and grouped through supervised and unsupervised generative models, action items are detected, and discussions can be threaded across documents and email.

Multimodal reinforcement: Pen, speech, and text inputs combine to offer natural communications.

Meeting activity: Speech and note-taking activities combine to provide cross-training for recognizing meeting phases, and for tracking agendas and document usage.

3 Demo

We demonstrate how the meeting assistant captures speech, pen, and other meeting data using an ordinary laptop, produces an automated transcript, segments by topic, and performs shallow discourse understanding to produce a list of probable action items arising from a single, pre-recorded meeting. We then demonstrate a Meeting Rapporteur that provides a meeting summary and allows participants to review and organize the meeting transcript, audio, notes, action items, and topics—all while providing actions in a feedback loop that supports the meeting assistant's semi-supervised learning process. Finally, we also discuss the potential and current development of real-time capabilities that allow users to interact with the meeting assistant during an ongoing meeting.

References

Janin, A., Ang, J., Bhagat, S., Dhillon, R., Edwards, J., Marcias-Guarasa, J., Morgan, N., Peskin, B., Shriberg, E., Stolcke, A., Wooters, C., and Wrede B. 2004. The ICSI meeting project: Resources and research. In *Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP '04) Meeting Recognition Workshop (NIST RT-04)*.

Nijholt, A., op den Akker, R., and Heylen, D. 2005. Meetings and meeting modeling in smart environments. *AI & Society*, 20(2):202-220.

Waibel, A., Schultz, T., Bett, M., Denecke, M., Malkin, R., Rogina, I., Stiefelhagen, R., and Yang, J. 2003. SMaRT: The smart meeting room task at ISL. In *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, pp 752-755.