Facilitating Meetings with Playful Feedback

Ying Zhang
Marshall Bern
Juan Liu
Kurt Partridge
Bo Begole
Bob Moore
Jim Reich
Palo Alto Research Center
3333 Coyote Hill Road.
Palo Alto, CA 94304 USA
yzhang@parc.com

Koji Kishimoto
Fujitsu
Tokyo, Japan
Koji.koshimoto@jp.fujitsu.com

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

ACM 978-1-60558-930-5/10/04.

Abstract
Effective group meetings are important for the productivity of corporations. Various types of meeting facilitators have been developed over the past couple of years. We present a prototype that is unique because it captures both individual and group behaviors and provides *real time* playful feedback. The portable prototype includes a set of table-top microphones with an audio interface to a laptop PC, where audio data are processed and an avatar-based UI displays the shared state of individual and group behaviors during a meeting. The interface reveals not only level of participation, but also several other meaningful but harder to detect behaviors such as turn taking, interruptions, and group laughter. The presentation's design is deliberately playful to keep participants monitor, self-estimate and improve their meeting behavior.

Keywords
Meeting facilitation, conversation behavior detection, playful feedback, avatar-based UI

ACM Classification Keywords
H5.1. Information Multimedia Information. H.5.2. User Interfaces. H.5.3. Group and Organization Interfaces

General Terms

Design, Measurement, Human Factors

Introduction

In meetings, as in most forms of verbal interaction, the right to talk is a limited resource [10]. The participants must manage who will speak next locally and on a turn-by-turn basis [10]. We argue that one key part of poor communication in meetings is the ineffective and inefficient allocation of speaking turns. Some meeting participants talk too much, or at inappropriate times, while others may need encouragement to contribute, or are interrupted or blocked from speaking by more vocal participants [1]. One of the roles of a human meeting facilitator is to help the participants allocate turns-at-talk. Facilitators draw attention to participants who have not contributed and explicitly invite them to participate, at times shutting down other participants who are dominating the floor. In our research, we ask the question: Can an automated system detect turn-taking behavior in meetings and how can it help facilitate the allocation of turns-at-talk more effectively?

We designed and implemented a prototype of the "Ubiquitous Meeting Facilitator" (UMF) that monitors the level of participation, detects turn-taking and various types of interruptions, and shows the individual behaviors and group states visually in a shared display (Fig. 1). We intend to make the display not too personal or serious -- such a system could seem threatening or insulting. We use avatars to represent participants, sizes of avatars to represent level of participation, and facial expressions to represent individual behaviors, etc. Simultaneously, group states, such as prolonged silence, interruptions, overlapping conversation, and group laughter, are captured by background colors and icons. Note that laughter detection provides not only information about emotion, but it also reduces false alarms from overlapping speech. Meeting events are detected and recorded, so that the effectiveness of the meeting can be analyzed offline. The system is portable with table-top microphones connecting through an audio interface to a laptop. In addition to real-time facilitation, the system can also use recorded audio files to analyze past meetings.

figure 1. The Ubiquitous Meeting Facilitator (UMF).

There has been ample previous work on capturing meetings for offline analysis. For example, the AMI and AMIDA projects [9] addressed the recognition and interpretation of multiparty meetings. The VACE project [11] developed systems and tools to automatically detect, extract, and report high interest people, patterns, and trends in visual content from foreign news, using video information capture, indexing, and retrieval technology. The CALO Meeting Assistant [8] is an integrated, multimodal system that captures speech and gestures and uses machine learning and robust discourse processing to provide a rich, browsable record of a meeting. Our system, UMF, which uses only audio, is simpler and lighter weight than these offline systems, yet it retains the core functionality of capturing meeting events and conversational status.

There has also been work on real-time GUIs for meeting participation. Conversation Clock [2] displays a visualization of each speaker's activity over time, however the shared display only shows level of participation from individuals, while conversational behaviors such as turn-taking and interruptions are not clearly depicted. DiMicco et. al. [3] describe a shared display for influencing group participation by exposing each speaker's total speech time. Again, no conversational behaviors are shown other than the total amount of speech time from individuals. The MIT Media Lab's Meeting Mediator (MM) [6] consists of portable sociometric badges for capturing data and cellphones for displaying meeting status. MM emphasizes balance, using the position of a colored circle to depict the dominating behavior intuitively. However, there are no visual highlights for interruptions, common laughter or group behaviors. UMF, on the other hand, displays a rich set of high-level features based on Conversation Analysis, including turn-taking, domination, interrupts, laughter, and other group behaviors such as long overlapping or silence. UMF can be used for both online facilitation and offline meeting analysis. Although the current system is implemented for a shared display, it should adapt easily to individual displays on cellphones.

UMF System Design

The design objective for UMF was to build a prototype of a meeting facilitator for group brainstorming meetings, with at most 12 participants. At the outset, we determined that the system should:

- be portable, and cheap and easy to install;
- be humorous, non-intrusive, and not too personal;
- provide real-time feedback of meeting progress;
- detect over- and under-participation;
- detect and display interruptions;
- distinguish laughter from speech;
- detect and display group state and individual behavior; and
- enable after-meeting statistics and meeting quality analysis.

The software components of UMF are depicted in Fig. 2, and described in detail in the rest of this section.

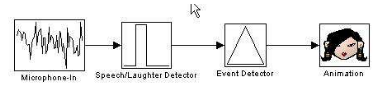

figure 2. UMF software components, consisting of speech/laughter detection, conversation behavior detection and avatar-based animation.

Speech and Laughter Detection

The low-level signal processing algorithms perform simple detection and classification operations. The system breaks down a raw audio stream into short segments, and classifies each segment into a category (silence vs. utterance and speech vs. laughter). The classified segments are then fed to the event detection mechanism to detect high-level events such as turn-taking and interruptions. The low-level signal processing and classification algorithms that we implemented in the UMF prototype were built on the state of the art in speech [7] and laughter detection [5], and were chosen for simplicity and speed. The algorithms work in real-time, enabling the overall system to give real-time feedback to meeting participants.

The UMF system handles crosstalk, which arises when one person speaks the signal is detected by all the nearby microphones. To suppress the utterance detection from crosstalk, we correlate the time-series power across different microphones. If a microphone has higher power than all its correlates, it is considered the "primary channel", and its detection suppresses the detections from all other channels with sufficiently high correlation coefficients. In this way, we ensure that each speaker is detected by only one microphone.

The UMF system distinguishes laughter from speech. This classification is especially important for distinguishing impolite interruptions, which should be discouraged, from collegial laugher, which should be encouraged (at least up to a point). For speech/laughter classification, we used features proposed in [5], along with some newly devised features. A number of features are extracted from the audio segments, such as the first and sixth cepstrum coefficients, which measure the relative energies in speech formants (stripes in the spectrogram Fig. 3).

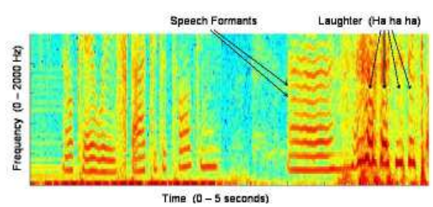

figure 3. Spectrogram showing distinct features of speech and laughter.

One of our newly devised features measures time periodicity, because laughter is often characterized by "hah-hah-hah-hah" sounds with a period of 100-300 msec. Another new feature simultaneously measures both time and frequency periodicity (a checkerboard pattern in the spectrogram), because each "hah" resembles a voiced sound, with strong formants. We used quadratic discriminant analysis (QDA) to combine these features into a binary classifier for distinguishing speech and laughter. The classifier's performance (approximately 80% correct classifications) is comparable to that reported [5].

Conversation Behavior Specification and Detection

We have used a formal specification approach for meeting behaviors and designed and implemented timed-automata-based detection algorithms. A speech sequence *Si* from a channel *i* is a mapping from time to a tri-state: 0, 1, or 2 where 0 means silence, 1 means normal speech, and 2 means laughter. Given speech sequences from two channels, properties of the sequence or the relationship between the two sequences can be formally specified. Some major conversation events are defined as follows (Fig. 4).

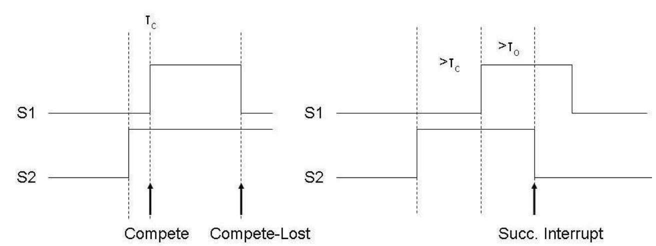

figure 4. Formal event specification: Left - compete, Right - successful interrupt.

Competes are defined for two sequences *S1* and *S2*, as occasions when two individuals try to speak at the same time.

Interrupts are defined as occasions when two segments *S1* and *S2* overlap for an extended period of time. There are two possibilities: successful interrupt and unsuccessful interrupt.

Each specification corresponds to a timed-automaton in which the time entering and leaving a state is used in conditions of state transitions.

Avatar-based GUI
The main display is a clock face with a maximum of 12 avatars, each representing a participant (Fig. 5). There is an extendable set of different styles of avatars (ours are designed using [4]) to choose from for a meeting, and each participant can choose his/her avatars. Each avatar has four types of faces: normal, talking, laughter and frown. The size of an avatar represents the relative value of the exponentially weighted moving average. The hand of the clock points to the current floor holder. A face is talking or laughing if its corresponding participant is talking or laughing, respectively. A face frowns if its corresponding participant was successfully interrupted by someone or lost a compete to speak; a hat displayed on the head of a face indicates the seriousness of offensive behaviors, i.e., successfully interrupting others many times. There are currently two levels of the repeated interruption indicator, represented by two types of hats. The level of interruption fades with time, and hats change or disappear as the level falls below its threshold. The background of the clock has five colors: gray, red, blue, green and white. Gray represents one speaker at a time, red means there were overlapping conversations, blue means there are people being interrupted or lost in a compete, green means there is group laughter, and white means there is an overly long period of silence. In practice, laughter played an important role in the meetings we studied when designing the system. Laughter frequently occurs concurrently with speech, but must be prevented from counting as an interruption or change of turn. In cases of group laughter, where multiple people laugh nearly simultaneously, we consider this to be a positive sign of group cohesion, and these group laughs are counted and indicated in our UI by a ``smiley" token that is added to a tally at the top of the screen. Meetings with too few of these tokens may be overly serious and one could hypothesize that effort to lighten up the meeting might increase creativity.

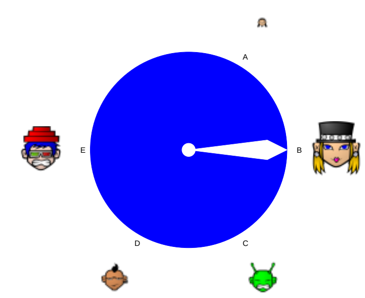

figure 5. UMF Avatar-based GUI: head size represents the level of participation, red (E) and black (B) hats represent level-1 and level-2 aggressive behaviors, respectively. Frown faces (C and E) indicate lost competition or being interrupted. The white arrow indicates the current floor holder.

Meeting Quality Analysis

Events such as turns, race/competition, overlaps and interrupts are detected and recorded over time. Meeting quality metrics are calculated based on statistics of events. Number of turns, total length of turns, and average length of turns are obtained for each participant. Interactive behaviors, such as interrupts and competes, are represented by interaction matrices.

Conclusion

UMF is a meeting facilitator with a playful GUI presenting a wide variety of conversational metrics. The idea is to encouraging good meeting behaviors such as turn taking and laughter, and discouraging domination and shyness. UMF does not only provide real time feedback, but can also be used to analyze meetings offline.

References

[1] S. Barsade. The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior Administrative. *Science Quarterly*, 47:644--675, 2002.

[2] T. Bergstrom and K. Karahalios. Conversation Clock: Visualizing Audio Patterns in Co-located Groups. In *Proceedings of the 40th Annual Hawaii international Conference on System Sciences*, January

[3] J. M. DiMicco, A. Pandolfo, and W. Bender. Influencing Group Participation with a Shared Display. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, November 2004.

[4] Dookyweb. Avatar creation, 2007. http://www.buscarmessenger.com/avatars.swf.

[5] Lyndon S. Kennedy and Daniel Ellis. Laugher Detection in Meetings. In *Proceedings of NIST ICASSP04*, pages 118--121, Montreal, Canada, May 2004.

[6] Taemin Kim, Agnes Chang, Lindsey Holland, and Alex Pentland. Meeting Mediator: Enhancing Group Collaboration using Sociometric Feedback. In *Proceedings of ACM Computer Supported Cooperative Work (CSCW08)*, 2008.

[7] Lie Lu, Hao Jiang, and HongJiang Zhang. A Robust Audio Classification and Segmentation Method. In *Proceedings of the 9th ACM International Conference on Multimedia (MM01)*, pages 203--211, 2001.

[8] J. Niekrasz and P. Ehlen. Multimodal Meeting Capture and Understanding with the CALO Meeting Assistant Demo. In *4th Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI)*, June 2007.

[9] Steve Renals, Thomas Hain, and Herv'e Bourlard.

Recognition and Understanding of Meetings The AMI and AMIDA Projects. In *Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU'07*, Kyoto, 9 December 2007.

[10] H. Sacks, E. A. Schegloff, and Gail Jefferson. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696--735, 1974.

[11] H. Wactlar, S. Stevens, A. Hauptmann, M. Christel, and A.J. Bharucha. A System of Video Information Capture, Indexing and Retrieval for Interpreting Human Activity. In *IEEE International Symposium on Image and Signal Processing and Analysis (ISPA'03)*, Special Session on System Perspectives in Information Retrieval, September 2003.