Second Messenger: Increasing the Visibility of Minority Viewpoints with a Face-to-face Collaboration Tool

Joan Morris DiMicco, Walter Bender
MIT Media Lab
20 Ames St
Cambridge, MA 02139 USA
+1 617 253 9326
{joanie, walter}@media.mit.edu

ABSTRACT

This paper introduces the application Second Messenger, a tool for supporting face-to-face meetings and discussions. Second Messenger uses a speech-recognition engine as an input method and outputs filtered keywords from the group's conversation onto an interactive display. The goal of this interface is to improve the quality of a group discussion by increasing the visibility of diverse viewpoints.

Categories and Subject Descriptors

H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces -- *computer-supported cooperative work, synchronous interaction, theory and models.*

General Terms

Human Factors.

Keywords

CSCW, intelligent assistance, speech processing, face-to-face conversations.

1. INTRODUCTION

Second Messenger is an application that augments a face-to-face discussion by providing a display of the verbal comments made during the discussion. By observing the communication between individuals, Second Messenger uses a combination of speech-recognition technology and semantic analysis to display a real-time text summary of the group's comments. The content on the shared display can be manipulated with a pointing device, making Second Messenger a real-time meeting tool that can be used by the group to organize their discussion.

The core goal of Second Messenger though is to influence the way a group communicates in a face-to-face setting. Extensive research in social psychology and organizational behavior, reviewed in [4], has shown that increasing the amount of discussion around minority viewpoints can increase the quality of group decisions. Second Messenger attempts to bring about this through emphasizing the minority viewpoints in the group by selectively filtering *whose* keywords appear on the display. By increasing the visibility of group members who speak less frequently, and filtering out the comments of group members who are verbally dominating, the application attempts to amplify the voice of the group's minority members.

IUI '04, Jan. 13-16, 2004, Madeira, Funchal, Portugal.
ACM 1-58113-815-6/04/0001.

2. USAGE SCENARIO

Imagine the following scenario. While a research group discusses an upcoming potluck dinner, the Second Messenger display shows the comments of different people slowly falling down the screen: "cashew chicken," "beef stew," "salad," "fruit salad." The group uses the display to organize the list of food items according to entrée type to see what items are missing. Conveniently, side comments from the group that don't relate to the topic of food do not get displayed. When one quiet member of the group says, "Umm, I have some food allergies. I'm allergic to nuts," rather than restricting the displayed phrase to "food, nuts," Second Messenger displays "food allergies, allergic nuts." By both hearing and seeing this new information, instead of overlooking or ignoring the quiet comment, the group turns the conversation to the topic of what foods he is allergic to. (See Figures 1 and 2 for screenshots of the display and client applets.)

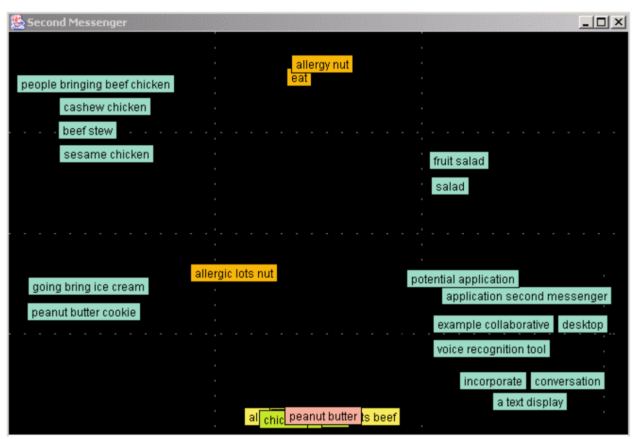

Figure 1: Second Messenger's shared display. The spoken phrases fall down the center of the screen. Users can move the phrases around the screen to organize their ideas. Untouched phrases collect at the bottom of the display.

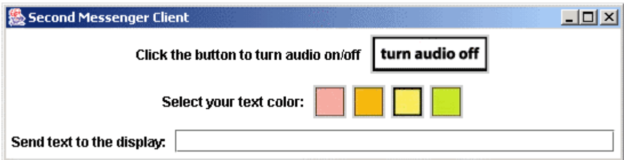

Figure 2: Second Messenger's client. Users can turn their audio feed on/off and choose the color of their transcribed text. They also have the option to send text directly to the display.

3. SECOND MESSENGER

The following sections explain how the application currently works: first assembling a transcript of the conversation; and then filtering it according to semantic and social facilitation goals. Figure 3 provides an overview of this process.

3.1 Conversation Transcription

Today's speech-recognition technology is far from ideal, creating a significant challenge to deploying an application that detects informal, non-dictation speech. To achieve the highest accuracy possible and to eliminate the challenge of speaker identification, we've required that each user of Second Messenger wear an individual microphone that sends his/her speech to an individually-trained speech model. We use IBM's ViaVoice engine to convert each speaker's audio to text and then stream this transcript to a central server for processing.

Second Messenger does not need a complete transcript of a conversation, but rather a collection of the relevant keywords. Thus we perform part-of-speech filtering on each incoming utterance to limit the text down to nouns, verbs, and adjectives. This type of filtering removes the non-verbal utterances (ah's and um's) and many of the uninteresting comments, such as simple affirmations and confirmations.

3.2 Semantic Filtering

The next step is to determine which text to display to the group. Our opinion is that the most interesting comments are the ones that haven't been said before, yet have some relation to previous comments made by others. Therefore when we perform a semantic analysis of the phrases using WordNet, we calculate how close the current phrase is to prior phrases spoken by *other* members of the group. Specifically, we locate mutual conceptual parents of the spoken words. For example, after semantic filtering, the phrase "I saw some good peanut butter cookies at Star market yesterday" becomes "peanut butter cookies" if other people have previously used words connected to the parent concept "food." Our objective is to encourage contributions from everyone in the group, therefore the emphasis here is to display the ideas that relate to each other that originated from different people, preventing one person from controlling which text appears on the display.

3.3 Social Facilitation

As stated in the introduction, the primary goal of this project is to encourage equal participation in a meeting for the purpose of increasing the diversity of the discussion, and ultimately the quality of the decision-making. Utilizing the fact that the system knows exactly how many words each user has spoken, Second Messenger is able to calculate if someone is under- or over-contributing to the conversation. Imagine a meeting scenario as described earlier where most of the group is able to state their opinions, yet one member contributes disproportionately less. When Second Messenger detects this type of imbalance in the verbal contributions, rather than displaying the spoken text with the same imbalance in contributions, it visually emphasizes the less-vocal member's contributions. When someone is contributing one-fourth the amount of the other group members, their text phrases are no longer filtered through the semantic filter, so that all of their spoken nouns, verbs, and adjectives are displayed. If someone is dominating the discussion, by saying twice as much as all other group members combined, they will no longer have any of their words displayed on the screen. These contribution ratios were chosen as a baseline and we plan to refine them as we observe more users. We hypothesize that this type of filtering will make the less-vocal members' contributions more apparent and the dominating individuals seek out others' ideas and opinions.

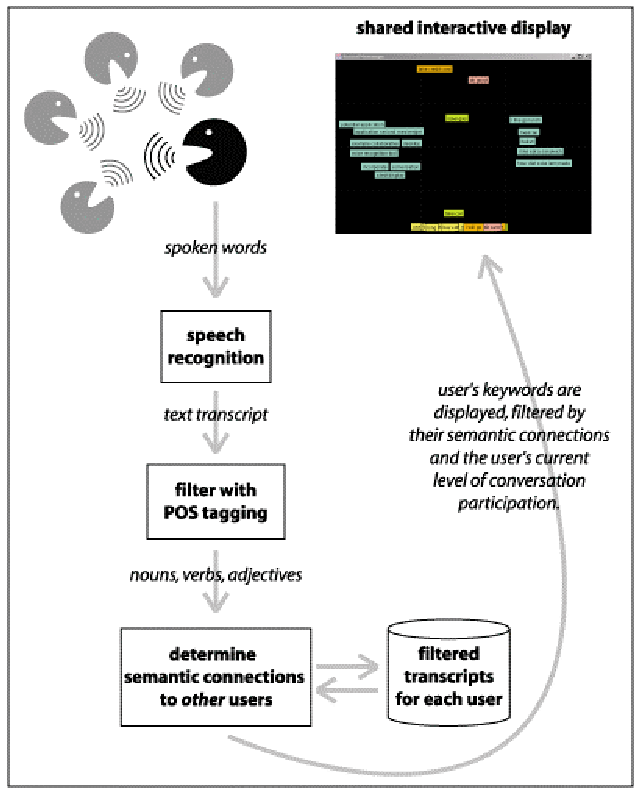

Figure 3: Second Messenger's architecture.

4. RELATED WORK

While there has been a great deal of prior work on face-to-face meeting support, beginning with the *Colab* project [6], there has been relatively little work using verbal content as input to a shared workspace. One project that strives towards this was built by Jebara, et al., [2] where their system performed real-time topic spotting by comparing the voice-recognition transcript to a fixed corpora derived from newsgroups. During a face-to-face conversation, the current topic of conversation (limited to the corpora) was displayed on a shared screen.

Eagle, et al., [1] further developed Jebera's work on topic spotting by using the Open Mind Common Sense database [5]. With OMCS they located the 'common sense' connections between spoken words to eliminate errors in voice-recognition and provide a topic-level summary of a conversation. While this was not a real-time support tool, it offers a promising method for analyzing speech-recognizer output.

Another interesting meeting-support tool using voice recognition was developed by Kristjansson, et al. [3]. Their application augmented an audio recording of a meeting with an outline of the meeting's structure derived from a discourse analysis of the conversation. Like the previous example, this application was not real-time, but it demonstrates how discourse analysis techniques can be applied to a fuzzy dataset to derive meeting structure.

5. CURRENT PROGRESS

To date, we have deployed the system as described and used it informally during several of our weekly research meetings. Through this process, we have confronted the reality that poor voice-recognition results can distort Second Messenger's text output. To prevent this, we have started to filter ViaVoice's output based on the "phrase score" returned for each word. Although IBM states that the phrase score is not a confidence value, Vemuri, et al. [7] demonstrated that a word's phrase score is a good predictor of word accuracy and we have been pleased with the results thus far. By setting the phrase score threshold quite high, we can ensure that the words sent through the semantic filter were actually spoken.

Currently, we are preparing to run more formal behavioral experiments to test our hypothesis that using Second Messenger during a face-to-face meeting will result in more equal participation amongst the group members.

6. FUTURE DIRECTIONS

We believe there is great potential for applications that influence face-to-face interactions and there are many directions to expand this research. Our future plans include improving Second Messenger's concept of conversational topics and individual intention.

Currently the system does not have a concept of conversation topic, yet with this knowledge it could track topic changes and who contributed which ideas within each topic. While limiting the influence of dominate members may help less dominant people to contribute, we would prefer if the system would reward new *ideas* rather than new speakers. One method to detect and track topics would be to incorporate a common-sense data source into the analysis of the text, as done by [1]. As an aspect of topic tracking, the system could also be improved with a modeling of time and a sliding window of observation to give new ideas more weight.

Another direction we would like to explore is incorporating feedback other than voice into the system. During a meeting there is an abundance of information found in what people look at, the objects they manipulate, and the text they write down. By capturing these observations, Second Messenger could augment its model of topics and individual contributions to build an understanding of the intentions of the group. With a model of individual versus shared intentions, Second Messenger could further assist the group in exploring new ideas and unique approaches to problem solving.

7. CONCLUSION

For technology to support more productive meetings and face-to-face discussions, a first step is to design technology that understands what is going on during the meeting. Towards this, Second Messenger utilizes speech-recognition and semantic-analysis techniques to gather and then display the conversational keywords onto a shared workspace. Going one step further, Second Messenger utilizes its knowledge of who said what to assist in finding the relevant keywords and to emphasize the contributions of "minority" group members. In this manner, this application attempts to equalize the contributions of individuals during a meeting in order to increase the diversity of ideas discussed and increase the quality of the decisions made.

8. ACKNOWLEDGMENTS

This work was supported in part by the information: organized research consortium at the MIT Media Lab.

9. REFERENCES

[1] Eagle, N., Singh, P. and Pentland, A. (2003): 'Common Sense Conversations: Understanding Casual Conversation using a Common Sense Database,' in *Proceedings of the Artificial Intelligence, Information Access, and Mobile Computing Workshop at IJCAI*, Acapulco, Mexico, August 2003.

[2] Jebara, T., Ivanov, Y., Rahimi, A. and Pentland, A. (2000): 'Tracking Conversational Context for Machine Mediation of Human Discourse,' in *Proceedings of the AAAI Fall 2000 Symposium*, August 2000.

[3] Kristjansson, T., Ramesh, P., Huang, T.S. and Juang, B.H. (1999): 'A Unified Structure-Based Framework for Indexing and Gisting of Meetings,' in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMCS)*, Florence, Italy, pp. 572-577.

[4] Nemeth, C.J. and Staw, B.M. (1989): 'The tradeoffs of social control and innovation in groups and organizations,' *Advances in Experimental Social Psychology*, vol. 22.

[5] Singh, P. (2003): Open Mind Common Sense, MIT, http://commonsense.media.mit.edu.

[6] Stefik, M., Foster, G., Bobrow, D., Kahn, K., Lanning, S. and Suchman, L. (1987): 'Beyond the chalkboard: computer support for collaboration and problem solving in meetings,' *Communications of the ACM*, vol. 30, no. 1, pp. 32-47.

[7] Vemuri, S., DeCamp, P., Bender, W. and Schmandt, C. (2004): 'Improving Speech Playback Using Time-Compression and Speech Recognition,' in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004)*, Vienna, Austria, April 2004.