GroupMedia - Using Wearable Devices to Understand Social Context

Anmol Madan, Ron Caneel, Alex "Sandy" Pentland

Massachusetts Institute of Technology
MIT Media Laboratory
20 Ames Street, Cambridge 02139

{anmol,rcaneel,sandy} @ media.mit.edu

1. Introduction

The gap between wearable technologies in the research laboratory and commercial devices has been reducing rapidly. Given the current convergence between hand-helds, cell phones, and wearable computers, and the pervasive presence of these computationally powerful devices, we believe that we can extend their utilization by building applications that are aware of their user's *group context.*

This paper explains how we use speech features and biometric data for a dynamic group, to quantify conversational interest and improve group dynamics within social settings.

2. Motivations and Problem Statement

The MiThril wearable system [1] is an elaborate wearable device, which has much of the widespread ease-of-use characteristics of PDAs and cell phones. However, it has substantial untapped potential when it is extended into a group-centric system wherein multi-user applications can take advantage of the individual entities. The form of this interaction can be as commonplace as an audio groupware application for conversation on networked PDAs, to a full-fledged physiology-based coordination system for distance-separated teams.

A group-centric wearable system with real-time audio feature extraction and processing capabilities can capture communication patterns and styles between group members during a meeting or conversation. Based on the audio classifier we can calculate group interaction statistics such as speaking time, speaking frequency, and interruption behavior. A simple application of this idea is to augment a group meeting with a public display showing the speaking history of each member. Morris DiMicco [6] could show that such a public display resulted in more equal speaking times, enhanced group information sharing, and group satisfaction.

Our first design principle is to use aggregate group statistics, like overall voice energy, head nodding and speech timing characteristics to estimate the level of interest and involvement within the group. This can be relayed back to the participants via a public display and might also be used by outsiders to determine if they want to join the conversation. We use a GUI application *RateIt!* on Zaurus handhelds, in order to be able to correlate explicit interest ratings with group behavior, and thus derive an automatic estimate of group interest.

A problem often found in group interactions is negative influences of intra-group dynamics. Particularly, if emotions are going high in a discussion people tend to forget how they interact, and this behavior is often problematic. Our second design principle is that by sending each participant a) his/hers personal interaction pattern and b) the overall group interaction behavior, we can improve the interaction pattern. For example if an individual can discover that he/she interrupts other people much more often than the average, or that he/she speaks more than everybody else and now can adjust his/her behavior, in real-time.

Each member can also specify some basic behavior limits and will be alerted via PDA display, audio cues or vibrato-tactile feedback if he/she deviates too much from their preferred behavior. Such reminders of out-of-range behavior can help prevent undesired behavior and thus improve the effectiveness of group communication.

3.1 GroupMedia Software and Hardware Overview

The MiThril wearable system is a combination of commercial off-the-shelf hardware with the flexibility offered by in-house engineering. The system software and architecture is can be scalable from a full-fledged system to just a PDA with audio support. These gave us the advantage of being able to rapidly prototype and deploy applications; for example, up to forty of these systems have been used in classroom applications [9].

Fig 1. GroupMedia - MiThril system featuring a Zaurus PDA, SAK 2, accelerometer [on hat] and IR tag reader packaged in an easy-carry holster.

As shown above, the system core is the Sharp Zaurus PDA, which runs Linux on an Intel Strong ARM processor [Xscale processor for the SL 5600]. The system uses the Swiss Army Knife, version 2.0 [SAK2] sensor acquisition board, to connect to various I2C and analog sensors. The SAK2 also features a low-power 2.4 Ghz transceiver to communicate with wireless sensors, and a compact flash storage for sensor data in case of loss of connectivity.

The various sensors that have been connected with the system via the SAK2 include accelerometers [Analog Devices ADXL, 3-axis], IR tags [face-to-face identification and overall location], electrocardiogram [EKG], galvanic skin response [GSR] and temperature. MiThril software implements the Enchantment whiteboard and signaling system [3], a light weight low cost means of routing information transparently across processes or distributed devices. The MiThril system is described in more detail here [1]. The GroupMedia systems use Sqlite on the Zauri to build a long-term history of interaction, for example over a few days. Information is accepted and displayed over the Zaurus PDA interface using Qtopia Graphical User Interface.

3.2 Dynamic Group Formation

Fig 2. Forming a dynamic group using ad hoc wireless

The ability to form dynamic groups, using these wearable devices, is key to understanding social context. This can be done in various ways on our system, by using infra-red beacons [on MiThril], Bluetooth [Class 2, compatible with our system], 2.4 Ghz wireless networking [Nordic chip on-board the SAK2], etc. with varying levels of resolution and location inference. We have used ad hoc 802.11 wireless networking, to be able to rapidly prototype applications for a group within a room, we use the security provided within TCP/IP and SSL to deal with security issues. By tying this implementation with our Enchantment multi-user communications architecture [4], we are able to leverage its security provisions and flexibility as shown in the figure above; the system can scale

easily from client-server to peer-to-peer architectures, over infrastructure or ad hoc wireless networks.

3.3 Machine Learning

We use machine learning methods to extract meaningful features from the audio signals and bio-signals such as head-nodding from accelerometers or `interest' spikes from GSR sensors, all in real-time.

The real time audio classifier is based on the algorithm introduced by Basu [5], and uses two hidden Markov models (HMMs). The first HMM determines whether a frame is voiced or unvoiced, by extracting the maximal peak value, number of peaks and spectral entropy, for each 32 ms frame from the audio signal. The second HMM uses voiced segments to determine if a person is actually speaking or not. Due to the processing limits of the Sharp Zaurus we use a fixed-point-integer Fast Fourier Transform [FFT] for feature extraction. This audio classification system allows us to accurately detect speaking and not-speaking transitions even in a noisy environment.

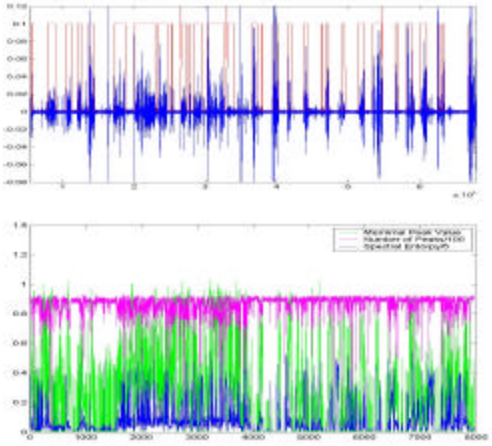
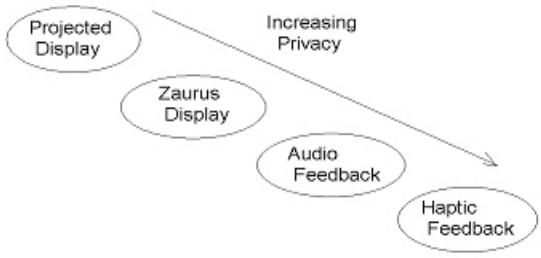


Fig 3. Audio signal and speaking / not-speaking classification [top] and corresponding real-time features [bottom]

The above figure shows the speech classifier results on an audio signal, which is based on the features represented below it. Since we implement real-time speech feature analysis and classification, we do not need to record audio signals to understand social context and group interest. Hence, we address possible privacy concerns with regard to audio recording.

The head-nodding classifier is a gaussian mixture model, with 31-dimensional multi-variate gaussian components based on the frequency-domain representation of accelerometer signals [2]. The system currently can accurately detect [vertical] head nodding by the typical frequency of movement. The real-time classification engine computes the FFT on the PDA and identifies the correct class based on the mixture model. The model parameters are calculated by Estimation Maximization offline in MATLAB and are fed into the real-time mixture model classifier.

3.4 Real-Time Feedback



In order to give real-time feedback to members of the group, we are the following different methods,

- A projected display with all or selected group statistics
- Individual information sent to the PDA display of every person.
- Audio cues over headphones to inform people
- Vibrato-tactile feedback using pancake motor mounted at the shoulder [6], outside the direct cone of attention.

4 Understanding Social Context

The GroupMedia system to understand conversational interest has been deployed in ten preliminary sessions so far, which have involved the same group of four to eight people, each of about 1-2 hours duration.

Initial user studies with this system indicate that the speech characteristics [speaking time, interrupting] and head nodding can be used to understand and improve the dynamics of the group. By the time of this workshop, we hope to have results linking conversational interest with the various features.

Using the *RateIt!* application we expect to be able to extend this capability further by combining an audio classifier, a head-nodding classifier, speech features, GSR, and a GUI interface for expressing opinion, in order to build an automated, person-independent mechanism for estimating conversational interest in an ad hoc group.

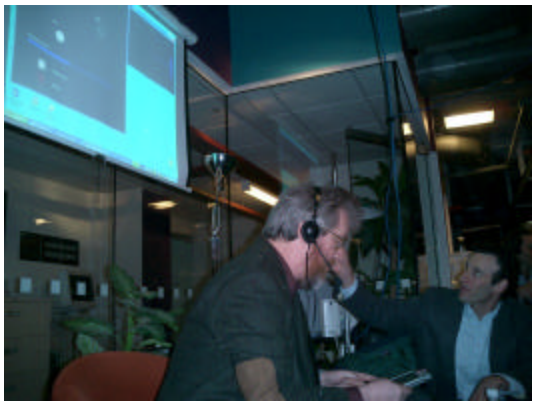


Fig 4. RateIt! Application and group context visualisation

5. Conclusions and Opportunities

We believe that a real-time system of this sort can radically affect social interaction, group communication effectiveness, and coordination. Some examples are,

- Audio characteristics and body language can reflect the mood of the social situation [8], and by evaluating them for a varying group, we can begin to understand group context as a whole. As any good marketing person would know, if you could tell in real-time that the audience was not really interested in your pitch, you could adapt your style.
- By looking at speech characteristics and better understanding the group dynamics in real-time in meetings situation, participants can improve communication effectiveness.
- By combining this system with skin conductivity, which is a very good indicator of arousal and attention [7], we hope to improve the accuracy of this system.

References

[1] R.W. DeVaul, M. Sung, J. Gips, S. Pentland, "MIThril 2003: Applications and Architecture", Proc. ISWC '03, White Plains, N.Y., October 2003, pp 4- 11.

[2] R. W. DeVaul and S. Pentland, "The MIThril Real-Time Context Engine and Activity Classification", Technical Report, MIT Media Lab, 2003

[3] E. Keyes, Enchantment White paper, MIT Media Laboratory

[4] S. Basu. "Conversation Scene Analysis", in Dept. of Electrical Engineering and Computer science. Doctoral. 2002, MIT.

[5] Morris DiMicco, J., "Designing Interfaces that Influence Group Processes". Doctoral Consortium Proceedings of the Conference on Human Factors in Computer Systems (CHI 2004), Vienna, Austria, April 2004.

[6] A. Toney, L. Dunne, " A Shoulder pad insert for Vibrotactile Display", Proc ISWC '03, White Plains, N.Y., October 2003.

[7] R.W. Picard, "Affective Computing", MIT Press, Cambridge

[8] G. Tom, P. Petterson, et al "The Role of Overt Head Movement in the formation of Affect", Basic and Applied Psychology, 1991

[9] M. Sung, et. al. "MIT.EDU: System Architecture for Distributed Real World Application in Classroom Settings"