GroupMedia: Distributed Multi-modal Interfaces

Anmol Madan
Ron Caneel
Alex "Sandy" Pentland
Massachusetts Institute of Technology
MIT Media Laboratory
20 Ames Street, Cambridge 02139
+1 617 258 5956
anmol@media.mit.edu
+1 617 253 4662
rcaneel@media.mit.edu
+1 617 253 3818
sandy@media.mit.edu

ABSTRACT
In this paper, we describe the GroupMedia system, which uses wireless wearable computers to measure audio features, head-movement, and galvanic skin response (GSR) for dyads and groups of interacting people. These group sensor measurements are then used to build a real-time *group interest index*. The group interest index can be used to control group displays, annotate the group discussion for later retrieval, and even to modulate and guide the group discussion itself. We explore three different situations where this system has been introduced, and report experimental results.

Categories and Subject Descriptors
J.4 [**Computer Applications**]: Social and Behavioral Science – *Psychology, Sociology*

General Terms
Measurement, Experimentation and Human Factors.

Keywords
Human behavior, interest, speech features, galvanic skin response, head nodding, prosody, influence model

1. INTRODUCTION AND MOTIVATIONS
Cell phones and PDAs are converging, and the pervasive presence of these computationally powerful devices makes likely they will become increasingly important interfaces for human-computer interaction. They are also evolving into powerful wearable computers, making it possible to implement *distributed* multi-modal applications with commodity hardware. We believe that we can improve the usefulness of these small wearable devices by building applications that harness their distributed sensing and computing power to create systems that are aware of their user's *group context.*

ICMI '04, October 13–15, 2004, State College, Pennsylvania, USA.

Sociologists have developed powerful theories of group behavior and interaction [1,18] that could potentially be used to build socially aware devices. However these theories are generally not quantitative and are not grounded in machine perception capabilities. We believe that by using a cell phone/PDA's ability for the wearable, unobtrusive sensing of group context, real-time processing, and wireless communications, there is a tremendous opportunity to produce a quantitative *social intelligence* for these ubiquitous devices. Such social intelligence could greatly improve the usability, flexibility and scalability of wearable devices.

Eagle and Pentland [10] showed how interest ratings could be associated with group speaking patterns to analyze the dynamics of the group and annotate `interesting' comments. In this work participants used a PDA to continuously rate their interest in the conversation, and these interest ratings were summed into a group interest rating.

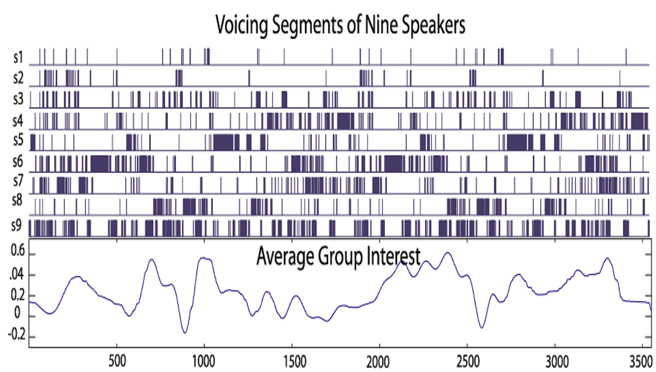

Figure 1. Voicing segments for a one-hour class mapped above the aggregate interest level

Figure 1 shows a typical data set collected in a classroom setting. From these basic features profiles of a student's typical classroom behavior are built over time, using conversation features such as speaking rate, energy, duration, participants, interruptions, transition probabilities, and time spent holding the floor.

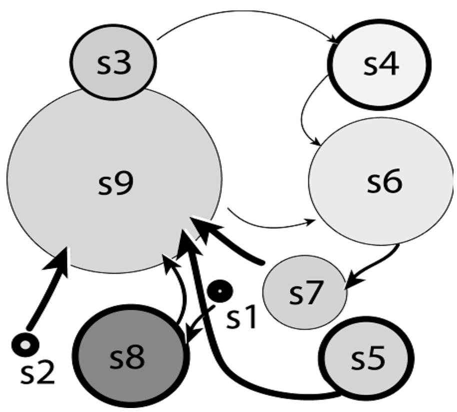

Figure 2: Metrics for classroom interaction analysis and visualization of the turn-taking dynamics between the professor (s9) and students. Speaking Time = Circle size, Transition Probability =Width of the link, Average Interest Level = Circle color (individual) and Circle border (group)

In this manner the system uncovers information concerning the dynamics of the class. The information collected can include a list of the peers that a student typically sits by, avoids, talks to, interrupts, and how the conversation transitions. As can be seen from Figure 2, a professor (s9) is obviously the dominant member while his advisees (s2, s7, s8) concede the floor to him with relatively high probability - indicative of his influence.

Subjective feedback can be pooled and shared with the participants via a public display, e.g., real-time graphics illustrating participant information (as shown in Figure 2) are projected onto a screen using an LCD projector. A simple version of this sort of participant feedback (incorporating only percent speaking time) has been evaluated by Morris DiMicco [17] and found to significantly improve the objective performance of a four-person group engaged in a typical problem-solving task. This experiment demonstrates that the ability to capture real-time, rich data about group behavior can be used to improve the dynamics of the group's interaction.

The main limitation of this previous work is that participants had to provide continuous interest ratings during the discussion. This sometimes distracted them from the discussion, and at other times (unfortunately often the most interesting times) the participants got so `wrapped up' in the discussion that they forgot to provide interest ratings. Another limitation of this previous work is that the only machine perception was to determine speaking vs. not speaking. No attempt was made to measure vocal features such as stress, prosody or `body language'.

In this paper we describe a distributed, wearable sensing and computation system that can make real-time measurement of a wide range of vocal features, aspects of body language, and physiological signals. We are using this system to develop a behavior-based *group interest index*, by comparing the sensor data with user-generated interest ratings across a variety of situations.

The organization of this paper is first to describe our distributed sensing and computation platform, then our main sensing modalities, and finally our experiments in modeling dyadic and group interest.

2. THE GROUPMEDIA SYSTEM

Our system extends the MiThril wearable computer, a proven accessible architecture that combines inexpensive commodity hardware, a flexible sensor/peripheral interconnection bus, and a powerful light-weight distributed sensing, classification, and inter-process communications software layer to facilitate the development of distributed real-time multimode and context-aware applications [9]. Our recent efforts have been towards extending MiThril to a multi-user, group-centric framework.

The major components of the MIThril-GroupMedia architecture are: a PDA-centric hardware configuration, the Enchantment software network and resource discovery API, the real-time machine learning inference infrastructure, and secure ad-hoc networking over 802.11 wireless. Each of these components is briefly described below.

2.1 Hardware

The GroupMedia system uses the Zaurus SL series PDA-based mobile device. This system allows applications requiring real-time data analysis, peer-to-peer wireless networking, full-duplex audio, bluetooth wireless communication, local data storage, graphical interaction, and keyboard/touch-screen input.

The system uses the Hoarder [12] sensor hub, to connect to an accelerometer (triaxial Analog Devices ADXL) and GSR leads. The GSR analog signal is amplified using a daughter biometric board for the Hoarder, before digitization. The daughter biometric board also supports measurement of heart rate [using a serial Polar chest heart-rate monitor] and EKG. The Hoarder board can also function independently as a compact flash based data acquisition platform.

Ergonomic considerations for certain experiments (like speed-dating) require wireless sensing and a smaller form-factor, which led us to use BodyMedia physiology measurement devices and bluetooth accelerometers[7]. We are also testing Motorola A760 Linux cell phones and bluetooth GSR sensors [22] to create the next-generation platform for this research.

Figure 3. The GroupMedia system featuring a Zaurus PDA, Hoarder sensor board, bluetooth accelerometer [on hat], GSR leads and a microphone

2.2 Software and Networking

The Enchantment whiteboard and signaling software [15] is a light weight low cost means of routing information transparently across processes or distributed devices, and allows for easy routing and storing of raw signals and machine-learning classification results. Enchantment is intended to act as a streaming database, capturing the current state of a system, person or and on modest embedded hardware can support many simultaneous clients distributed across a network and hundreds of updates a second. Information transfer has SSL security and can be used by the real-time context classification engine or saved for offline analysis. The GroupMedia systems are also capable of using SQLite on the Zaurus to build a long-term history of interaction, for example, over a few days, on a SD card that can store up to a gigabyte of data.

To enable a mobile multi-modal system, we need to build elements of dynamic social context by proximity detection of other users, implemented using 802.11 wireless in ad-hoc networking mode. The system can scale easily from client-server to peer-to-peer architectures, over infrastructure or ad-hoc wireless networking.

2.3 Context Classification

GroupMedia context classification consists of real-time speech feature analysis using hidden markov modeling and head-nodding detection from accelerometer signals using gaussian mixture modeling. The important design features of the system are simplicity, modularity, flexibility, and implementability under tight resource constraints. The MiThril Inference Engine [8] used for head-nodding detection, abstracts the data analysis into distinct steps, including the transformation of raw sensor data into features more suitable for the particular modeling task, the implementation of statistical and hierarchical, time-dependent models that can be used to classify a feature signal in real time, and the development of Bayesian inference systems which can use the model outputs for complex interpretation and decision-making.

3. FEATURE SET AND ANALYSIS

To form our group social index we combine three types of sensor measurements from every individual in the group, and then use machine learning techniques to compare these sensor measurements to the average interest expressed by each of the subjects using a hand-held `rate-it' interest meter or post-session feedback dialog. The sensor measurements we are currently using are a selection of speech features, head-movements and nodding, and GSR. By using these features we attempt to capture elements of non-verbal communication like body language, prosody, empathy, mirroring behavior and physiology, representative of interest in an activity or person.

3.1 Speech Features

The system calculates speech features (devoid of any conversational content) of two different types - individual features and the group features. The individual features are extracted solely from the audio signal of one person, and represent the overall speaking style of the individual for the measured duration. Group features are calculated collectively and represent the dynamics of interaction between individuals.

3.1.1 Individual features

The individual features include fraction of speaking time, standard deviation of energy, standard deviation of formant frequency, standard deviation of spectral entropy (both associated with stress measurement) and the voicing rate, and we use them as measures of speaker prosody and emphasis (although we calculate a few others like MFCC, they are not used in our analysis). We use a multi-level HMM structure to classify the voiced/non-voiced sections and hence the speaking /not-speaking regions of the audio stream of every individual. The algorithm was introduced by Basu [2] and is particularly robust even in settings with background noise.

The speech feature extraction has been implemented in real-time on the Sharp Zaurus PDA. This allows us to stream low bandwidth speech classifier results, rather than high-bandwidth audio signals, over a wireless network. This proves to be extremely valuable when we scale the experiment to a larger number of users. Due to the processing limits of the Sharp Zaurus we use a fixed-point-integer FFT for real-time feature extraction.

3.1.2 Group features

From the basic speech features we next measure the dynamics of the individuals and their interactions. We look at the influence parameters and the number of short back-and-fourth exchanges.

The ability to learn of a dynamic model and the interpretability of such a model greatly depends on its parameterization. The requirement for a minimal parameterization has motivated our development of Coupled Hidden Markov Models (CHMMs) to describe interactions between two people, where the interaction parameters are limited to the inner products of the individual Markov chains [3]. This allows a simple parameterization in terms of the *influence* each person has on the other.

This influence parameter expresses then how strong the overall state for an actor A is depending on the state of actor B. In this case we use a simple two-state model of speaking vs. not-speaking to model individual dynamics, and then measure the influence parameter to determine the coupling between speakers. In Choudhury [6] this measure of influence was shown to have an extremely high correlation with one measure of the social novelty (and thus presumably the interestingness) of the information being presented.

We label short interjections of a time scale less than 1 second as back-and-forth exchanges (typically single words like ‘okay’, ‘aha’, ‘yup’ or sounds of approval), and consider them as mirroring behavior, where one person unconsciously mimics the other’s prosodic pattern. Our analysis further on describes their importance.

3.2 Galvanic Skin Response (GSR)

GSR has been used as a measurable parameter of a person's internal "state" [21]. In terms of physiology, GSR reflects sweat gland activity and changes in the sympathetic nervous system and measurement variables. It is measured as change in the relative conductance of a small electrical current between the electrodes placed at the fingertips [or the palm]. The activity of the sweat glands in response to sympathetic nervous stimulation (increased sympathetic activation) results in an increase in the level of conductance. According to [11], there is a relationship between sympathetic activity and emotional and cognitive arousal, although one cannot identify the specific emotion being elicited. Fear, anger, startle response, orienting response and sexual feelings are all among the emotions that may produce similar galvanic skin responses.

Studies in psychophysiology have used galvanic skin responses (or electro-dermal activity) as a physiological index for multiple individuals exposed to the same cognitive or emotional events. They have gone to the extent of looking at correlations in galvanic skin responses and using it as an objective measure of relationships between married couples or counseling empathy between patients and therapists during psychotherapy [16].

In our research, we attempt to capture this *group interest index,* by seeking similar GSR responses within the dynamic set of people. Although any one GSR spike by any one individual could have a large number of causes, we propose that GSR spikes experienced *by the entire group* at should be correlated to cognitive or emotional events that the entire group has experienced

With galvanic skin response signals, we are essentially seeking two types of information. Instantaneous spikes have been correlated with startle and strong emotional physiological responses [13]. On the other hand, longer duration GSR rise and fall trends have been related to relatively longer term changes in empathy [11] between individuals. The raw galvanic skin response signal is pre-filtered and smoothed using a median average to eliminate high-frequency noise in measuring equipment.

In order to measure the extent to which GSR responses in several people are correlated to cognitive or emotional events that they all experienced, we calculate the cross-correlation co-efficient on adjusted magnitude over a window of 5 seconds. This allows us to capture longer-term trends as well as short-term responses. We can isolate instantaneous peaks by differentiating the signals or doing matched filter correlations. To analyze long-term trends, we also correlate the slopes for the GSR signals.

3.3 Head Nodding

Across cultures head nodding has been observed as an indicator of agreement or understanding while a headshake has indicated a negative signal for disbelief or disapproval. Kapoor [14] mentions that head nodding also plays a role in conversional feedback where both the speaker and the listener nod synchronously, to keep the conversation flowing. At the most basic level, this behavior can be compared to the reptilian principle *isopraxism*, where one animal imitates the behavior of the other.

Another possible role of head nodding may be derived from the chameleon effect as described by Chartrand and Bargh [5]. They observed that people mimic body movements of their conversational partners and that this is reflected more in the behavior of a seemingly more empathetic person. In a recent study Briñol and Petty [4] showed that head nodding could affect the attitude in the person itself. All these findings can explain how unconscious head nodding could be an index of or even affect the conversational interest and dynamics between both the speaker and the listener.

In the proceeding analysis of conversational interest, we largely look at head-movement and head nodding as an aggregate group statistic. We have evaluated various characteristics like the group head nodding energy, variance and means, and classifier results, and compare them to interest ratings given by users of the system.

The head-nodding classifier is based on a gaussian mixture model [8], with 31-dimensional features generated from the frequency domain representation (32 point FFT) of the raw accelerometer signal. 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 Gaussian mixture model training is done offline in MATLAB, and the results are fed into a real-time classifier running on the Zaurus PDA, that computes the FFT and identifies the correct class based on the mixture model.

3.4 User Interest Rating

In order to be able to correlate user-reported interest with these sensed features, we have the user input an interest rating on a numeric scale, on a touch-screen interface on a Zaurus PDA. This PDA also does data acquisition for all sensors connected to the user.

The application *Rateit!* on the Zaurus is used to accept objective user interest ratings on a scale of one to ten, via the touch-screen display. This GUI was particularly designed so that the users can change interest ratings without significant cognitive load and without having looking at the PDA touch-screen to use it.

4. EXPERIMENTS AND RESULTS

4.1 Speech Features and Speed Dating

Scenario: Ten couples are at a speed-dating party at a popular local bar. A closer look reveals that they all are using Sharp Zaurus PDAs with microphones, and some are seen wearing BodyMedia GSR monitors. The PDA's perform real-time speech-feature analysis and collect user feedback at the end of each five-minute 'dating' session that can be correlated with 'liking' or 'getting-along' with potential matches.

Speed dating is relatively new way of meeting many potential matches during an evening. Participants interact for five minutes with their date, at the end of which they decide if they would like to stay in touch with him/her, and then move onto the next person. A 'match' is found when both singles answer yes, and they are later provided with mutual contact information. Since each speed-dating session is a short, intense conversation with objective results, we though this was an interesting scenario to understand speech features.

The GroupMedia systems were used for speech-feature analysis for a total of 57 high-confidence, five-minute speed-dating sessions. For each session, participants gave feedback on four questions – if they were romantically interested in the person, if they would like to stay in touch just as friends, if they would like to stay in touch for a business relationship and how well they got along on a scale of one to ten. (The first three questions were yes/no type questions)

We used the speaking/not-speaking HMM to calculate both individual and dynamic group features- fraction of speaking time, standard deviation of energy, standard deviation of format frequency, standard deviation of spectral entropy, voicing rate, influence parameters and back-and-forth exchanges. These features (without any understanding of the content) can explain more than 1/3rd of the variance in ratings. For a female 'liking' a male during a 5 minute session, the correlation values with combined male and female speech features are r = 0.67, p=0.05. Similarly, significant correlations are obtained for the 'friendly' (r=0.63, p=0.04) and 'business' relationship yes/no questions (r=0.7, p=0.03) for both female and males.

An interesting observation was that female features alone showed more correlation with both male (r=0.5,p=0.02) and female (r=0.48, p=0.03) ratings, than male features alone with male or female ratings (very low). In other words, female speech features and speaking styles are more important for a couple 'liking' or 'getting along' with each other, than male features. In terms of individual features, the most important feature for female 'liking' ratings was her own voicing rate, and for male 'liking' ratings it was the female's speaking time.

The influence parameters for the speed-dating data set have a better correlation with ratings for friendship and business, than romantic interest. Essentially, the influence parameters capture turn-taking dynamics between people, since the alpha values represent the "influence" all speakers present speaking/not-speaking states have on a particular speaker's next-state. We have seen some relation between the influence parameters, and empathy and exchange of information in the past [19,20], and possibly representing situations where people need to show empathy or control. Also, the back-and-forth exchanges were statistically important for female ratings, but show little or no importance for male ratings.

The next step is to build classifiers that can model these correlations. Our first attempt, a two-class linear classifier based on regression coefficients has a cross-validation accuracy of about 70% for predicting whether someone will 'like' you, based on speech feature analysis. By using a three class linear or SVM classifier, seems like for 80% of the people we can predict their `yes/no' with very high accuracy, and for the remaining 20% it is too close for us to call (and we can tell if a subject is in the 80% group or the 20% group). We have similar results for the yes/no question about friendly and business relationships, and the business relationship question is like the `gut feel' people have about hiring someone.

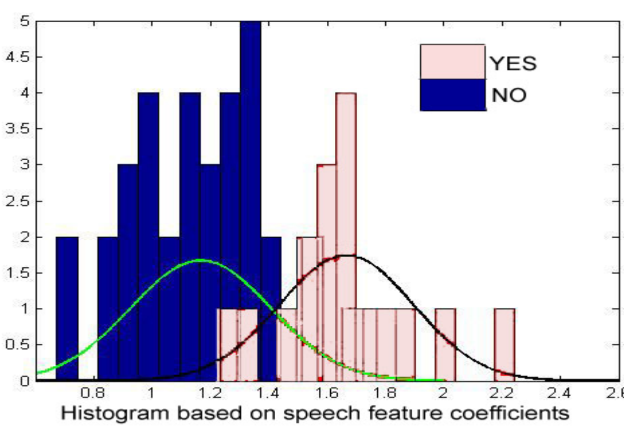

Figure 4. Histogram plot for female business relationship question (yes/no)

The graph above shows a histogram plot for the women's answers for 'business relationship' (yes/no). The X axis is calculated from the feature coefficients, and the two colors indicate actual responses of women. As seen, they are separable and it is possible to assign high probabilities of yes/no answers to sessions on the extremities. The gaussians fitted on the samples illustrate one simple way of modeling these probabilities.

4.2 Group Conversational Interest

Scenario: Group deliberations and decision-making are an integral aspect of Sloan Business School. Four business school students are keenly involved in an animated discussion to find a class project they have to execute as a team. Individuals are using Rateit! on their Zaurus PDAs to give an objective rating of how interesting they find the ideas. This can be correlated with their head-movement and nodding, speech features and physiology, to understand the socio-physiometrics of brainstorming and idea-generation.

The GroupMedia system has been used to measure conversational interest in ten sessions of an MIT class called Digital Anthropology. Each session involved a group of 3-4 people, for durations ranging from ten minutes to an hour. The students engaged in conversation and brainstorming sessions, while we evaluated their physio-metrics and subjective interest ratings.

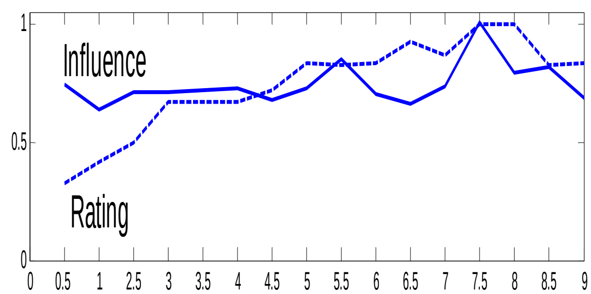

Figure 5. Plot of participant interest ratings from the PDA application Rateit! and the corresponding influence parameters.

The above graphs show a typical data from these meetings. The graphs show group interest ratings during an eight-minute session, and the corresponding influence parameters calculated from the audio features. A long-term rise in interest is observed, along with a corresponding long-term rise in influence parameters; the influence parameters are a statistically significant predictor of user interest ratings with a correlation of r=0.5. More importantly, the 'bumps' in the interest rating correspond to `bumps' in the influence parameters. This supports the ideas that as individuals find the conversation more interesting they begin to influence each other more (were more `engaged'). Using the mathematical influence parameters and the objective interest ratings, we begin to see these trends relating *social influence* and *interest.*

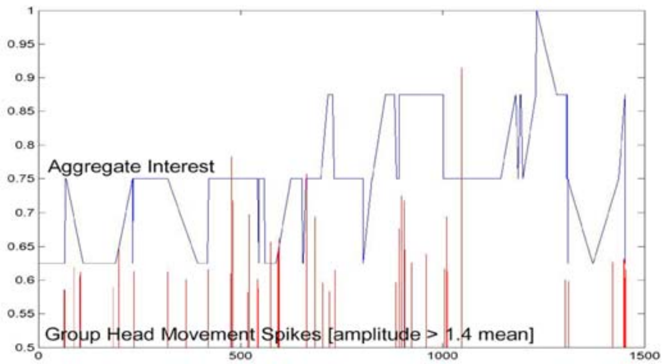

Figure 6. Plot of interest ratings from the PDA application Rateit! and the corresponding overall group head-nodding energy

There was also a correlation between the overall head-movement in the group and objective interest. Figure 6 shows clusters of head nodding behavior when there are changes in interest ratings. Bursts of group head-nodding correctly identify 80% of the changes in group interest level, with a 1/3 false alarm rate. Head nodding is not a perfect indicator of interest change, nor does it give the sign of the change, but it does provide the very useful information that we should look for changes in participant behavior and interest.

4.3 Movie Audience Interest

Scenario: A few friends are intently watching an exciting movie together. The scene looks perfectly normal, until one notices that their baseball hats have accelerometers on them, and they are wearing small GSR leads on their fingertips. Their animated reactions to various scenes in the movie, reflected in their physiology could be invaluable information for the movie's producers and editors.

The GroupMedia system has been used to measure audience behavior and reactions to short movies and commercials, in the form of their galvanic skin response, head-movements and speech features (intermittent conversation, laughing, and conversation after the clip).

In our experimental protocol for short films, a total of 15 subjects (in groups of 3-4) were shown a three-minute short film, followed by five minutes of conversation, and then by another nine-minute film. The subjects used Rateit! on their PDA's to give a moving scale of interest, that could be correlated with their features.

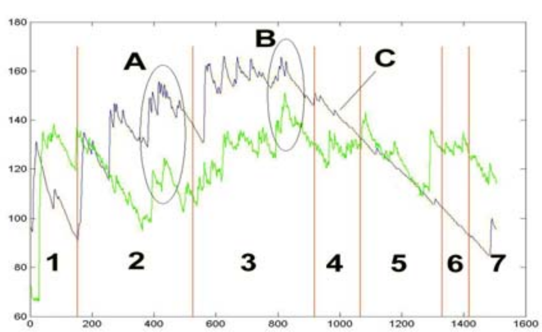

Figure 7. GSR signals for two individuals, where vertical lines indicate contextual events in the epoch, like watching a short film (1), conversation (2), watching another short film (3 to 6), and filling out survey forms (7). It can be seen that several regions, e.g. A and B, are highly correlated, whereas other regions, e.g., C, are not.

Figure 7 shows GSR responses for two individuals, measured for about twenty-five minutes while the subjects watched the two short films. The first short film was generic, and the second was specifically related to the academic institution of one of the two subjects. The vertical lines in figure 7 indicate changes in activity. Section 1 was when subjects were watching a short film, section 2 was when they engaged in conversation, sections 3 to 6 was when they watched another short movie and during section 7 they filled out final survey forms and spoke informally.

As seen, there is a sharp rise in GSR values at the start of section 2 for both subjects; this is when they started conversing at the end of the first short film, and one of them asked a question. Points A [during conversation] and B [during the second short film] reflect typical areas of strong correlation in GSR responses of both individuals experiencing identical startle, cognitive or emotional events. On the other hand, C represents the start of a long-term drop in GSR response, for a subject who was not related to the academic institution that the second film was about. By using the survey forms and the interest-ratings, we were able to correlate these strong physiological reactions with particular scenes and effects in the movie, and this feedback was provided to the movie's creators.

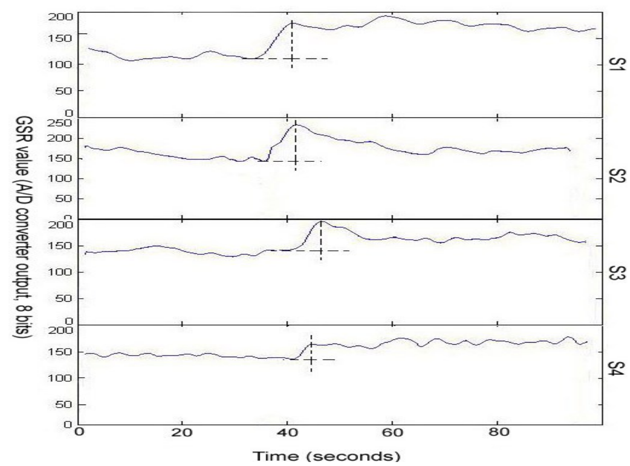

Figure 8: Almost simultaneous GSR spikes observed in 2 sessions (subjects S1.S2 and S3, S4) watching a clip with an unexpected, shocking ending

expected to be extremely funny and amusing, and the third had an unexpected shocking ending. The graph above shows the GSR spikes for the clip with an unexpected shocking ending for 4 subjects viewing it (S1 and S2 together, S3 and S4 together) - prominent spikes at the surprise ending can be detected with slope detection on the GSR signal.

This form of physiological "annotation" and feedback for movies and commercials is the first step towards a link between behavior, physiology and interest for movie audiences.

5. VISUALIZATION AND FEEDBACK

The motivation behind a real-time system that captures multi-modal signals and does real-time analysis is to be able to give feedback to individuals within the conversational setting and impact their performance.

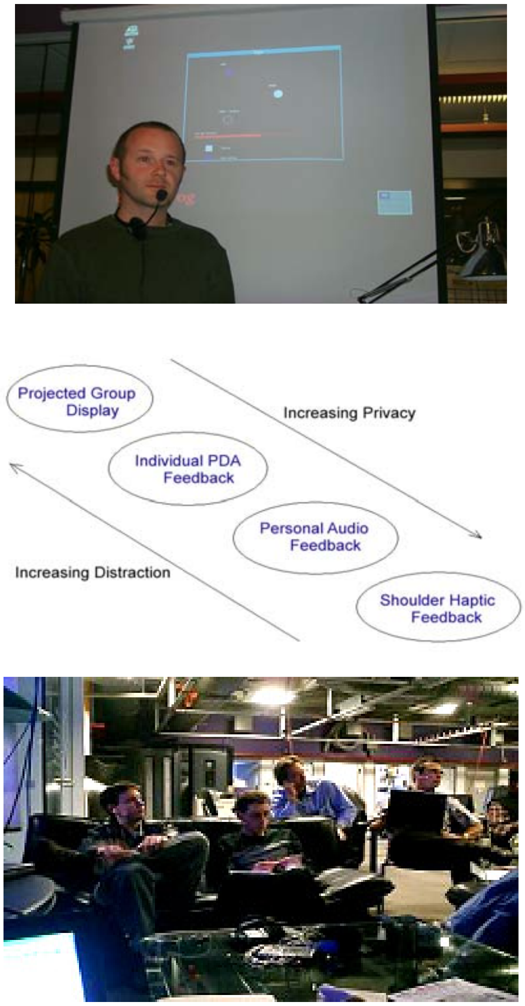

Figure 9: (top) a GroupMedia public display showing users, real-time speaking and head nodding results, (middle) types of feedback, (bottom) a small class using GroupMedia.

We have tried several forms of projected displays; at the very least the display shows a real-time rolling graph of overall interest within the group, calculated from the *Rateit!* application on the Zaurus PDA. A more elaborate display shows individual states, like speaking, interrupting, not-speaking [based on the real-time speech classifier], head-nodding [results of the head-nodding classifier on the Zaurus] and also individual GSR signals as socio-physiological feedback. Based on real time audio feature extraction we also can display the influence parameters graphically. During some of the sessions, we have varied the projected visualization of interest and group statistics, used alternatives or removed feedback altogether.

The difference between feedback on projected display and personal feedback on a Zaurus PDA is in the privacy of information conveyed. If the individual only knows his feature analysis and the overall interest other people are experiencing in his contributions to the exchange, he can modify his behavior. Projected displays, on the other hand, may violate this perceived feeling of privacy, since people are known to modify their social behavior when placed under open scrutiny [17]. Informally we have found that using a more private display is quite effective. The individual's statistics are displayed on the person's PDA screen and are updated every minute. Most effective seem to be a statistic about his/her participation in the conversation, the frequency at which his/her comments occurred, and the group responses to them (in terms of whatever measurements are available). On a similar note, a real-time personal feedback system for speed dating would be quite compelling.

Audio cues over headphones can also be used to give the same information in an even more private manner. We also tested haptic [vibrato-tactile] feedback by mounting coin-cell motors on the shoulder. [23].

6. CONCLUSION AND FUTURE WORK

We have developed a framework that helps us understand the correlation between prosodic audio features, head-movement, some physiology and participant ratings of interest in social interactions. We have shown how speech features can significantly explain more than 1/3rd of the variance for 'liking' or 'getting along' with a partner in speed dating. We have argued that correlation in galvanic skin responses of multiple individuals may be used to understand common interest states within a group. We have shown that there seems to be a correlation between the turn-taking influence parameters and interest. Finally, changes in interest seem to be marked by significant head-movement and nodding within the group. Although the experimental scenarios may focus more on some feature than others, these building block experiments help us understand the combination of our features into a real-time *group interest index*.

Our long-range goal is to open the way for more *social intelligence* in these wearable devices. Our results can be made more meaningful by adding contextual constraints to the measured behavior. We are pursuing this approach in several types of social situation, like speed dating, negotiations, poker games, and live performance audiences. By looking at behavior within these particular contexts, we hope to come up with a broad understanding of the relation between behavior, various types of physiology, and mental states such as interest.

The potential implications of this technology are enormous. Physiology based *behavior-metrics* can offer a much more fine-grained measurements for movie productions, live performances, advertising campaigns, focus group studies, etc. than any paper-based surveys ever could. Also, by overlaying social network information with a real-time understanding of interest, we would be able to relay information about individual interest levels over their social network, and can start to build distance-separated *interest-networks*. Social intelligence in pervasive mobile devices could give users real-time feedback, annotate important sections and connect distance-separated users at relevant moments.

7. ACKNOWLEDGMENTS

The authors would like to thank Dr. Carl Marci, M.D., Mass. General Hospital for his extensive knowledge about human physiology and for helping us with GSR equipment calibration and Dr. Martin C. Martin, whose work on speech feature analysis guided our research. Also, we want to thank Joost Bonsen for suggesting the project name and his untiring efforts in getting our projects organized successfully, and our UROP Shaun Foley for his able engineering skills.

8. REFERENCES

[1] Bales, R. F. Social interaction systems: Theory and measurement. New Brunswick, NJ:Transaction Publishers, 1999.

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

[3] Brand, M., Oliver, N., and Pentland, S., “Coupled hidden Markov models for complex action,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR97), 1996.

[4] Briñol, P. and Petty, R.E., “Overt head movements and persuasion: A Self-validation analysis.” Journal of Personality and Social Psychology, *84,* 1123-1139, 2003.

[5] Chartrand, T.L., and Bargh, J.A., “The chameleon effect: The preception-behavior link and social interaction.” Journal of Personality and Social Psychology, 76, 893-910, 1999.

[6] Choudhury, T., “Sensing and Modeling Human Networks.” PhD thesis, Dept of MAS, MIT, 2003.

[7] Contexual Computing Group, Georgia Tech, GA Bluetooth accelerometer design (http://www.cc.gatech.edu/ccg/resources/btacc/index.html)

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

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

[10] Eagle, N. and Pentland, A., "Social Network Computing", Ubicomp 2003: Ubiquitous Computing, Springer-Verlag Lecture Notes in Computer Science 2864, pp. 289-296, 2003.

[11] Fuller, G. D., “GSR History, & Physiology” taken from Biofeedbacks Methods and Procedures in clinical practice, 1977.

[12] Gerasimov, V., Hoarder Data Acquisition System, (http://vadim.www.media.mit.edu/Hoarder/Hoarder.htm)

[13] Healey, J. and Picard, R. W.,“StartleCam: A Cybernetic Wearable Camera”, In Proc. ICWC ‘98, 1998

[14] Kapoor, A. and Picard, , R. W., “A real-time head and nod shake detector”. In IEEE PUI, Orlando, FL, 2001.

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

[16] Marci, C., ”Psychotherapy Correlates of Empathy, An investigation of the patient-clinical dyad”, 2002

[17] 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.

[18] Miller, J.G., Living System, McGraw-Hill, New York.

[19] Pentland A, Curhan J, Eagle N, Martin M., “Money and Influence”, Media Lab Technical Note 577, May 2004

[20] Pentland A, “Social Dynamics:The Voice of Power and Influence”, Media Lab Technical Note 579, May 2004

[21] Picard, R.W., “Affective Computing”, MIT Press, Cambridge

[22] Reynolds C. Strauss M. , Picard R.W., Bluetooth GSR sensor design, (http://arsenal.media.mit.edu/notebook/archives/cat_projects.html)

[23] Toney, A. and Dunne, L., “A Shoulder pad insert for Vibrotactile Display”, Proc ISWC ’03, White Plains, N.Y., October 2003