Socially Aware Media

Alex (Sandy) Pentland
MIT Media Laboratory
Room E15-387, 20 Ames St
Cambridge, MA 02139
pentland@media.mit.edu

ABSTRACT

Face-to-face communication conveys social context as well as words, and it is this social signaling that allows new information to be smoothly integrated into a shared, group-wide understanding. By building machines that understand social signaling and social context we can begin to make communication tools that keep remote users `in the loop,' and can dramatically improve collective decision making.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: Collaborative Computing, Computer-supported cooperative work, Organizational Design, Theory and Models

General Terms: Management, Measurement, Design, Experimentation, Human Factors, Standardization.

Keywords: Social Signals, Affect, Non-linguistic communication.

1. INTRODUCTION

Imagine a small group of individuals on the prehistoric African veldt. Each day the adults go out gathering and hunting, and in the evening return to sit around a central clearing where they recount the events and observations of the day, and discuss what to do tomorrow. During the group discussion social signaling…reflecting the power hierarchy as well as individual desires…accompanies the new information and the collective social signaling communicates to each individual what the group thinks about it. At the end of this social discussion collective decisions have been made, and the required individual behaviors are enforced by the iron hand of social pressure.

Psychology has firmly established that the same sort of social processes are still a dominant part of our modern lives. Instead of talking about the tribal spirit we now speak of the corporate spirit, and instead of dominance displays we have office politics. But the core facts are the same: we are social animals, and the meaning of information depends on its social context.

Unfortunately, technology has so far focused either on the isolated individual, or has treated the person as just another cog in an information processing machine. The result is that current communications technology doesn't feel very good. Buzzing pagers, ringing cell phones, and barrages of e-mails are leashes that keep people tethered to their job, and people worry that we are being assimilated into some sort of unhappy Borg Collective.

Technologists have responded with interfaces that pretend to have feelings or that call us by name, filters that attempt to shield us from the digital onslaught, and smart devices that organize our lives by gossiping behind our backs. But the result usually feels like it was designed to keep us isolated, wandering like a clueless extra in a cold virtual world.

These solutions, while well meaning, ultimately fail because they ignore the core problem: Computers are socially ignorant. Technology must account for this by recognizing that communication is always socially situated, and that discussions are not just words but also part of a larger social dialog. Successful human communicators universally recognize that communication is part of an evolving social process, and use this fact to their advantage. Digital communications can begin to do the same by trying to quantify social context and understanding how this context can be used to select successful interaction behaviors.

At MIT, my research group and I are working to automatically quantify social context in human communication. We have developed three `socially aware' platforms that objectively measure several aspects of social context, including nonlinguistic *social signals* measured by analyzing the person's tone of voice, facial movement, or gesture [1]. We have found these nonlinguistic social signals to be particularly powerful for analysis and prediction of human behavior, sometimes exceeding even expert human capabilities. These tools for measurement of social context permit the communications system to support social and organizational roles instead of viewing the individual as an isolated entity. Example applications include automatically patching people into socially important conversations, instigating conversations among people in order to build a more solid social network, and reinforcing family ties.

2. SOCIAL SIGNALS

Psychologists have firmly established that social signals are a powerful determinant of human behavior and speculate that they may have evolved as a way to establish hierarchy and group cohesion [2,3]. Most culture-specific social communications are conscious, however other social signals function as a *sub*conscious collective discussion about relationships, resources, risks, and rewards. In essence, they become a subconscious `social mind' that interacts with the conscious individual mind. In many situations the nonlinguistic signals that serve as the basis for this collective social discussion are just as important as conscious content for determining human behavior [1,2,3,4,5].

2.1 Affect, Paralinguistic Signals, Attitude

Social interaction has commonly been addressed within two different frameworks. One framework comes from cognitive psychology, and focuses on emotion. Ekman and Friesen [17] are

MM'05, November 6–11, 2005, Singapore.

the most well-known advocates of this approach, which is based roughly on the theory that people perceive others' emotions through stereotyped displays of facial expression, tone of voice, etc. This simplicity and perceptual grounding of this theory has recently given rise to considerable interest in the computational literature [13]. However serious question about this framework remain, including the question of what counts as affect? Does it include cognitive constructs such as interest or curiosity, or just the base dimensions of positive-negative, active-passive? Another difficulty is the complex connection between affect and behavior: adults are skilled at hiding emotions, and seemingly identical behaviors may have different emotional roots.

The second framework for understanding social interaction comes from liguistics, and treats social interaction from the viewpoint of dialog understanding. Kendon and Argyle are among the best known pioneers in this area [15,16], and the potential to greatly increase the realism of humaniod computer agents has generated considerable interest from the computer graphics and human-computer interaction community [14]. In this framework prosody and gesture are treated as annotations of the basic linguistic information, used (for instance) to guide attention and signal irony. At the level of dialog structure, there are linguistic strategies to indicate trust, credibility, etc., such as small talk and choice of vocabulary. While this framework has proven useful for conscious language production, it has been difficult to apply it to dialog interpretation, perception, and for unconscious behaviors generally.

In this paper I propose a new computational framework, that of social signaling, in which speaker *attitude* or *intention* is conveyed through the amplitude and frequency of prosodic and gestural activities. This framework is based on the literature of personality and social psychology, and is different from the linguistic framework in that it consists of non-linguistic, largely unconscious, signals about the social situation, and different from the affect framework in that it communicates social relation and not speaker emotion. It is different in another way as well: it happens over longer time frames than typical linguistic phenomena or emotional displays, treating gestures more like a motion texture than individual actions, and it appears to form a largely independent channel of communication. In the language of the affect framework, these signals are sometimes identified by the oxymoronic label `cognitive affect', whereas in the linguistic framework they might be related to dialog goals or intentions. Social signaling is what you perceive when observing a conversation in an unfamiliar language, and yet find that you can still `see' someone taking charge of a conversation, or establishing a friendly interaction [3].

2.2 Predicting behavior

Social psychologists have found social signals to be extremely powerful in predicting human behavior across a wide range of school, business, government, and family situations. With only a few minutes of observation, an expert psychologist can regularly predict behavioral outcome with about 70 percent accuracy [3]. Amazingly, observation of such `thin slices' of behavior can accurately predict even important life events—divorce, student performance, and criminal conviction—even though these events might not occur until months, sometimes years, later.

Following the social psychologists' example, I reasoned that a test for our ability to automatically measure social signals would be our ability to predict outcomes from `thin slice' observation of human interactions. Could we predict human behavior without listening to words or knowing about the people involved?

Together with my research group I have built a computer system that objectively measures a set of non-linguistic social signals, such as engagement, mirroring, activity, and stress, by looking at `tone of voice' over one minute time periods [1]. Unlike most previous research, I wanted to measure signals of speaker attitude rather than trying to puzzle out the speakers instantaneous internal state, and consequently treated prosody and gesture as a longer-term `motion texture' rather than focusing on individual motions or accents. Although people are largely unconscious of this type of behavior, other researchers [2,3,6,7] have shown that similar measurements are predictive of infant language development, judgments of empathy, attitude, and even personality development in children.

Using this `social perception machine' we could "listen in" to the social signals within conversations, while ignoring the words themselves. We found that after a few minutes of listening in this way, we were able to accurately predict:

- who would exchange business cards at a meeting;
- which couples would exchange phone numbers during a speed dating event in alocal bar;
- who would come out ahead in a negotiation;
- who was a connector within their work group; and
- a range of subjective judgments, including whether or not a person felt a negotiation was honest and fair or a conversation was interesting.

In a typical case with a three-class linear decision (yes, not enough information, no) the yes/no accuracy is almost 90 percent. Accuracy is typically around 80 percent with a two-class linear decision rule, where we make a decision for every case. More generally, linear predictors based on the measured social signals typically have a correlation of r=0.65, ranging from around r=0.40 to as much as r=0.90. Most experiments involved around 90 participants, typically 25 to 35 years old, with a third being female (for more detail see http://hd.media.mit).

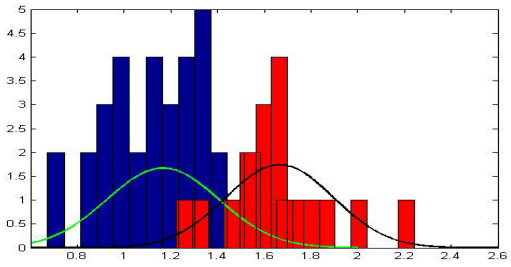

Figure 1: Measuring Prediction Accuracy.

The histogram shown in Figure 1 shows a typical case, in this instance the data is `would you like to work with this person or not?'. The blue parts of the histogram are `no' answers; the red parts are `yes' answers. The vertical axis is frequency of data points, and the horizontal axis is our predictor, with greater values meaning a `yes' is more likely. If you place the yes-no boundary at 1.4, you get a 72 percent decision accuracy.

Achieving this level of accuracy is pretty amazing, especially since experiments using human experts have typically shown considerably less accuracy. Moreover the decisions we examined are among the most important in life: finding a mate, getting a job, negotiating a salary, and finding your place in a social network. These are activities for which we prepare intellectually and

strategically for decades. Yet the largely subconscious social signaling that occurs at the start of the interaction appears to be more predictive than either the contextual facts (attractiveness and experience) or the linguistic structure (strategy chosen, arguments employed, and so on).

3. QUANTIFYING SOCIAL SIGNALS

The machine understanding community has studied human communication at many time scales --- e.g., phonemes, words, phrases, dialogs --- and both semantic structure and prosodic structure have been analyzed. However, the sort of longer-term, multi-utterance structure associated with signaling of social attitude (e.g., interested, attracted, confrontational, friendly, etc). has received little attention from the machine understanding research community. To quantify these social signals I began with a broad reading of the voice analysis and social science literature, and eventually developed texture-like measures for four types of social signaling, which were designated activity level, engagement, stress, and mirroring [1]. By using these measurements to tap into the social signaling in face-to-face discussions, we can then anticipate outcomes by use of learned statistical regularities. Although in this article we will discuss mostly vocal measures of social signaling, I have also developed face and hand gesture equivalents to the audio features, and experiments using these visual features are underway

3.1 Activity level

Activity level is the simplest measure: it is how much you participate in the conversation. For the activity-level measure, we use a two-level Hidden Markov Model (HMM) to segment the speech stream of each person into voiced and non-voiced segments, and then group the voiced segments into speaking versus nonspeaking. We then measure conversational activity level by the percentage of speaking time.

3.2 Engagement

In broad terms, engagement is how involved a person is in the current interaction. Is he driving the conversation? Is she setting the tone? Engagement is measured by the influence each persons pattern of speaking vs. not-speaking has on the other person's pattern, that is, engagement is a measure of who drives the pattern of conversational turn-taking. When two people are interacting, their individual turn-taking dynamics influence each other, which we can be model as a Markov process [6]. By quantifying the influence each participant has on the other, we obtain a measure of their engagement. To measure these influences we model their individual turn-taking by an Hidden Markov Model (HMM) and measure the coupling of these two dynamic systems to estimate the influence each has on the others turn-taking dynamics [8]. Our method is similar to the classic method of Jaffe et al. [6], who found that engagement between infant-mother dyads is predictive of language development. Our formulation relaxes Jaffe et al.'s parameters so that we can calculate the direction of influence and analyze conversations involving many participants.

3.3 Stress

Stress is the variation in prosodic emphasis. For each voiced segment we extract the mean pitch (frequency of the fundamental format), and the spectral entropy. Averaging over longer periods provides estimates of the mean-scaled standard deviation of the formant frequency and spectral entropy. The sum of these standard deviations becomes a measure of speaker stress; such stress can be either purposeful (prosodic emphasis) or unintentional (caused by discomfort). Similar measures of vocal stress have been used to detect deception and also to predict the development of personality traits such as extroversion in very young children.

3.4 Mirroring

Mirroring is when one participant unconsciously copies the prosody and gesture of the other. Mirroring is considered a signal of empathy, and can positively influence the outcome of a negotiation and other interpersonal interactions [7]. In our experiments the distribution of utterance length is often bimodal. Sentences and sentence fragments typically occurred at several-second and longer time scales. At time scales less than one second there are short interjections ("uh-huh"), but also back-and-forth exchanges typically consisting of single words ("OK?" "OK!" "Done?" "Yup"). The frequency of these short utterance exchanges is our measure of mirroring behavior.

4. SOCIALLY AWARE SYSTEMS

We have incorporated these social signaling measurements into the development of three `socially aware' communications systems. Figures 2-4 show these systems in use; the Uberbadge is a badge-like platform [9], GroupMedia is based on the Sharp Zaurus PDA [10], and Serendipity is based on the Nokia 6600 mobile telephone [11]. In each system the basic element of social context is the identity of people in the users' immediate presence. The systems use several methods to determine identity, including Bluetooth-based proximity detection, infrared (IR) or radio-frequency (RF) tags, and vocal analysis.

To this basic context, one can add estimates of individuals' social attitude (e.g., interested, determined, cooperative, attracted, pessimistic, etc.) , using audio feature analysis, sensors for head and body movement, and even biosignals such as galvanic skin response (GSR). These sensing capabilities provide a quantitative measure of the attitudes displayed by people in the user's immediate, face-to-face social context. The resulting system can identify face-to-face interactions, capture collective social information, extract meaningful group descriptions, and transmit the group context to remote group members.

Figure 1: The UberBadge.

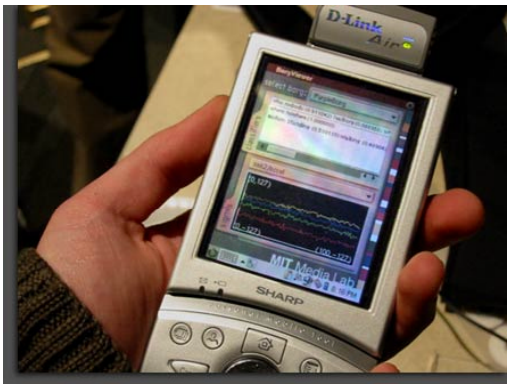

Figure 3: The GroupMedia System.

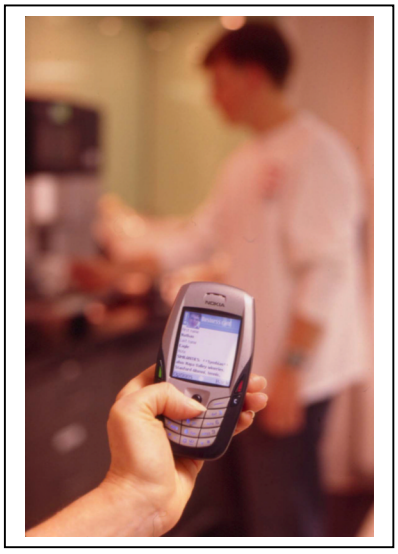

Figure 4: The Serendipity system.

When the system detects a face-to-face interaction, defined as the combination of proximity and conversational turn taking, it defines a group context that consists of the participants' identities, attitudes estimated from social signals as described above, and the compressed audio (and possibly video) information stream. It then creates a *social gateway* that contains the group context information and that lets preapproved members of the social or work group access the ongoing conversation and group context information. The social gateway uses real-time machine learning methods to identify relevant group context changes. A distance-separated user can then access these group context changes.

4.1 A new level of communication

How can `socially aware' systems change human communications? How can knowing social context and in particular speaker attitude help? Simple uses are to provide people with feedback on their own interactions. Did you adopt a forceful attitude during a negotiation? Were you interested when you were talking to your spouse? Did you project a helpful, empathic attitude during the teleconference? Such feedback can potentially head off many unnecessary problems.

The same sort of analysis can also be useful for robots and voice interfaces. While word selection and dialog strategy are very important to achieve a successful human-machine interaction, our experiments and those of others show that social signaling may be even more important.

4.1.1 What was that name?

An obvious use for social context is to help build your social network. At some time, nearly everyone has met an interesting person and then has lost that person's business card or forgotten that person's name. On the basis of an audio analysis and observations of body motion, our UberBadge system [9] can keep track of all interactions where you seem interested in the other person, and email you the names and particulars of those individuals at the end of the day.

4.1.2 Building social capital

Social capital is the ability to leverage your social network by knowing who knows what, and to whom you should talk to get things done. It is perhaps the central social skill for any entrepreneurial effort, yet many people find it difficult. We are therefore building systems that can help a person build social capital.

One example is Serendipity, a system implemented on Bluetooth-enabled mobile phones [11] and built on BlueAware, an application that scans for other Bluetooth devices in the user's proximity. When Serendipity discovers a new device nearby, it automatically sends a message to a central gateway server with the discovered device's ID. If it finds a match, it sends a customized picture message to each user, introducing them to each other.

The real power of this system is that it can be used to create, verify, and better characterize relationships in online social network systems, such as Friendster or Orkut. If two people hang out after work, they are probably social friends. If they meet only at work or not at all, they are likely to have a very different relationship. The system can refine the relationship characterization by analyzing the social signaling that occurs during phone calls between the two people. The phone extracts the social signaling features as a background process so that it can provide feedback to the user about how that person sounded and to build a profile of the interactions the user had with the other person.

4.1.3 Staying in the loop

A major problem with distributed workgroups is keeping yourself in the loop. Socially mediated communications, such as our GroupMedia system [10], can help with this problem by patching people into important conversations. When it detects a potentially interesting conversation, it could notify distant group members. Whether or not a certain member receives notification depends on measured interest levels, direction of information flow, and group membership.

Once the distant group member receives a notification, that person has several options. He can subscribe to the information and begin to receive the raw audio signal plus annotations of the social context or he can choose to have the system notify him only in case of especially interesting comments or he can store the audio signal with social annotations for later review.

Suppose, for example, most of your workgroup has gathered, the information flow is from the boss, and the interest level is high. You might be wise to patch into the audio and track the measured level of group interest for each participant's comments. The group context

information and the linking-in notification that the system gateway provides can increase both the group cohesion and your understanding of the raw audio.

The same framework could also enhance the social life of close friends. Suppose two or three of your closest friends have discovered an amazing band at a bar, and are having a great time. The system could detect the situation and, given appropriate prior authorization, automatically send you an invitation to join them. Although such a system wouldn't be to everyone's taste, the idea generally gets a thumbs-up from college undergraduates.

4.1.4 Group dynamics

Social scientists have carefully studied how groups of people make decisions, and the role of social context in that process. Unfortunately, what they have found is that socially mediated decision-making has some serious problems, including group polarization, `groupthink,' and several other types of irrational behaviors that consistently undermine group decision-making [2,4]. To improve group function you need to be able to monitor the social communication and provide real-time intervention. Human experts can do that (they are called facilitators or moderators) but to date machines have been blind to the social signals that are such an important part of human group function.

The challenge, then, is how to make a computer recognize social signaling patterns. In salary negotiation, for example, we found that the lower status individual does better when they show more mirroring, which communicates that they are a team player. In a potential dating situation, the key variable was the female's activity level, indicating interest. By knowing that certain signaling patterns reliably led to these desired states, the computer can begin to gently guide the conversation to a happy ending by providing timely feedback.

Similarly, the ability to measure social variables like interest and trust ought to enable more productive discussions, while the ability to measure social competition offers the possibility of reducing problems like groupthink and polarization. If the computer can measure the early signs of problems, then it can intervene before the situation becomes unsalvageable.

To explore these ideas I gave every student in my Digital Anthropology seminar a GroupMedia system, so that our team could analyze the group interaction [12]. Real-time displays of participant interaction could be generated and publicly displayed to reflect the roles and dyadic relationships within a class. Figure 5 gives an illustration of what such a display might look like. As the professor (s9) I appear as the seminar's dominant member, and my advisees (s2, s7, s8) have a high probability of conceding the floor to me.

This type of analysis can help develop a deeper discussion. Comments that give rise to wide variations in individual reaction can cause the discussion to focus on the reason for the disparity, and those interested can retrieve these controversial topics for further analysis and debate later. The analysis also permits the clustering of opinions and comments using collaborative filtering. In this way, people can readily see opinion groupings, which sets the stage for inter- and intragroup debates.

4.1.5 Personal relationships

Social awareness may also be able to help reinforce family ties, an important capability in this age of constant mobility. By sensing when family members have had an unusually good, or unusually bad, experience we can promote supportive communication between family members. In one version, the system would randomly leave phone messages reminding family members to call each other. However, when it senses that there has been an unusual experience --- a serious argument, an especially fun conversation, or an unusually intense meeting – the system would always leave reminders for others to call. The system would not tell people exactly why they should call, because doing so could violate people's privacy. Instead, the reminders would strengthen the family network by encouraging conversations precisely when they are most likely to be appreciated.

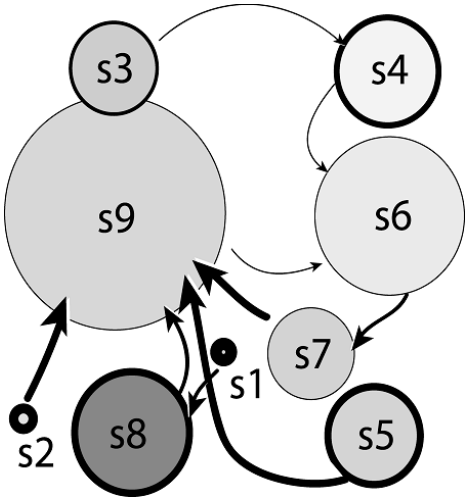

Figure 5. Display of group dynamics. Speaking Time is circle size, Transition Probability is width of the link, Average Interest Level is circle color (individual) or circle border (group).

5. SUMMARY

Social signaling seems to provide an independent channel of communication, one that is quantifiable and which can provide an important new dimension of communication support. The implications of a system that can measure social context are staggering for a mobile, geographically dispersed society. Propagating social context could transform distance learning, for example, letting users become better integrated into ongoing projects and discussions, and thus improve social interaction, teamwork, and social networking. Teleconferencing might become more reflective of actual human contact, since participants can now quantify the communication's value. Automatic help desks might be able to abandon their robotic, information-only delivery or their inappropriately cheerful replies.

Our current systems are just a first step toward generally useful communications tools. We must increase the reliability of our social context measurements and learn how to better use them to modulate communication. Much of our ongoing research is focusing on building meaningful mathematical models for estimating social variables and experimentally validating their use in a distance collaboration framework.

When considering the personal and societal effects of socially aware communications systems, I can't help but recall Marshall McLuhan's "the medium is the message." By designing systems that are aware of human social signaling, and which adapt themselves to human social context, we may be able to remove much of the

medium's message, and replace it with the traditional messaging of face-to-face communication. Just as computers are disappearing into clothing and walls, the otherness of communications technology might disappear as well, leaving us with organizations that are not only more efficient, but also better balance our formal, informal, and personal lives. Assimilation into the Borg Collective might be inevitable, but we can still make it a more human place to live.

6. ACKNOWLEDGMENTS

I thank my collaborators—Joost Bonsen, Jared Curhan, David Lazar, Carl Marci, M. C. Martin, and Joe Paradiso—and my current and former students—Sumit Basu, Ron Caneel, Tanzeem Choudhury, Wen Dong, Nathan Eagle, Jon Gips, Anmol Madan, and Mike Sung—for all the hard work and creativity they have added to this project. Thanks also to Deb Roy, Judith Donath, Roz Picard, and Tracy Heibeck for insightful comments and feedback. Parts of this article have appeared on Edge.org, *Proc. IEEE Int'l Conf. Developmental Learning, San Diego, Oct 2004*, and IEEE Computer, March 2005 pp. 63-70.

Code, data, and papers available at http://hd.media.mit.edu

7. REFERENCES

[1] Pentland, A. (2004) Social Dynamics: Signals and Behavior, Int'l Conf. On Developmental Learning, Salk Institute, San Diego, Oct. 20-22. See http://hd.media.mit.edu

[2] Nass, C., and Brave, S. (2004) Voice Activated: How People Are Wired for Speech and How Computers Will Speak with Us, MIT Press

[3] Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological Bulletin, 111(2), 256-274.

[4] Brown, R. (1986) Group Polarization, in Social Psychology (2d Edition), New York: Free Press

[5] Gladwell, M. (2000) The Tipping Point: How little things can make a big difference. New York: Little Brown

[6] Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., & Jasnow, M. (2001). Rhythms of dialogue in early infancy. Monographs of the Society for Research in Child Development, 66(2), No. 264.

[7] Chartrand, T., and Bargh, J., (1999) The Chameleon Effect: The Perception-Behavior Link and Social Interaction, J. Personality and Social Psychology, Vo. 76, No. 6, 893-910

[8] T. Choudhury, "Sensing and Modeling Human Networks." PhD thesis, Dept of MAS, MIT, 2003. Advisor: A. Pentland. See http://hd.media.mit.edu

[9] Laibowitz, M., and Paradiso, J. (2004) The Uberbadge Project, http://www.media.mit.edu/resenv/projects.html

[10] Madan, A., Caneel, R., Pentland, A., (2004) GroupMedia: Distributed Multimodal Interfaces, ICMI 2004. See http://hd.media.mit.edu

[11] Eagle, N., Pentland, A. (2004) Social Serendipity: Proximity Sensing and Cueing, MIT Media Lab Tech Note 580, May 2004. See http://hd.media.mit.edu

[12] Eagle, N., Pentland, A. (2003) Social Network Computing, Ubicomp 2003, Springer-Verlag Lecture Notes in Computer Science, No. 2864, pp. 289-296. See http://hd.media.mit.edu

[13] Picard, R. (1997), Affective Computing, MIT Press, Cambridge, 1997,

[14] Cassell, J., Bickmore, T. (2003) "Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents" *User Modeling and User-Adapted Interaction* 13(1-2): 89-132

[15] Kendon, A., Harris, R.M., & Key, M.R. (Eds). (1975). *Organization of behaviour in face to face interaction*. The Hague, Netherlands: Mouton

[16] Argyle, M. (1987). *Bodily communication*. Methuen.

[17] Ekman, P. and Friesen, W. (1977). *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA

.