Submitted to the International Conference on Information (ICIS). Seattle, WA. December 2003.

Dealing with Distance:
Capturing the Details of Collocation with Wearable Computers

RESEARCH IN PROGRESS PAPER

TRACK 1.

Technical Issues - Architecture, Systems, and Infrastructure

Nathan Eagle

MIT Media Lab
20 Ames St.
Cambridge, MA 02139

nathan@media.mit.edu

Abstract

This paper presents a method of harnessing a wearable computing infrastructure of PDAs and cell phones to capture detailed information on face-to-face interactions within the workplace. The system is designed to quantify the underlying dynamics of group behavior typically unavailable to distributed group members. It is our hypothesis that this additional source of social information will help make distributed teams more cohesive.

1. Introduction

Distributed teams will continue to proliferate in the near future. An expanding variety of communication tools and groupware continually make it easier for a team of geographically distant people to work together efficiently. However, many detriments of distance cannot be overcome despite the rapid advances in communication and collaboration technologies. The lack of persistent face-to-face rapport with collocated team members is thought to make a geographically isolated team member less cohesive (Hinds & Bailey, 2000).

Within these same organizations lies a ubiquitous wearable computing infrastructure that remains untapped. Mobile phones and PDAs have become part of standard corporate attire. Millions of people currently go to work *wearing* computers, sensors, microphones, and transceivers. And the functionality of these devices is continually increasing: handheld computers now have the potential for gigabytes of storage, wireless connectivity, and processors with speeds comparable to desktop PCs just a few years ago. Yet despite these capabilities, most organizations have not transcended the PalmPilot paradigm, limiting them to calendar and datebook -type functionality.

This paper bridges these seemingly unrelated topics by presenting a system that is able to quantify and represent the detailed dynamics of a collocated team by leveraging the existing wearable computing infrastructure within the workplace. Using custom analytic software and a mobile computing infrastructure consisting of over fifty wireless, linux-based handheld computers and an 802.11b network, we have created a testbed that we are using to learn the parameters that characterize group behavior and provide diagnostics for group efficiency. Although receiving daily information about these underlying team dynamics is by no means a substitute for face-to-face interactions, we hypothesize that this system will be able to supplement existing computer-mediated sources of social information for distributed team members.

2. Previous Work

Distributed Teams:

A wealth of research has revolved around distributed teams, and while the importance of cohesiveness is well established (Kiesler, 2002), methods to develop this cohesiveness remain an open question (Iacono & Weisband,1997). With collocated teams, the bandwidth for social information is so high that members are able to anticipate the strengths and shortcomings of their peers while monitoring group progress (Davenport, 1994; Walther, 97). The importance of such a fast social information transfer rate is one of the motivations for many managers to schedule periodic face-to-face interactions between all members of a distributed team (Maznevski, 2000). The large amounts of social information that are transferred during these encounters increases the efficiency of the team and trust between team members (Handy 1995).

A variety of mediated communication technologies have been studied as substitutes for collocation, including telephone, email, instant messaging, and sophisticated groupware applications (e.g., Sproull & Kiesler, 1991; Jarvenpaa & Leidner, 1999; Stough et al., 2000). However, the amount of social information that can be exchanged with existing meditated communication technologies pales in comparison to the transfer rate of face-to-face communication (Walther 1996-7). Our system attempts to capture a fraction of the social information transferred during these interactions. Just as it is beneficial to share information captured in meetings notes, communication logs and questionnaires with all the members of a team (Maznevski, 2000), we go a step further to propose a system that does away with the researcher entirely, and empowers the team to capture their own data and draw their own conclusions.

Quantifying Face-to-Face Interactions:

Self-report survey-based techniques have almost exclusively been used to quantify face-to-face interactions within the workplace (Wasserman & Faust, 1994). Recent advances in information technology, such as web-based surveys or even experience sampling methods on handheld computers (Barrett & Barrett, 2001), have helped facilitate gathering survey data, yet these techniques are unable to eradicate its fundamental flaws: bias and sparsity.

To overcome the bias and sparsity inherent in survey data, several infrared badges have been designed to detect face-to-face proximity. Choudhury et al. built a shoulder-mounted ‘sociometer’ that incorporated IR, sound level and accelerometers to establish social cliques within an organization. However, to move from social network mapping to capturing the details of social network function, many more features need to be extracted from these interactions.

3. Research Objectives

This system and study design began in response to two broad questions:

- Is it possible to capture a portion of the tacit dynamics and knowledge of a collocated team with existing handheld computers?
- Does disseminating information about these underlying behaviors of a core group reduce the social distance to peripheral group members?

Some of the most important aspects of a functioning team cannot be captured in a newsletter or progress report. For larger groups, these aspects are accumulated throughout an employee's extended career, for example: learning which people really influence results, who are the true experts on a subject, which people work well together, or who should connect with whom. On a smaller scale, these aspects could include information about who is currently excited about a given project, which people have lately been relatively isolated from the group, who has been sitting next to each other, or which members have been dominating recent conversations.

Potential beneficiaries of the system are not only the distributed team members, but also the collocated team and managers who assemble the team. The collocated group is able to get feedback on their efficiency and dynamics while also receiving individualized assessments about their personal social behaviors. These underlying dynamics can also help managers identify social problems and assemble future teams that are not only based on synergistic skill sets but also amenable social profiles.

4. The Reality Mining System

The system we designed to answer these questions is a combination of commercial hardware similar to existing devices in the modern workplace, and specialized 'sociometeric' software. The hardware is comprised of over fifty Sharp Zaurus PDAs, wireless CF cards, 256 MB SD cards, and either wired or wireless headset microphones. Applications were written to enable the Zaurus to stream high quality audio (22 kHz, 16-bit) to an available 802.11b network, or to store the audio locally when no network is detected. Besides streaming audio, packets in this wireless network could be 'sniffed' by the PDAs interested in determining who else is in the local proximity. Information regarding access point signal strength information was correlated with location using a static table look-up procedure. The system is typically kept in a participant's pocket, or for those with Bluetooth headsets, stored in a briefcase, purse, or backpack.

These handheld computers stream gigabytes of data each day to a central server for processing. This massive amount of audio is initially analyzed for speech. The speech detection algorithms use a variation of a multi-band center clipper over the 300 Hz to 3kHz frequency spectrum. Each audio stream is chunked and run through a bank of filters in the frequency domain. The output energy is thresholded to generate tentative speech segment labels (talking / not talking) over each second. An error-checking script correlates waveform segments streamed to the same access point over a short window to verify that the labeled speech regions were not due to another participant speaking loudly.

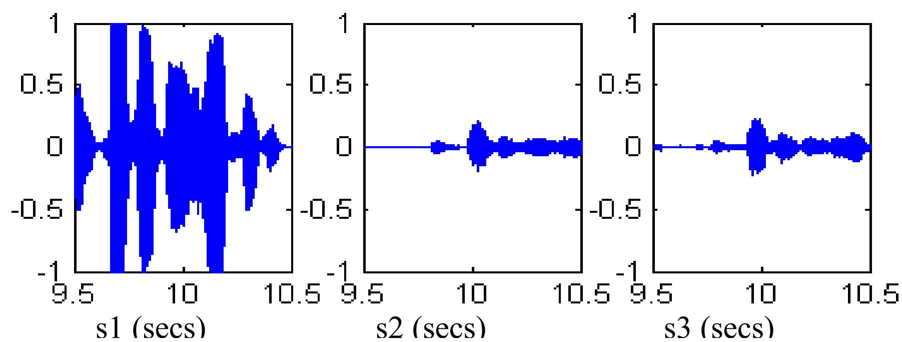

Fig. 1. These three waveforms sections are strongly correlated (c>.25), indicating that s2 and s3 are within earshot of the speaker (s1). The waveform's relative energy can also yield insight into the physical proximity between the speaker and the listeners.

Although a high correlation is indicative of proximity between two people, it is not a certain sign of a single conversation because the speech could be from two separate phone conversations. As shown in Figure 2 and described in (Basu, 2002), the voicing segment of one participant is the noisy compliment of the other participants' voicing segments. Measuring the mutual information between these binary voicing segments (v_1, v_2) has been shown to be an extremely reliable conversation detector:

$$\begin{aligned} a[k] &= I(v_1[t], v_2[t-k]) \\ &= \sum_{i,j} p(v_1[t]=i, v_2[t-k]=j) \log \frac{p(v_1[t]=i, v_2[t-k]=j)}{p(v_1[t]=i)\,p(v_2[t-k]=j)} \end{aligned} \tag{1}$$

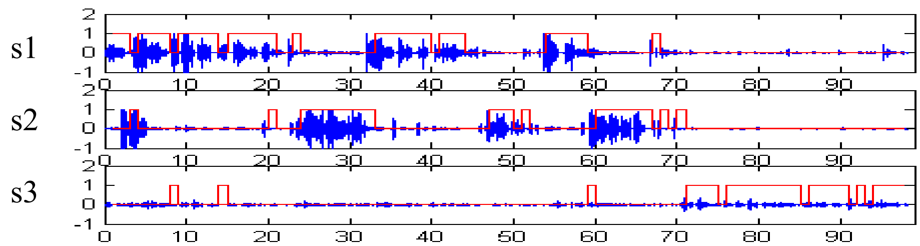

Fig. 2. A 90-second audio segment with turn-taking characteristics indicative of a single conversation

Once detected, the conversation audio streams are extracted and analyzed. Table 1 shows a selection of features that can be gleaned from this audio data. Profiles of a participant's typical social behavior are built over time using conversation features such as speaking rate, volume, duration, participants, interruptions, proximity of others, transition probabilities, and time spent holding the floor. During meetings, the devices can serve a dual purpose. The handheld computers also allow a user to input his or her interest level through an interface designed to minimize distraction.

The final component of the audio analysis is establishing the conversational situation: the topic and the surrounding context of a conversation. For situation understanding, our system combines a network of commonsense knowledge with contextual information (such as location, time and participants) that is obtained from the PDA. The best speech recognition engines today only get one correct word for every three during spontaneous speech, creating transcripts that even a human would have a difficult time interpreting. But just as additional contextual and common sense information can help a human infer the topic of a conversation, this type of information can be equally beneficial to a probabilistic model. Given a commonsense knowledgebase, along with contextual information from these mobile devices, creating a classifier to determine gist of noisy transcriptions has been shown to be tractable (Eagle et al., 2003).

5. Preliminary Results

The system was initially deployed in six one-hour meetings over the course of two months. These meetings consisted of the same nine people: one professor (s9), three of his advisees (s2, s7, s8), and five additional graduate students (s1, s3, s4, s5, s6). Audio and interest levels were captured and streamed from the PDAs to a centralized server for processing. The voicing segments can be seen in Figure 2, plotted above the group's aggregate interest level.

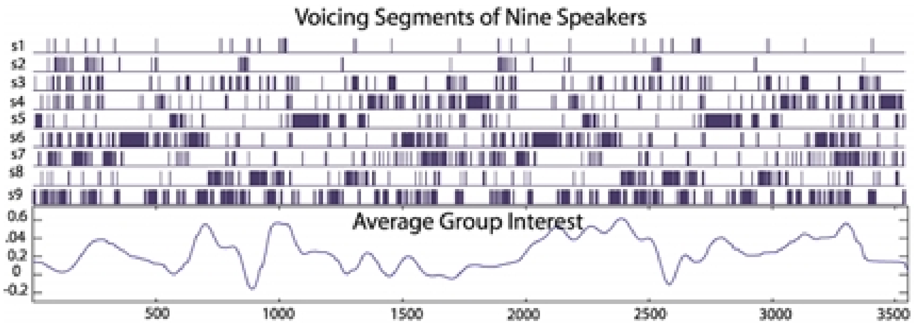

Fig. 3. A one-hour meeting with voicing segments mapped above aggregate interest level.

Each one-hour meeting consisted of almost 1 GB of data, primarily from the audio streamed from each headset microphone. These nine audio streams were chunked into 3,600 one-second blocks and analyzed using the metrics for speech and conversation detection described in Section 4. The features extracted from the audio streams included social behavior metrics for each individual including percentage of time talking, average comment length, most likely person to speak afterwards, closest person in physical proximity, average interest level, and the group's average interest when the individual is holding the floor. Unfortunately our interruption detector turned out to be ineffective for large group interactions due to the potential of more than one dialogue occurring simultaneously.

Speaker Number	Floor time (%)	Avg Comment (sec)	Nearest Neighbor	Transition (Speaker, %)	Avg Interest	Group Interest
s1	1.5	4.1	s8	s8-27	.21	.44
s2	2.2	2.2	s9	s9-47	.13	.36
s3	9.9	3.5	s9	s4-22	.20	.22
s4	11.4	9.6	s7	s6 -23	.05	.30
s5	12.8	8.8	s7	s9-37	.18	.33
s6	16.9	6.6	s4	s7-28	.09	.21
s7	10.1	6.6	s4	s9-30	.19	.24
s8	10.8	10.9	s1	s9-26	.40	.32
s9	24.4	6.9	s7	s6-22	.17	.25

Table 1. An analysis of the meeting from Figure 3

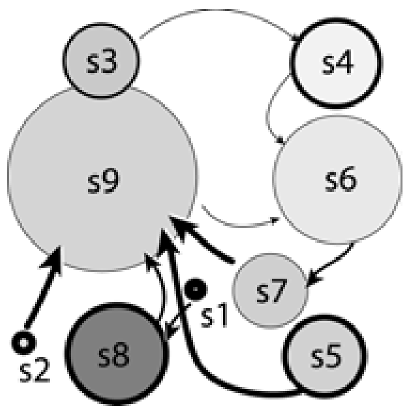

Fig. 4. A visual depiction of the group's dynamics from the meeting in Figure 3

What is striking about these preliminary results is how clearly the roles can be distinguished from the features listed above. The professor (s9) dominates the meeting while his advisees (s2, s7, s8) all concede the floor to him with relatively high probability - indicative of his influence (Atkinson & Heritage, 2003). Using topic-spotting methods mentioned in Section 4, work is also being done to find a correlation between how often a group incorporates the topics popular with an individual and the individual's influence within the group. Paradoxically, the interest levels of the group (represented by the circle borders in Figure 4) appear to be inversely proportional to the amount of time an individual holds the floor. When the most soft-spoken individuals (s1, s2) ventured to speak, they captured the highest interest level of the group, while those who dominated the meeting tended to get lower interest scores while holding the floor.

This type of analysis allows objective assessment of an individual's influence and contributions to the meeting, as well insight into the effectiveness of the group. Feedback to the most vocal speakers can be used to encourage them to share the floor with others. The more soft-spoken participants whose comments are appreciated by the group now have a means of receiving recognition of their contribution. Patterns in the behavior of dyadic pairs over time can be even more telling. Just as Allen was able to identify roles such as 'gatekeepers' in organizations by mapping social network information over an extended period of time (Allen, 77), the Reality Mining system has the capacity to capture and analyze vast amounts social information without surveys. Information about the people who interrupt, sit next to, or yield the floor to others, provides data that can be directly correlated with relationship.

6. Future Work

We have shown that it is possible to capture the tacit dynamics of a collocated team with existing handheld computers. However, it remains to be seen if this information will make the tie between a distributed member of that team more cohesive. We have initially tested the system on research groups within our lab, and are now planning a study on a collocated subset of a distributed team. Information about dynamics of the collocated team, similar to the results shown in Section 5, will be sent to peripheral team members who have had minimal face-to-face encounters with the team. We will give them daily information about who spent time with who, who interrupted whom, how the speaker transitions played out, which people the group found most interesting, and the system's assessment of each individual's current influence. These types of reports summarize some of the tacit social information that occurs throughout the day - information that typically never gets disseminated except in informal emails and chitchat, but typically contains the very items many people find most interesting. We will conduct a round of surveys each week monitoring the communication behavior of the peripheral members and their perceived cohesiveness with the group.

7. Conclusions

This work has created a foundation that has already exceeded the scope we initially envisioned. Social network analysts will soon have the potential to generate datasets never before possible in the realm of the social sciences. Architects have expressed interest in using this system to quantify how subtle changes to the interior of a workplace affect face-to-face communications. The measures we are currently developing to spot topics and identify questions enable a novel knowledge management system. Instead of surveys or mining static documents, this system can harness the information within face-to-face conversations by the water cooler, in a hallway, or at a meeting. Characterizing each person in an organization by local links to others, the questions they ask, and those that are posed to them, creates what is essentially a directed graph. This property allows the network to be queried for authoritative sources of information in a similar way internet search engines do today.

This research demonstrates our ability to capture extremely rich data on everyday human behavior, including interactions with others, movement, location, and activities, using hardware already worn daily by millions. In anticipation of these datasets, we are currently developing sophisticated probabilistic models that we hope will shed insight into the underlying parameters that govern group behavior. We are now instrumenting group activities such as negotiation, brainstorming, and weekly group meetings to derive relationship information, and using this information in controlled experiments to measure the extent to which it can be leveraged to create more effective distributed teams. Modern organizations will inevitably try to leverage their existing ubiquitous wearable computing infrastructures. Our system provides a testbed and baseline to build applications designed to capture a quantity of social information never before possible.

Acknowledgements

The authors would like to thank Jon Gips for developing the Zaurus interest poll application and Jonathon Cummings for sharing his knowledge about distributed teams. This work was partially supported by the NSF Center for Bits and Atoms (NSF CCR-0122419).

References

Allen, T. J. (1977). Managing the Flow of Technology, MIT Press, Cambridge, MA.

Atkinson, J., Heritage, J. (2003). Structures of Social Action: Studies in Conversation Analysis, Cambridge University Press, Cambridge, UK.

Basu, S. (2002). Conversation Scene Analysis, in Dept. of EECS. Cambridge, MIT

Choudhury, T., Clarkson, B., Basu, S., and Pentland, A.(2003). Learning Communities: Connectivity and Dynamics of Interacting Agents. To appear in the Proceedings of the International Joint Conference on Neural Networks - Special Session on Autonomous Mental Development. Portland, Oregon.

Eagle, N., Singh, P., Pentland, S. (2003). Gisting Conversations with Common Sense and Mobile Computing to appear in the Artificial Intelligence, Information Access, and Mobile Computing Workshop at the 18th International Joint Conference on Artificial Intelligence (IJCAI). Acapulco, Mexico.

Feldman Barrett, L., & Barrett, D. J. (2001). Computerized experience-sampling: How technology facilitates the study of conscious experience. Social Science Computer Review, 19, 175-185.

Handy, C. (1995). Trust and the virtual organization. Harvard Business Review, 73 (3), 40-50.

Hinds, P. J., and Bailey, D. E. (2000). Virtual team performance: Modeling the impact of geographic and temporal virtuality. Paper presented at the Academy of Management annual meeting, August,Toronto.

Iacono, S., & Weisband, S. (1997). Developing Trust in Virtual Teams, Proceedings of the 30th Hawaii International Conference on System Sciences. Maui, Hawaii.

Jarvenpaa, S., & Leidner, D. (1999). Communication and trust in global virtual teams. Organization Science, 10(6), 791-815.

Kiesler, S., & Cummings, J. (2002). What do we know about proximity and distance in work groups? In P. Hinds, & Kiesler, S. (Ed.), Distributed work (pp. 57-80). Cambridge, MA: MIT Press.

Maznevski, M., & Chudoba, C. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. Organization Science, 11(5), 473-492.

Singh, P.: The public acquisition of commonsense knowledge. Proceedings of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access. Palo Alto, CA: AAAI. (2002)

Sproull, L., & Kiesler, S. (1991). Connections: New ways of working in the networked organization. Cambridge, MA: MIT Press.

Stough, S., Eom, S. and Buckenmyer, J. (2000). Virtual teaming: a strategy for moving your organization into the new millennium, Industrial Management and Data Systems, vol. 100, no. 8, pp. 370-387.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. Communication Research, 23 (1), 1-43.

Walther, J. B. (1997). Group and interpersonal effects in international computer-mediated collaboration. Human Communication Research, 23 (3), 342-369.

Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press.