Submitted to the Fifth International Conference on Ubiquitous Computing (UbiComp). Seattle, WA. October 2003.

Social Network Computing

Nathan Eagle and Alex (Sandy) Pentland

MIT Media Lab, 20 Ames St, Cambridge, MA 02319
{nathan, sandy@media.mit.edu}
http://www.media.mit.edu

Abstract. A ubiquitous wearable computing infrastructure is now firmly entrenched within organizations across the globe, yet much of its potential remains untapped. This paper describes how the handheld computers and mobile phones in today's organizations can be used to quantify face-to-face interactions and to infer aspects about a user's situation, enabling more creative and transparent functioning of human organizations.

1 Introduction

Ubiquitous wearable computing has arrived in today's knowledge organizations. Handheld computers and mobile phones have been adapted as standard corporate attire across the globe. And the potential functionality of this new business uniform is dramatically increasing. Personal Digital Assistants (PDAs)[1], once computationally limited to storing calendar and contact information, now have wireless network connectivity and run at the speeds comparable to the desktop computers just a couple years ago. There are thousands of organizations comprised of millions of individuals who currently carry wireless transceivers, microphones, and three times the computational horsepower of an Intel Pentium I processor *in their pocket*.

Parallel to this wearable computing infrastructure lies what this paper refers to as an organization's social infrastructure. Although the traditional 'org chart' is meant to reflect the scaffolding of a social infrastructure, hierarchical job titles do very little to characterize an organization's underlying complex human network. More indicative of social infrastructure is the wisdom accumulated throughout an employee's extended career within an organization, 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.

The ubiquitous computing infrastructure within the workplace has the potential to augment organizational functioning by making social infrastructure more transparent. Using custom analytic software and a mobile computing infrastructure consisting of thirty wireless, linux-based handheld computers and an 802.11b network, we have created a testbed that we are using to learn how to make organizations that are more creative, efficient, and open.

[1] Throughout this paper, PDA is used interchangeably with the term 'handheld computer'.

2 Quantifying Face-to-Face Interactions within the Workplace

The social network research community has used survey data almost exclusively to establish the relationships between individuals within organizations. Despite the introduction of web-based surveys or experience sampling methods using surveys on handheld computers [7], the fundamental flaws inherent in self-report survey data remain: data bias and sparsity.

With the advent of corporate email and instant messaging, behavior measurement techniques are augmenting the surveys, enabling new social network datasets that require less direct participation from the participants [9]. Similarly, telephone logs can be analyzed to gain insight into the relationships between coworkers.

Quantifying the face-to-face interactions within an office environment is of particular interest, especially because complex information is rarely transmitted in an office environment by any other means [1]. If an individual requires a complex piece of knowledge from a colleague, he would use the telephone or email to set up a meeting, but then receive the information through a face-to-face interaction. Even outside the context of meetings, informal face-to-face conversations in the hall or by the water cooler are incredibly important for organizations [5]. Effectively harnessing this face-to-face communication channel has the potential to revolutionize the field of knowledge management.

Previous work at quantifying face-to-face interactions has been mainly with 'badges' that use infrared and RF to track individuals and their meetings. Choudhury and Pentland, for instance, built a shoulder-mounted 'sociometer' that incorporated IR, sound level and accelerometers in order to track interactions within an organization [4].

To move from social network mapping to social network function, we must also capture information about discussion content and context. Our system accomplishes this by capturing and analyzing each participant's audio, annotating the audio with subjective user feedback, extracting keyword-based topic and context information, and using audio and 802.11b data to establish location and other participants in local proximity. As shown in Table 1, an analysis of synchronized audio streams from coworkers can provide insight into individual social behavior as well as the efficiency of the collective.

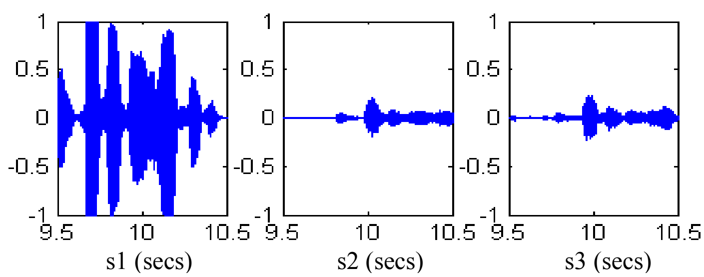
3 The Reality Mining System

Mirroring the ubiquitous wearable computing infrastructure in the modern workplace, the Reality Mining system is a combination of commercial hardware running specialized software. Thirty 802.11b-enabled PDAs containing standard personal information management applications were augmented with the ability to continuously stream and store audio and establish the proximity of others. The largest benefits of the system are realized as it scales. Detailed information regarding the dynamics of the face-to-face communication within the workplace can be quantified and correlated with the roles individuals play in an organization's social infrastructure.[2]

The heart of the Reality Mining system is the Sharp Zaurus. These linux-based, 206 MHz handheld computers were equipped with 802.11b CF cards and 256 MB of storage. Audio was captured from a variety of wired and wireless mobile phone headsets and lapel microphones that connect to the Zaurus through its audio jack. For all day use, interchangeable 1850 mAh battery packs were plugged into the AC adapter jack.

Applications for the Zaurus were created to record audio continuously, storing it locally until it could be transmitted to a server over an available 802.11b network. 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. More interactive applications were written for meeting analysis that accumulated and displayed aggregate interest statistics in real-time.[3] On the server-side, conversation detection, analysis, and inference software were written to process multiple large audio files in parallel.[4]

Conversation Detection. Our speech detection algorithm incorporated a variation of a multi-band center clipper.[5] 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 over a short window to verify that the labeled speech regions were not due to another participant speaking loudly.



[2] For example, 'gatekeepers', a term commonly found in the Organizational Behavior literature, are the connectors in an organization - typically the employees who are know most about the social infrastructure. They can be easily identified on a social network map as they have overwhelmingly more internal and external links than the average individual. [2]

[3] Thanks to Jonathan Gips for his work developing the Zaurus interest poll application.

[4] Streamed at 22KHz 16-bit, each person has a daily audio file of approximately 1 GB.

[5] Thanks to Mohan Sondhi of Avaya Labs for advice about designing a light weight speech detection algorithm

Fig. 1. These three waveforms sections are strongly correlated (*c*>*.25*), indicating that s2 and s3 are within earshot of the speaker (s1). An analysis of the relative energy between the waveforms can yield insight into the physical proximity of the speakers. It is also apparent from this figure that the audio synchronization is slightly off between s2 and s3

Establishing accurate vocalization labels is only the first part of conversation detection. The next step is to determine the proximity of the participants. This can be initially accomplished by comparing the access points to which the participants are streaming audio. The audio segments of participants who are near similar access points are then correlated to determine whether they are within earshot of each other. Essentially the concept uses the principle that a speaker's voice will not only be recorded in his own microphone, but also at a lower energy level in the microphones of the people around him. However, if there is a high correlation between two audio segments, this is not yet substantial evidence of a conversation. A correlation would occur even when two adjacent users are having separate conversations on mobile phones, or separate dialogues with individuals not using the Reality Mining system. As shown in Figure 2 and described in [3], the voicing segment of one participant is the noisy compliment of the other participants' voicing segments. Measuring the mutual information between these binary streams has been shown to be indicative of an actual conversation. [6]

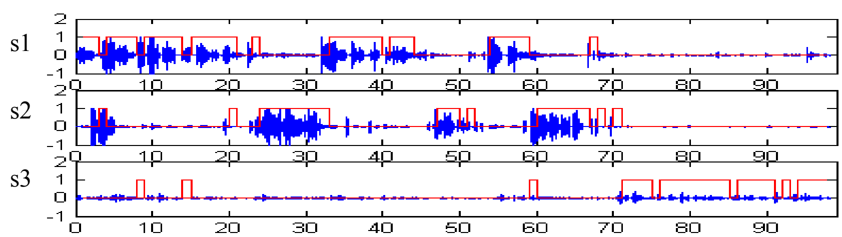


Fig. 2. The three speaking segments have high mutual information (*MI*>*.3*), indicative of a conversation between the speakers. It can be see that each voicing segment is the noisy complements of the remaining two

Conversation Analysis. Once detected, the audio streams of a conversation 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, energy, duration, participants, interruptions, transition probabilities, time spent holding the floor [3], and annotated by interest metrics. By comparing relative volume levels of a speaker's voice in multiple microphones, it even becomes possible to infer proximity of the participants to an approximate degree.

Throughout the meeting, the PDAs were serving a dual purpose. While streaming audio and wireless network information to a central server, the handheld computers

---

[6] Mutual information was also initially used to calculate approximate alignment between the audio streams.

were also enabling a user to input his or her interest level through an interface designed to minimize distraction.

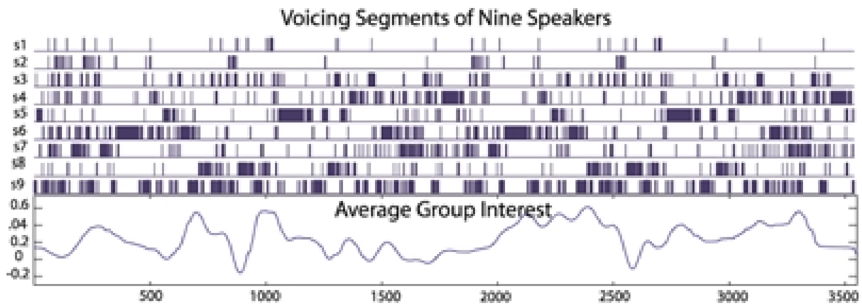


Fig. 3. The voicing segments and interest level of a one-hour meeting

Table 1. Meeting Analysis. A one-hour meeting in which the participants were wearing the Reality Mining system, streaming audio to a central server and simultaneously recording their interest level

Participant	Speaking time (%)	Avg (sec) Comment	Nearest Neighbor	Transition (Name, %)	Avg Interest	Group Interest
Ivan	1.5	4.1	Nathan	Nathan-27	.21	.44
Jon	2.2	2.2	Sandy	Sandy-47	.13	.36
Joost	9.9	3.5	Sandy	Jordan-22	.20	.22
Jordan	11.4	9.6	Mike S	Mike O-23	.05	.30
Leonel	12.8	8.8	Mike S	Sandy-37	.18	.33
Mike O.	16.9	6.6	Jordan	Mike S-28	.09	.21
Mike S.	10.1	6.6	Jordan	Sandy-30	.19	.24
Nathan	10.8	10.9	Ivan	Sandy-26	.40	.32
Sandy	24.4	6.9	Mike S	Mike O-22	.17	.25

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. Information about the people who interrupt, sit next to, or yield the floor to others, provides data that can be directly correlated with relationship. Using the topic-spotting methods described below, an individual's influence can also be correlated with how the group incorporates the topics popular with the individual.

Content and Context. The final component of the audio analysis is establishing the conversational situation: the topic and the surrounding context of a conversation. ViaVoice, a commercial speech recognition engine, is used to transcribe the audio streams, however its transcription accuracy often falls well below 40% for spontaneous speech recognition. For situation understanding, our system combines a network of commonsense knowledge with keywords and contextual information automatically obtained from the Zaurus. We make use of Push Singh's OpenMind network, containing over 250,000 commonsensical semantic relationships contributed from over 10,000 people across the web [8]. Despite this vast amount of data, the knowledge database can be compressed into fewer than 50 MB and easy stored locally on the PDAs. While the words the speech recognition engine gets correct tend to be grouped around neighboring semantically-related nodes, errors in the transcriptions turn out to be distributed randomly over this network. The nodes surrounding the largest clusters of keywords are assumed to be potential aspects of the speakers' situation. However, the robustness of the classifier comes from its ability to bias the prior probability of each node based on other contextual information from the PDAs, such as the user's location, conversation participants, or simply the people in his local proximity. Online learning algorithms incorporate subsequent observations into the classifier yielding a specialized model that better reflects an individual's behavior.

With only one correct word for every three, even a human would have a difficult time inferring the gist of a transcribed conversation. 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 becomes tractable [6].

Table 2. Conversational Inference. Two participants were standing in line, talking about what to order in the food court cafeteria. The situation classification with only the noisy transcript is shown in Table 2.1. Table 2.2 incorporates additional contextual information: the fact that the audio was streamed to the food court access point.

Table 2.1

Confidence	Classification with no context
5	Eat in restaurant
5	buy beer
5	talk with someone far away
5	eat in fast food restaurant
5	buy hamburger
4	go to hairdresser
4	wait in line

Table 2.2

Confidence	Classification with location context
27	eat in fast food restaurant
21	eat in restaurant
18	wait on table
16	you would go to restaurant because you
16	wait table
16	go to restaurant
15	know how much you owe restaurant

4 Applications

Once pocket-sized devices become more aware of the infrastructure in which they are part, a variety of exciting applications become possible. Three applications that are now being evaluated in classes at MIT include:

Meeting Miner: Participants continuously provide subjective feedback on comments and discussion using a 2D touch pad. The feedback interface converts the task of providing continuous feedback into a low-attention, secondary task. By correlating peaks in interest/approval with the individual audio inputs, the system can automatically provide a summary audio track consisting of comments that had high approval or interest ratings, and to employ speech analysis to identify topics that had high (or low) ratings.

OpinionMetrics: Subjective feedback is pooled and shared with the participants via a public display. Comments that give rise to wide variations in opinion cause the discussion to focus on the reason for disparate opinions, and controversial topics can be retrieved for further analysis and debate. Opinions and comments can also clustered using 'collaborative filtering', to display groupings of opinion, allowing within-group and between-group debate

GroupMapper: Dynamic maps of social infrastructure can be generated and publicly displayed to reflect the roles and dyadic relationships that individuals have within a work group. It is hoped that such analysis will help with such tasks as determining who to ask for help, identifying isolated cliques, and a gaining a deeper insight into the underlying dynamics of the organization. Architects have expressed interest in using this system to monitor how small changes to the interiors of buildings have an effect on the office communications.

Privacy. Although our system uses encryption and permissions to address some problems of privacy, significant concerns remain. To deploy the system at the enterprise level, additional work needs to be spent to assuage the concerns of the more privacy conscious. Quick modifications will include a ten-minute delete and temporary mute button. Another potential modification to the system would be to stream the audio to the participants' personal computers. At the end of each week, the conversation inference algorithms could be used to summarize a user's interactions, creating a list of the week's conversations including location, topic, people in proximity, and duration. Along with each interaction, would be a checkbox to mark if the conversation is private, public or should be permanently deleted. A private conversation would be stored only the user's computer as part of his private conversation diary. Relevant information from the public conversations would be uploaded to the central server for processing. The ability to have their entire weekly interactions quantified and displayed can provide insight into an individual's personal time management and may create enough value to justify the system on its own. This could be especially important for organizations of individuals who need to keep careful track of how they spend their time for billing purposes.

4 Future Research and Conclusion

This project 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. 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 teams and collaborations.

Such a data-driven model of social network functioning offers the potential to transcend the traditional org-chart, perhaps by drawing parallels to ad-hoc network optimization. Forming groups based on heretofore unrecognized inherent communication patterns rather than an orthodox hierarchy may yield significant insights to the organizational structure community. We believe that modern organizations will inevitably try to leverage their existing ubiquitous wearable computing infrastructures. Our system provides a testbed and baseline to build and demonstrate future social network applications.

Acknowledgements

This work was partially supported by the NSF Center for Bits and Atoms (NSF CCR-0122419).

References

1. Allen, T. J.: Organizational Structure for Product Development, Working Papers. (2002)
2. Allen, T. J.: Managing the Flow of Technology, MIT Press, Cambridge, MA (1977)
3. Basu, S: Conversation Scene Analysis, in Dept. of EECS. Cambridge: MIT (2002)
4. Choudhury, T., Clarkson, B., Basu, S., and Pentland, A.: 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. (2003)
5. Cohen, D., Prusak, L.: In Good Company: How Social Capital Makes Organizations Work, Harvard Business School Press, Cambridge, MA (2001)
6. Eagle, N., Singh, P., Pentland, S.: to be published in the Artificial Intelligence, Information Access, and Mobile Computing Workshop at the 18th International Joint Conference on Artificial Intelligence (IJCAI). Acapulco, Mexico. (2003)
7. Feldman Barrett, L., & Barrett, D. J.: Computerized experience-sampling: How technology facilitates the study of conscious experience. Social Science Computer Review, 19, 175-185. (2001)
8. 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)

9. Tyler, J. R., Wilkinson, D. M. & Huberman, B. A.: Email as spectroscopy: automated discovery of community structure within organizations. Preprint: http://xxx.lanl.gov/arXiv:cond-mat/0303264 (2003)