The *Sociometer*: A Wearable Device for Understanding Human Networks

Tanzeem Choudhury and Alex Pentland
Human Design Group
20 Ames Street, E15-384C
Cambridge, MA02139 USA
+1 617 253 0370
tanzeem@media.mit.edu

ABSTRACT
In this paper, we describe the use of the *sociometer*, a wearable sensor package, for measuring face-to-face interactions between people. We develop methods for learning the structure and dynamics of human communication networks. Knowledge of how people interact is important in many disciplines, e.g. organizational behavior, social network analysis and knowledge management applications such as expert finding. At present researchers mainly have to rely on questionnaires, surveys or diaries in order to obtain data on physical interactions between people. In this paper, we show how noisy sensor measurements from the sociometer can be used to build computational models of group interactions. Using statistical pattern recognition techniques such as dynamic Bayesian network models we can automatically learn the underlying structure of the network and also analyze the dynamics of individual and group interactions. We present preliminary results on how we can learn the structure of face-to-face interactions within a group, detect when members are in face-to-face proximity and also when they are having a conversation. We also measure the duration and frequency of interactions between people and the participation level of each individual in a conversation.

Keywords

Organizational behavior, social network analysis, wearable computing, Bayesian networks.

MOTIVATION AND INTRODUCTION

In almost any social and work situation our decision-making is influenced by the actions of others around us. Who are the people we talk to? For how long and how often? How actively do we participate in the conversations? Answers to these questions have been used to understand the success and effectiveness of a work group or an organization as a whole. Can we identify the differences between people's interactions? Can we identify the individuals who talk to a large fraction of the group or community members? Such individuals, often referred to the *connectors*, have an important role in information diffusion [1]. Thus, learning the connection structure and nature of communication among people are important in trying to understand the following phenomena: (i) diffusion of information (ii) group problem solving (iii) consensus building (iv) coalition formation etc. Although people heavily rely on email, telephone and other virtual means of communication, research shows that high complexity information is mostly exchanged through face-to-face interaction [2]. Informal networks of collaboration within organizations coexist with the formal structure of the institution and can enhance the productivity of the formal organization [3]. Furthermore, the physical structure of an institution can either hinder or encourage communication. Usually the probability that two people communicate declines rapidly with the distance between their work location [2, 4]. Being able to measure the relationship between communication networks and different environmental and organizational attributes will enable us to create better workplaces with improved communication and collaboration among their members.

We believe that wearable sensor data combined with pattern recognition techniques will play an important role in sensing and modeling physical interactions. These techniques can complement and augment existing manual techniques for data collection and analysis. The results can be used for understanding human communication patterns studied in organizational behavior and social network analysis. The knowledge of people's communication networks can also be used in improving context-aware computing environments and coordinating collaboration between group and community members.

SENSING AND MODELING HUMAN COMMUNICATION NETWORKS

As far as we know, there are currently no available methods to automatically model face-to-face interactions within a community. This absence is probably due to the difficulty of obtaining reliable measurements from real-world

interactions. One has to overcome the uncertainty in sensor measurements. This is in contrast to modeling virtual communities where we can get unambiguous measurements about how people interact – the duration and frequency (available from chat and email logs) and sometime even detailed transcription of interactions [5, 6].

We believe sensing and modeling physical interactions among people is an untapped resource. In this paper, we present machine-learning techniques that use wearable sensor data to make reliable estimates about a user's interaction state (e.g. who is she talking to, how long did the conversation last, etc.). We use these results to infer the structure and dynamic relationships that exists in groups of people. This can potentially be much cheaper and more reliable than human-delivered questionnaires. It is also more easily scalable to larger groups, and does not depend on personal recall or interpretation. Automatically discovering the high-level group structures within an organization can also provide a sound basis for then exploring more fine-grained group interactions using questionnaires or interviews.

In the following sections we will describe how we use wearable sensors to measure and build models of interactions. In summary, we are seeking to discover how information about social network relationships can be derived by applying statistical machine learning techniques to data obtained from wearable sensors. We hope to lay the groundwork for being able to automatically study how different groups within social or business institutions connect, understand how information propagates between these groups and analyze the effects of new policy or new technology on the group structure.

EXPERIMENTAL DETAILS

The first step towards reliably measuring communication is to have sensors that can capture interaction features. For example, we need to know who is talking to whom, the frequency and duration of communication. To record the identity of people in an interaction, we equip each person with an infra-red (IR) transceiver that sends out unique ID for the person and receives ID from other people in her proximity. We use microphones to detect conversations.

In this section we describe a pilot experiment we have recently completed in our lab. A group of eight people at the MIT Media Lab agreed to wear the sociometer – the wearable sensor that measures social interactions. It is an adaptation of the hoarder board, a wearable data acquisition board, designed by the wearable computing group at the Media lab [7]. The sociometer is especially packaged by wearable designer Brian Clarkson [8] for the comfort of the wearer, aesthetics, and placement of sensors that are optimal in getting reliable measurements of interactions. The users have the device on them for six hours a day (11AM –5PM) while they are on MIT campus. We collected 10 days (two full work weeks) of data from each subject, which amounts to 60 hours of data per subject.

The Sociometer

The sociometer has an IR transceiver, a microphone, two accelerometers, on-board storage, and power supply. The wearable stores the data locally on a 256MB compact flash card and is powered by four AAA batteries. A set of four AAA batteries is enough to power the device for 24 hours. Everything is packaged into a shoulder mount so that it can be worn all day without any discomfort.

The sociometer stores the following information for each individual

(i) Information about people nearby (sampling rate 17Hz – sensor IR)
(ii) Speech information (8KHz - microphone)
(iii) Motion information (50Hz - accelerometer)

Other sensors (e.g. light sensors, GPS etc.) can also be added in the future using the extension board. For this paper we do not use the data obtained from the accelerometer.

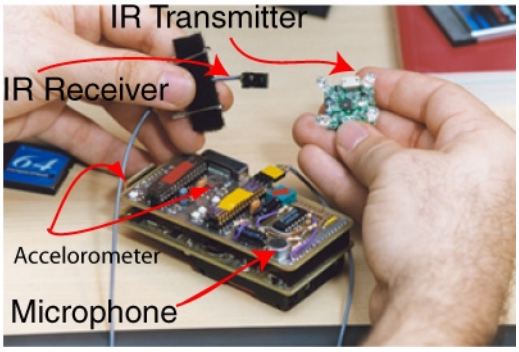

Figure 1 - The wearable sensor board

The success of IR detection depends on the line-of-sight between the transmitter-receiver pair. The sociometer has four low powered IR transmitters. The use of low powered IR transmitters is optimal because (i) we only detect people in close proximity as opposed to far apart in a room (as with high-powered IR) and (ii) we detect people who are facing us and not people all around us (as with RF transmitter). The IR transmitters in the sociometer create a cone shaped region in front of the user where other sociometers can pick up the signal. The range of detection is approximately six feet, which is adequate for picking up face-to-face communication. The design and mounting of the sociometer places the microphone six inches below the wearer's mouth, which enables us to get good audio without a headset. The shoulder mounting also prevents clothing and movement noise that one often gets from clip-on mics. All of the eight users were very satisfied with the comfortable and aesthetic design of the device. None of the

subjects complained about any inconvenience or discomfort from wearing the device for six hours everyday.

Despite the comfort and convenience of wearing a sociometer, we are aware that subject's privacy is a concern for any study of human interactions. Most people are wary about how this information will be used. To protect the user's privacy we agree only to extract speech features, e.g. energy, pitch duration, from the stored audio and never to process the content of the speech. But, to obtain ground truth we need to label the data somehow. Our proposed solution is to use garbled audio instead of the real audio for labeling. Garbled audio makes the audio content unintelligible but maintains the identity and pitch of the speaker [9]. In future versions of the sociometer we will store encrypted audio instead of the audio, which can also prevent unauthorized access to the data.

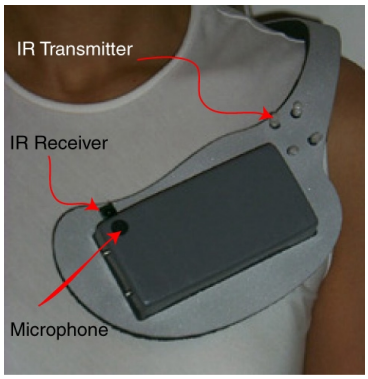

Figure 2 - The shoulder mounted sociometer

Data Analysis Methods and Preliminary Results

The first step in the data analysis process is to find out when people are in close proximity. We use the data from the IR receiver to detect proximity of other IR transmitters. The receiver measurements are noisy – the transmitted ID numbers that the IR receivers pick up are not continuous and are often bursty and sporadic. The reason for this bursty signal is that people move around quite a lot when they are talking, so one person's transmitter will not always be within the range of another person's receiver. Consequently the receiver will not receive the ID number continuously at 17Hz. Also each receiver will sometimes receive its self ID number. We pre-process the IR receiver data by filtering out detection of self ID number as well as propagating one IR receiver information to other nearby receivers (if receiver #1 detects the presence of tag id #2, receiver #2 should also receive tag id #1). This pre-processing ensures that we maintain consistency between different information channels. However, we still need to able use the bursty receiver measurements to detect the contiguous time chunks (an episode) people are in proximity. Two episodes always have a contiguous time chunk in between where no ID is detected. A hidden Markov model (HMM) [10] is trained to learn the pattern of IR signal received over time. Typically an HMM takes noisy observation data (the IR receiver data) and learns the temporal dynamics of the underlying hidden node and its relationship to the observation data. The hidden node in our case has binary state - 1 when the IDs received come from the same episode and 0 when they are from different episodes. We hand-label the hidden states by labeling 6 hours of data. The HMM uses the observation and hidden node labels to learn its parameters. We can now use the trained HMM to assign the most likely hidden states for new observations. From the state labels we can estimate the frequency and the duration that two people are within face-to-face proximity. Figure 3 shows five days of one person's proximity information. Each color in the sub-image identifies a person to whom the wearer is in close proximity of and the width is the duration contact. Note that we are also able to detect when multiple people are in close proximity at the same time.

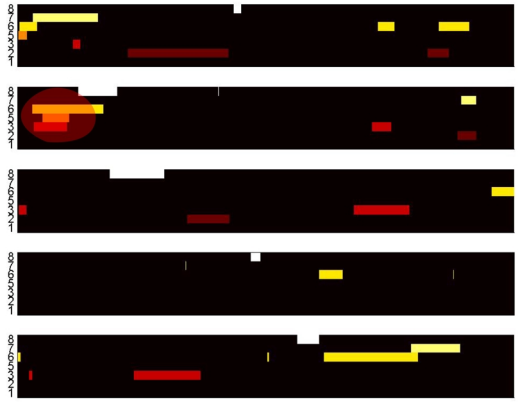

Figure 3 - Proximity information for person # 1. Each sub-image shows one day's information

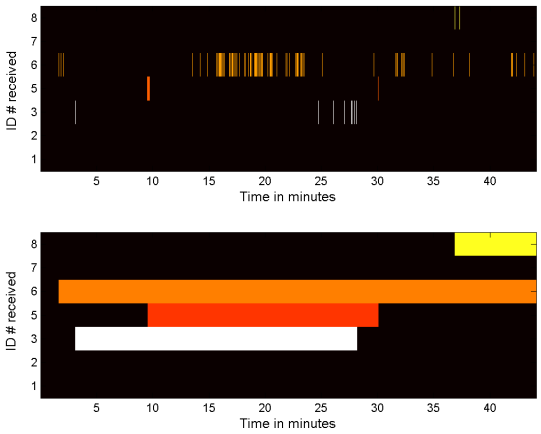

Figure 4 - Zoomed into the red shaded region from day two in figure 3. Upper panel: Bursty raw data from IR receiver. Lower Panel: Output of the HMM which groups the data into contiguous time chunks.

The IR tag can provide information about when people are in close face-to-face proximity. But it provides no information about whether two people are actually having a conversation. They may just have been sitting face-to-face during a meeting. In order to identify if two people are actually having a conversation we first need to segment out the speaker from all other ambient noise and other people speaking in the environment. Because of the close placement of the microphone with respect to the speaker's mouth we can use simple energy threshold to segment the speech from most of the other speech and ambient sounds. It is been shown that one can segment speech using voiced regions (speech regions that have pitch) alone (ref sumit). In voiced regions energy is biased towards low-frequency range and hence we threshold low-energy threshold (2KHz cut off) instead of total energy. The output of the low-frequency energy threshold is passed to another HMM as observation, which segments speech regions from non-speech regions. The states of hidden node are the speech chunks labels (1 = a speech region and 0 = non-speech region). We train our HMM on 10 minutes of speech where the hidden nodes are again hand labeled.

Figure 5 shows the segmentation results for a 35 second audio chunk. In this example two people wearing sociometers are talking to each other and are interrupted by a third person (between t=20s and t=30s). The output of low frequency energy threshold for each sociometer is fed into the speech HMM which segments the speech of the wearer. The red and green lines overlaid on top of the speech signal show the segmentation boundaries for the two speakers. Also note that the third speaker's speech in the 20s-30s region is correctly rejected, as indicated by the grayed region in the figure. In Figure 6 we show the spectogram from the two speakers' sociometers overlaid with the results from the HMM.

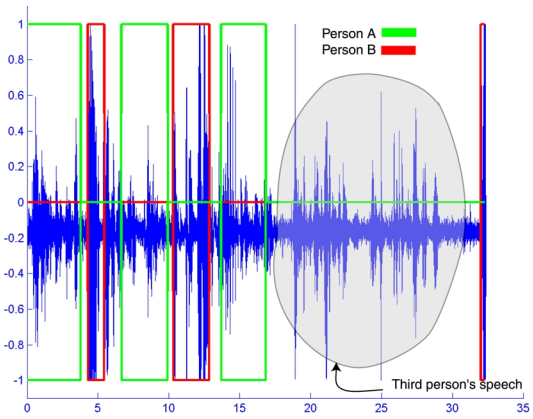

Figure 5 - Speech segmentation for the two subjects wearing the sociometer.

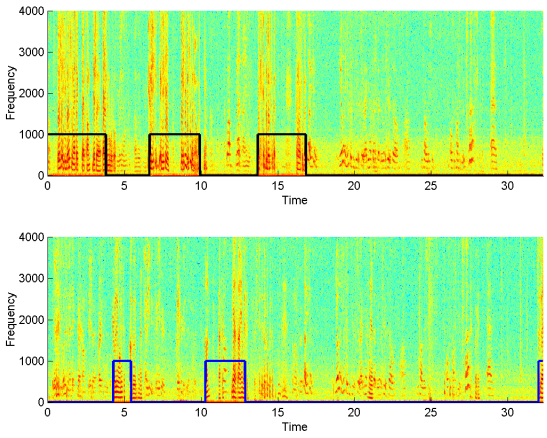

Figure 6 - Spectogram from person A and person B's microphone respectively with the HMM output overlaid in black(top) and blue(bottom).

Now we have information about when people are in close proximity and when they are talking. When two people are nearby and talking, it is highly likely that they are talking to each other, but we cannot say this with certainty. Recent results presented in the doctoral thesis of Sumit Basu[11] demonstrate that we can detect whether two people are in a conversation by relying on the fact that the speech of two people in a conversation is tightly synchronized. Basu reliably detects when two people are talking to each other by calculating the mutual information of the two voicing streams, which peaks sharply when they are in a conversation as opposed to talking to someone else. We are in the process of using his techniques for detecting if two people talking in close proximity are actually talking to each other.

Once we detect the pair-wise conversation chunks we can estimate the duration of conversations. We can further break down the analysis and calculate how long each person talks during a conversation. We can measure the ratio of interaction, i.e. (duration of person A's speech)**:**(duration of person B's speech). We can also calculate what fraction of our total interaction is with people in our community, i.e. inter vs. intra community interactions. This may tell us how embedded a person is within the community vs. how much the person communicates with other people. For example, someone who never talks to his work group but has many conversations in general is very different from someone who rarely talks to anyone.

A first pass picture of the network structure can be obtained by measuring the duration that people are in close face-to-face proximity. Figure 7 shows the link structure of our network based on duration, i.e. the total length of time spent in close proximity. There is an arrow from person A to person B if the duration spent in close proximity to B accounts for more than 10% of A's total time spent with everyone in the network. The thickness of the arrow scales with increasing duration. Similarly, Figure 8 shows the link

structure calculated based on frequency, i.e. the number of times two people were in close proximity. We are currently working on combining the audio and IR tag information and re-estimating the link the structure. Then we will be able look at network structure along various dimensions – i.e. based on frequency and duration of actual conversations people have with each other. We can also analyze the structure based on the dynamics of interaction (e.g. interaction ratio) we mentioned earlier in the section.

There are a few interesting points to note about differences in the structure based on duration vs. frequency. The two main differences are that in the frequency network there are links between ID #1 and ID # 7 and there are extra links connecting ID #6 to many more nodes than the duration network. The additional link to ID # 6 was created because person # 6 sat mostly in the common space through which every one passed through frequently. Consequently, ID# 6 was picked up by most other receivers quite often, but the duration of detection was very short. However, if we combined IR with the presence of audio these extra links would most likely disappear. But, the links between ID 1 an ID 7 are more interesting – although these two people never had long discussions they quite often talked for short periods of time. These links would probably remain even when combined with audio.

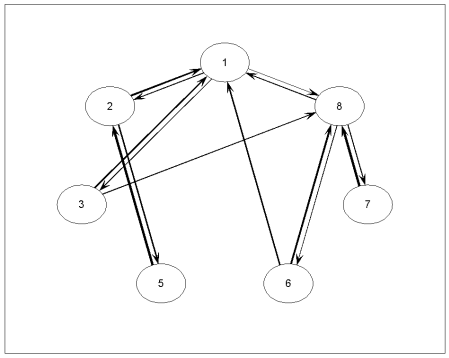

Figure 7 - The link structure of the group based on proximity duration.

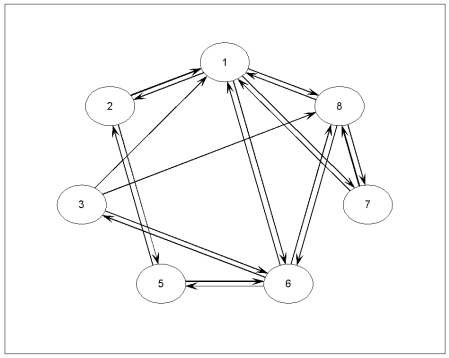

Figure 8 - The link structure of the group based on proximity frequency.

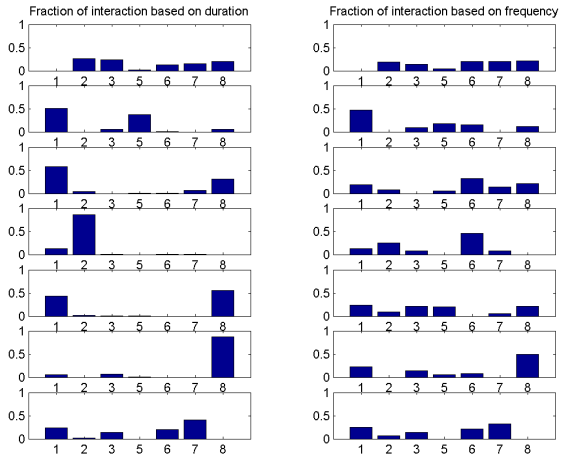

Figure 9 – Interaction distribution based on proximity duration (first column) and proximity frequency (second column). Each row shows results for a different person in the network

Figure 9 shows the fraction of time each individual spends with other members in the group based on duration and frequency. Person 1 talks to all other members regularly and is the most connected person as well (see Figure 7 and Figure 8). Person 2-6 have more skewed distribution in the amount of time they spend with other members, which means they interact mostly with a select sub-group of people. These are only a few examples of looking at different characteristics of the network. Analysis along different dimensions of interaction is going to be one the main advantage of sensor-based modeling of human communication networks.

DISCUSSION AND CONCLUSION

In this paper we have presented preliminary results from our efforts in sensor-based modeling of human communication networks. We show that we can automatically and reliably estimate when people are in close proximity and when they are talking. We demonstrate the advantage of continuous sensing of interactions that allows us to measure the structure of communication networks along various dimensions – duration, frequency, ratio of interaction etc. We are working on combining the proximity and audio information channels and obtaining quantitative results for our algorithms by comparing the accuracy of our algorithms to hand-labeled ground truth data of the interactions. We are also working on modeling the evolution and dynamics of the network as a whole and quantitatively measuring the influences people have on each other [12]. Within the next two months we also plan to scale-up our experiment to include a group of 25 people who belong to different research groups and different physical locations within the MIT campus. We can then begin to model how information propagates between groups in a community or an organization and analyze the effects of new policies or new technologies on the dynamics of the group.

ACKNOWLEDGMENTS

The authors would like to especially thank Brian Clarkson for designing the shoulder mount for the sociometer. Thanks to Sumit Basu whose work on Conversational Scene Analysis has guided our work on audio processing.

Also thanks to Leonardo Villarreal who spent many hours making and testing the 25 sociometers.

REFERENCES

1. Gladwell, M., *The Tipping Point: How little things make can make a big difference*. 2000, New York: Little Brown.
2. Allen, T., *Architecture and Communication Among Product Development Engineers*. 1997, Sloan School of Magement, MIT: Cambridge. p. pp. 1-35.
3. Huberman, B. and Hogg, T., *Communities of Practice: Performance and Evolution.* Computational and Mathematical Organization Theory, 1995. **1**: p. pp. 73-95.
4. Allen, T., *Organizational Structure for Product Development*. 2000, Sloan School of Management, MIT: Cambridge. p. pp 1-24.
5. Adar, E., Lukose, R., Sengupta, C., Tyler, J., and Good, N., *Shock: Aggregating Information while Preserving Privacy*. 2002, HP Laboratories: Information Dynamics Lab.
6. Gibson, D., Kleinberg, J., and Raghavan, P. *Inferring Web communities from link topology*. In *Proc. 9th ACM Conference on Hypertext and Hypermedia*. 1998.
7. Gerasimov, V., *Hoarder board*. 2002. http://vadim.www.media.mit.edu/Hoarder/Hoarder.htm.
8. Clarkson, B., *Wearble Design Projects*. 2002. http://web.media.mit.edu/~clarkson/.
9. Marti, S., Sawhney, N., Jacknis, M., and Schmandt, C., *Garble Phone: Auditory Lurking*. 2001.
10. Jordan, M. and Bishop, C., *An Introduction to Graphical Models.* 2001.
11. Basu, S., *Conversation Scene Analysis*, in *Dept. of Electrical Engineering and Computer science*. Doctoral. 2002, MIT. p. pp 1-109.
12. Basu, S., Choudhury, T., Clarkson, B., and Pentland, A. *Towards Measuring Human Interactions in Conversational Settings.* In *In the proceedings of IEEE Int'l Workshop on Cues in Communication at the Computer Vision and Pattern Recognition (CVPR) Conference*. 2001. Kauai, Hawaii.