Sensing and Modeling Human Networks

by

Tanzeem Khalid Choudhury

B.S., Electrical Engineering,
University of Rochester (1997)
S.M., Media Arts and Sciences,
Massachusetts Institute of Technology (1999)

Submitted to the Program of Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirement for the degree of

DOCTOR OF PHILOSOPHY IN MEDIA ARTS AND SCIENCES
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2004

Signature of Author____

Program in Media Arts and Sciences
August 28, 2003

Certified by____

Alex P. Pentland
Toshiba Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by____

Andrew B. Lippman
Chairman, Departmental Committee on Graduate Studies
Program in Media Arts and Sciences

Sensing and Modeling Human Networks

by

Tanzeem Khalid Choudhury

Submitted to the Program of Media Arts and Sciences,
School of Architecture and Planning,
on August 28, 2003 in partial fulfillment of the
requirement for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

Knowledge of how groups of people interact is important in many disciplines, e.g. organizational behavior, social network analysis, knowledge management and ubiquitous computing. Existing studies of social network interactions have either been restricted to online communities, where unambiguous measurements about how people interact can be obtained (available from chat and email logs), or have been forced to rely on questionnaires, surveys or diaries to get data on face-to-face interactions between people.

The aim of this thesis is to automatically model face-to-face interactions within a community. The first challenge was to collect rich and unbiased sensor data of natural interactions. The "sociometer", a specially designed wearable sensor package, was built to address this problem by unobtrusively measuring face-to-face interactions between people. Using the sociometers, 1518 hours of wearable sensor data from 23 individuals was collected over a two-week period (66 hours per person).

This thesis develops a computational framework for learning the interaction structure and dynamics automatically from the sociometer data. Low-level sensor data are transformed into measures that can be used to learn socially relevant aspects of people's interactions – e.g. identifying when people are talking and whom they are talking to. The network structure is learned from the patterns of communication among people. The dynamics of a person's interactions, and how one person's dynamics affects the other's style of interaction are also modeled. Finally, a person's style of interaction is related to the person's role within the network. The algorithms are evaluated by comparing the output against hand-labeled and survey data.

Thesis Supervisor: Alex P. Pentland
Title: Toshiba Professor of Media Arts and Sciences

Sensing and Modeling Human Networks

By

Tanzeem Khalid Choudhury

Thesis Committee:

Advisor:____

Alex P. Pentland
Toshiba Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Thesis Reader:____

Thomas J. Allen
Howard W. Johnson Professor of Management
Sloan School of Management
Massachusetts Institute of Technology

Thesis Reader:____

Michael Kearns
Professor, Computer and Information Science
Co-Director, Institute for Research in Cognitive Science
University of Pennsylvania

DEDICATION

TO ABBU

YOU MADE IT POSSIBLE FOR ME TO LIVE THE LIFE I IMAGINED.

Acknowledgments

During the six years I have been at MIT, I have been influenced by many people around me, who have shaped the way I think and taught me a lot about how to do research.

Among those, I want to first thank Sandy for providing me with guidance and support, and always the encouragement to try something new and different. Thanks to the rest of my committee members, Prof. Tom Allen and Prof. Michael Kearns, for the valuable advice they have given me and the time and effort they have put into this process.

Next I would like to thank the amazing group of graduate students with whom I have had the good fortune to work – the Pentlandians. Of them, two deserve very special thanks – Sumit Basu and Brian Clarkson. They have been amazing friends and the greatest colleagues. My grad school experience would not be the same without them. I have learned so much from our discussions, on topics ranging from machine learning and human behavior modeling to career choices, life and love.

I would like to express my appreciation and gratitude to the following people who have either contributed towards or inspired the work presented in this thesis. Vadim Gerasimov, Rich DeVaul and Josh Weaver designed the hardware we used for the data collection. Sumit's thesis on 'Conversational Scene Analysis' has guided some of the speech and conversation detection work. Brian's brilliant design skills were invaluable in building the sociometer. The influence model learning algorithms described in Chapter 3 was developed jointly with Sumit and Brian. Thanks to my UROP Leonardo Villarreal for spending long hours in the machine shop with me building the sociometers. This thesis would not have been possible without the subjects who trusted me and who so willingly participated in my experiment. Thank you for your help.

Thanks to Jim Rehg, who has been a great mentor to me and always a source of good advice. And of course my friends who have given me the strength and support and with whom I had many thought provoking conversations. Thank you Normin, for being my friend for so many years now. Jen, Natalia, Vikram, Raul, Ashish, Cynthia, Mathilee, Tania, and Marina – you have all enriched my life with your different perspectives on life. Special thanks to Apu who has never failed to be there for me. My parents who have loved me and supported me unconditionally – thank you so much. I can not express in words the gratitude I have for my dad, who has made it possible for me to follow my dreams and taught me to never be deterred by obstacles. I would like to dedicate this work to you - Abbu. And finally thanks to Raj, who makes my life so much more interesting, fun, stimulating and full of love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

"*Social scientists, particularly economists, have a fatal attraction for working on theoretical propositions with mathematical models and avoiding the real world.*"

Richard Cyert (1994)

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 we have? Can we identify people's interactions automatically? Can we identify the individuals who talk to a large fraction of the group or community members? Such individuals, often referred to the connectors, play an important role in information diffusion [1]. Answers to these questions have been used to understand the success and effectiveness of a work group or an organization as a whole. Thus, knowing the connection structure and nature of communication among people within a community 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 interactions [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. The knowledge of people's communication patterns can also be used in improving context-aware computing environments that is not only aware of the user but also how she is embedded in the larger collective.

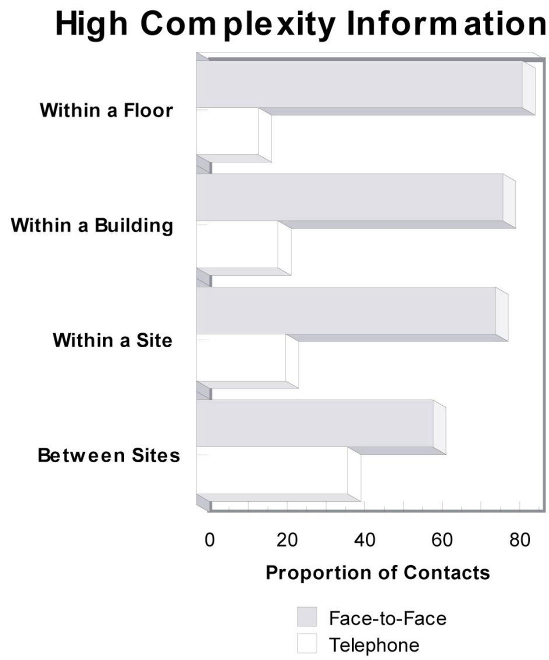

Figure 1-1: Role of from face-to-face interaction in high complexity information exchange (from Prof. Allen's paper [2])

To date, most of the research in social network analysis and organizational communication that analyze interaction patterns are based on self-reports. However, studies show that self-reports correspond poorly to communication behavior as recorded by independent observers [5]. Getting observational data on communication is important in the studies of networks in organizations and other social collectives. But human interactions are diffuse and unpredictable and thus very difficult to observe directly. Existing methods that use humans observers to record communication in groups are impractical, as it would require observers to be placed in every possible location conversations may occur, or have observers follow every member of the group. The observation of face-to-face verbal message exchange behavior in large collective groups can only be made practical if an acceptable replacement for human coders is found.

As wearable computers become more and more integrated into our daily life it provides the opportunity to use appropriate sensors to capture information about how groups of people interact over periods of weeks, months and even years. This work demonstrates the feasibility of using wearable sensor packages that are uniquely designed to collect auditory, motion and identification data from people wearing them and methods to detect and identify face-to-face communication patterns in organizations and other social groups.

As many of our interactions over the course of weeks and months are repetitive, if we are able to collect sensor data from people over an extended period of time it will be an ideal dataset for using statistical pattern recognition algorithms to reliably learn reoccurring patterns in communication among group members. The main hypothesis of this thesis is that these patterns of communication in groups can be captured by wearable sensors worn by the group and recognized by machine learning methods, thus serving as a viable alternative for human coders.

Our first task is to design a wearable platform that is acceptable and adopted by a large user base and captures reliable data from the user. The collected data from the wearable is used to extract robust interaction features from the data which can identify when people are talking, who they are talking to and the duration and nature of our face-to-face interactions. The next task is to use these features to statistically estimate the group structure and the dynamics of communication within the network. Doing this will allow us to answer some of the following questions: are some people better connected than others? Does the physical layout of groups have noticeable effect on communication? How does the communication pattern vary over the course of the day? How do individual interaction dynamics affect each other during conversations?

In summary, this work outlines a complete system that enables a sensor-based approach to measuring interactions between groups of people, addressing issues from sensor selection to sensor packaging to robust feature extraction and dynamic models of group interaction.

1.1 Motivation and Background

> *"We like to think of ourselves as individuals capable of making up our own minds about what we think is important, and how to live our lives. Particularly in the United States, the cult of the individual has gained a large and devoted following, governing both our intuitions and our institutions. Individuals are to be seen as independent entities, their decisions are to be treated as originating from within, and the outcomes*

they experience are to be considered indicators of their innate qualities and talents.

It's a nice story, implying as it does not only the theoretically attractive notion that individuals can be modeled as rationally optimizing agents, but also morally appealing message that each person is responsible for his or her own actions. However, there is a difference between holding someone accountable for their actions and believing the explanation for those actions is entirely self-contained. Whether we are aware of it or not, we rarely, if ever, make decisions completely independently and in isolation. Often we are conditioned by our circumstances, our particular life histories and our culture. We also can not help but be influenced by the mesmerizing pool of universally available, often media-driven information in which we continually swim. In determining the kind of person that we are and the background picture against which our lives play out, these generic influences determine both the expertise and the preferences we bring into and decision-making scenario. But once we are in the scenario, even our experience and predispositions may be insufficient to sway us entirely one way or the other. This is where the externalities – whether information, coercive, market or coordination externalities – enter to play a crucial role. When push comes to shove, humans are fundamentally social creatures, and to ignore the role of social information in human decision making – to ignore the role of externalities – is to misconstrue the process by which we come to do the things we do."[6]

This work was greatly motivated by research done in understanding human behavior and interactions in four separate disciplines - (i) social network analysis (ii) mathematical models of network topology and dynamics (iii) social psychology and (iv) ubiquitous and pervasive computing. One of the goals of this work is to make connections and build upon these somewhat disconnected research areas. Below we discuss these areas of research in more detail and make connections to the work done in this thesis.

Social network analysis focuses on uncovering the patterns of people's interactions from data. Many believe that the success or failure of societies and organizations often depends on the patterning of their internal structure. This area of research characterizes the relationship among groups of people[7-10]. Based on network data gathered primarily from surveys, self-reports, or online behavior, the goal is not only to learn the topology but also to understand the roles and position of the members of the network. The emphasis is on analysis techniques that can explain the behavior of the collective – be it in an organization, school or communities[11-14]. It is important to note that many of these techniques make very insightful discoveries by asking simple but relevant questions or queries about the nature of communication of groups. For example, Tyler and Huberman [14] could identify could identify both formal and informal networks within an organization just using 'to' and 'from' fields of email logs. The were also able to identify leadership roles within the communities. Adamic and Adar used the Stanford University online alumni community to measure different social network phenomena [15]. However one of the pressing problems in analyzing the face-to-face social network is data collection which is mostly done by tedious manual input from members of the network being studied or by human observers. We address this problem of data collection and ways to move the burden of observing human communication from human-observer to machines. Once we are able to do this, previous work in social network analysis provides rich set of tools on how to interpret and understand the data.

Recently, significant amount research efforts has been in the 'science of networks' - deriving models that can explain how information propagates, contagious disease spreads, and even terrorist organizations operate. A popular approach is to propose a model and then validate it based on how well the model explains various phenomenon observed by the system [16-19]. Some work focuses on the generalizability of the models to different domains starting from statistical physics to social networks to the worldwide web. Another important direction of this research is how one can use local/partial information about the network structure to propagate information or search for targets efficiently [20-22].

Statistical approaches to understanding the topology and dynamics of how groups of people exchange ideas or propagate information has been done for online communities and the worldwide web [23, 24], where there is an abundance of data and measurements of system parameters are uncertain. But using similar techniques to study the informal networks that arise from face-to-face interactions have not been explored much. A few reasons why there has been so much emphasis on working with online/virtual communities are – (i) online data is more readily available for analysis (ii) there is much less ambiguity in online data compared to sensor measurements of face-to-face interactions. Whereas it may be trivial to figure out whether A sent an email to B but quite difficult to figure out whether A spoke to B using sensors such as microphones, cameras etc. So there are two stages of studying face-to-face communication – (i) extracting reliable measurements from sensors and then (ii) using these features to extract meaningful parameters that explain the pattern of interactions.

There is also a long history of work in the social sciences aimed at understanding the interactions between individuals and influencing their behavior. In the psychology community, there are many instances of work studying these effects. At a simple level, in the work of Wells and Petty [25], the authors show how a person's opinion could be significantly influenced by repeated head nodding from the audience. At a more complex level the structural relationships between people and their physical proximity can effect the collaboration pattern and adoption of new ideas[4, 6]. Studies of this kind give us interesting insights into the workings of human dynamics. In many cases, the experimenters have been able to take quantitative measures of behavior changes by looking at task performance, questionnaire responses or by manual annotation of audio-visual data. However, interaction trends and dynamics are hard to infer from small amounts of hand annotated data. Current technology enables us to record massive amounts of data which can be a treasure trove of information if we can sieve through this data automatically. At the same time, referring back to psychology literature we can identify interaction features that are meaningful and informative in interpreting behavior and serve as guideline on the types of features that we need to extract from our data in order to do analysis.

In the field of ubiquitous and pervasive computing researchers have been working for years in sensor-based approaches to modeling humans and dealing with sensor uncertainty and noise. There has been work in trying to understand the identity and activity of people in meeting rooms, offices and even in open public-spaces [26-31]. The basic notion of context-aware computing is to give computers the ability to understand one's surroundings. However, so far the focus is on the single-person perspective: to understand what and who is around the user of the system and to better provide information or resources to that person. One important aspect often ignored is how others fit into a person's interaction network which can be an important factor in designing a better context-aware agent or environment.

Some research in using sensors to support face-to-face communication comes close to being able to analyze how people tend to interact within a community. The work with active badges focus mainly on presence rather than actions of people – e.g. the Olivetti badge system developed by AT&T could identify a person's location in a building but not whether they were talking or with whom [32]. A few initiatives in utilizing or learning social-network structure using sensors are beginning to emerge [33], however, so far these systems only look use proximity information to make inferences. One of the earliest such systems is the 'lovegety' which is more of facilitator of face-to-face conversation. The 'lovegety' would connect individuals of the opposite sex within five meter radius of the user based on the settings of the users. 'IntelliBadge' is a radio frequency tracking device that can be use to track movement of people and visualize the pattern of their movement. As the identity and some attributes are associated with the device it can be used to do some social network analysis – i.e. are users with similar interests more likely to congregate in certain areas? Two other systems called 'nTag' (based on the thesis work of Rick Borovoy [34]) and 'Spotme' [33] are tools that facilitate networking among participants of large events and monitor the pattern of movements and virtual communication with the users. None of these systems have the ability to identify and analyze face-to-face communication.

Figure 1-2: (a) Olivetti badge (b) Lovegety (c) InteliiBadge

Figure 1-3: (a) nTag (b) SpotMe

In the interests of studying the interactions between humans and the influences of various experimental variables, we worked on an experimental setup we called the "Facilitator Room". Our goal was to use computer vision, speech processing, and machine learning tools towards quantifying how much one person is influencing another's behavior[35]. The main problem we encountered in the Facilitator room was obtaining natural interactions. If the goal is to map, explain and eventually influence human communication and interactions the ability to get sensor measurements from people in their regular setting without imposing any special constraints or requirements is critical. In order to overcome the constraints imposed by the Facilitator room, we built the 'sociometer' to measure interactions in groups without changing their daily patterns of interaction.

1.2 Roadmap

We are now ready to describe our work on sensing and modeling human interactions. As mentioned above this work covers a broad range of material from different research areas. We will describe and refer to the appropriate literature when describing our experiments, methods and analysis in the following chapters. This chapter provided the user with the motivation behind this work and positioned the work amongst the research efforts in social network analysis, social psychology, artificial intelligence and ubiquitous computing. The remainder of the thesis is organized as follows:

- Chapter 2 introduces the sociometer, the wearable sensor package that is used for data collection and describes the data collection process and how we address issues such as user acceptability and privacy. It includes a summary of our subjects' reaction to the sociometer.
- Chapter 3 describes the features and learning methods that we use in the thesis to process low level sensor data and extract information that are relevant for modeling interactions .
- Chapter 4 presents an analysis of the social network based on proximity and interactions. We analyze some individual and group properties that arise in our dataset. We compare the outputs from the sensor data with hand labeled ground truth and also survey data collected during our experiment.

- In Chapter 5, we explore the dynamics of people's interactions and how we quantitatively measure how much a person's style of interaction affects another during an interaction.
- The concluding Chapter 6 outlines the contributions and future directions.

Chapter 2: The Sociometer

To our knowledge, there are currently no sensor-based data-driven methods for modeling 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 [36, 37].

In the following subsections we will describe how we use wearable sensors to collect face-to-face interaction data. Data collected using wearables can be cheaper and more reliable than the data obtained using human-delivered questionnaires, because a sensor-based approach is free from recall failures and personal interpretation bias of surveys, and does not need to use expensive human coders/observers. The ability to discover face-to-face communication networks automatically allows researchers to gather interaction data from larger groups of people. This can potentially remove one of the current bottlenecks in the analysis of human networks: the number of people that can be surveyed using manual techniques.

2.1 Design Issues and Wearability of the Sociometer

One of the key aspects of wearable systems is their *wear-ability*. The design of wearables needs to be suited to the human body and should not interfere with fluid human motion. Very often wearable devices are built just to meet the technical requirements and little attention is given to the design issues. However, if wearable computing and sensing devices are to become popular among the mainstream population, the interaction between the wearable and the human body needs to be smooth and seamless. In building the sociometer we took the design of the device seriously as we needed to have subjects other than the researchers wear it for an extended period of time.

The sociometer has been designed for the comfort of the wearer, aesthetics, and placement of sensors that are optimal in getting reliable measurements of interactions. A very useful guideline on the design considerations for a wearable is outlined in 'Design for Wearability' [38] which was the result of seven years of study with a variety of commercial, military and industrial wearable and mobile devices. Below we list the key principles described in 'Design for Wearability' that one needs to keep in mind while designing a wearable and comment on how well the sociometer meets these design criteria:

1. **Unobtrusive placement** – Areas that (i) are relatively same size across adults (ii) that have low movement/flexibility and (iii) and have large surface area are good candidates for the placement of a wearable. The figure below depicts the regions that the researcher found best suited for placing wearables. For the sociometer, we chose the collar and shoulder area which did not interfere with the dynamics of the body. The area is close to the mouth and hence placing the microphone in a suitable area to pick up the speech of the wearer.

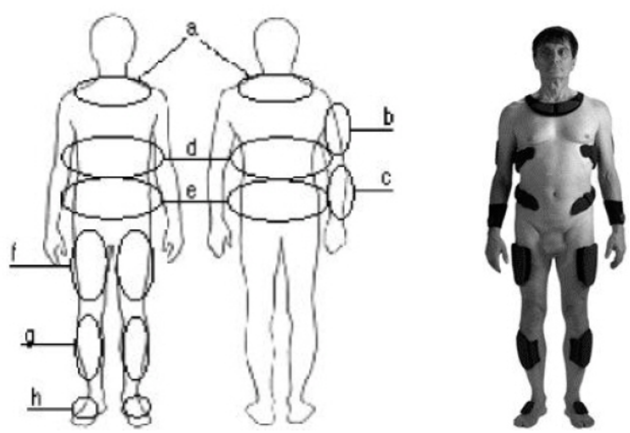

Figure 2-1: Areas in the body that are best suited for placing wearables. Images from [38]

2. **Defining the shape** – Having concavity against the body, convexity outside the surfaces of the form and soft edges ensures comfort and stability . The shape of the sociometer conforms to all these design requirements.

3. **Human movement** – Wearables should be designed around more active areas so that the form does not interfere with changes in body form resulting from movement. The collar placement did not restrict the movement of our subjects in typical activities performed in office/work environment.
4. **Human perception of size** – there is an aura around the human body that the brain perceives as part of the body and typically between 0-5 inches from the body. If a wearable fit within the perceived size the wearer will feel more comfortable and at ease. The sociometer is about 2.5 inches of the body.

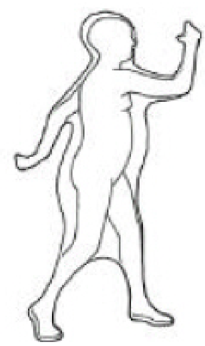

Figure 2-2: Aura around the human body. Image from [38]

5. **Size variation** – The device should be able fit as many types of users and also adjust to the weight gain/loss of the single user. The convexity of the sociometer can be adjusted to fit the varying sizes of the shoulder area.
6. **Attachment and weight** – Wrapping wearables around the body is more comfortable than a single point attachment. The sociometer wraps around the left shoulder and neck area for better weight distribution and user comfort.
7. **Weight** – The weight of a wearable should not hinder the movement or balance of the human body. Our wearable design achieves this through the wrap around attachment.
8. **Aesthetics** – A wearable is a very personal device that a person has on her body. It is important that the design has some aesthetic appeal and is not a conglomeration of the necessary hardware. The sociometer made a stylistic statement which was

received mostly with positive reviews. As with any design, it is almost impossible to get an objective measure of how good our design was.

2.2 Sensors

To reliably measure verbal face-to-face communication we need sensors that can capture interaction features. We need to know when someone is speaking, who they are speaking to, the duration of their interaction etc. To record the identity of people in an interaction, the sociometer is equipped with an infra-red (IR) transceiver that sends out unique ID for the person and receives ID from other people within the wearer's proximity. The device has a microphone that records the audio.

The sociometer hardware is an adaptation of the hoarder board, which is a wearable data acquisition board, designed by the electronic publishing and wearable computing group at the Media lab [39, 40]. The board has an IR transceiver, a microphone, two accelerometers, on-board storage, and its own 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 that it can be worn all day without any discomfort.

The sociometer records the following information for each individual:

1. Information about people nearby (sampling rate 17Hz – sensor IR)
2. Speech information (8KHz - microphone)
3. Motion information (50Hz - accelerometer)

Other sensors (e.g. light sensors, GPS etc.) could be added in the future using the extension 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 microphones.

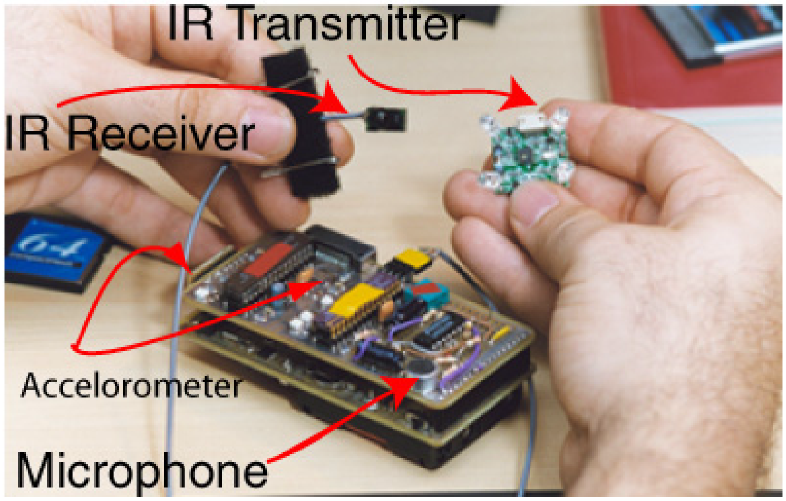

Figure 2-3 The wearable sensor board

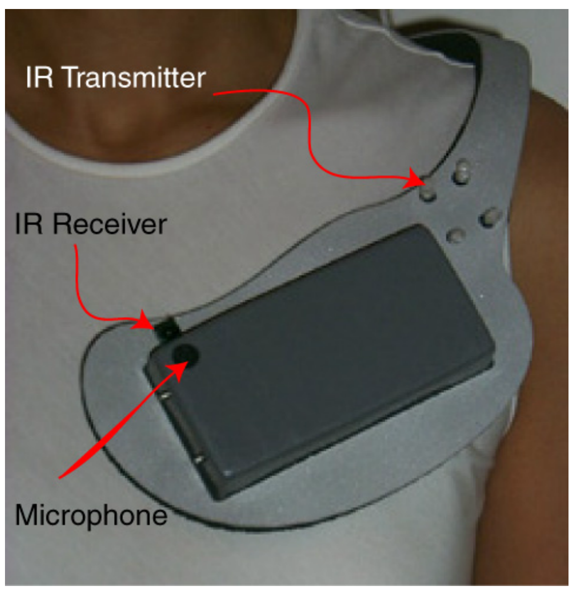

Figure 2-4 The shoulder mounted sociometer

Figure 2-5: Close-up of the sociometer showing the concavity against the body, soft edges and wrap around effect for better wearability

2.3 The data collection method

The formal structure and physical layout of an organization has been shown to impact the collaboration and communication patterns of people [2]. Thus we wanted our experimental design to include variations in these variables. In our experiment we selected participants such that we had (i) participants that belonged to different groups within the overall organization, (ii) had the majority of the members within a group participate in the experiment and (iii) the participants were physically distributed within the building.

The data collection was done at the MIT Media Lab were different research group members volunteered to wear the sociometer. The subjects represented four separate research groups at the lab, and most of the members within a research group participated in the study. The participants were a mix of students, faculty and administrative support staff and were distributed across different floors of the building. The Media lab is a four story high building with research groups housed on the lower level and third and fourth floors. Of the four groups participating in our study, two were from the third floor, one group from the fourth floor and the other from the lower level. **Figure 2-6** show the distribution of the participants in this study overlaid on the floor plan of the building, each color represents a different research group. The users had the device on them for six hours a day (11AM – 5PM) while they were on or around the MIT campus. The subjects did wear the device both indoors and outdoors, e.g., they had the device on when they went outside the building for lunch or to go to the bookstore etc. We conducted the experiment in two stages – (i) the single group stage where 8 subjects from the same research group wore the sociometers for 10 days (two full work weeks - 60 hours of data per subject). (ii) the multi-group stage where 23 subjects from 4 different research group wore the sociometers for 11 days (over two full work weeks and 66 hours of data per subject) and a total of 1518 hours of interaction data. The initial data collection for the single group stage (the author's own research group) was done primarily to debug our system and ensure all the experimental flaws were fixed before we collected the larger data.

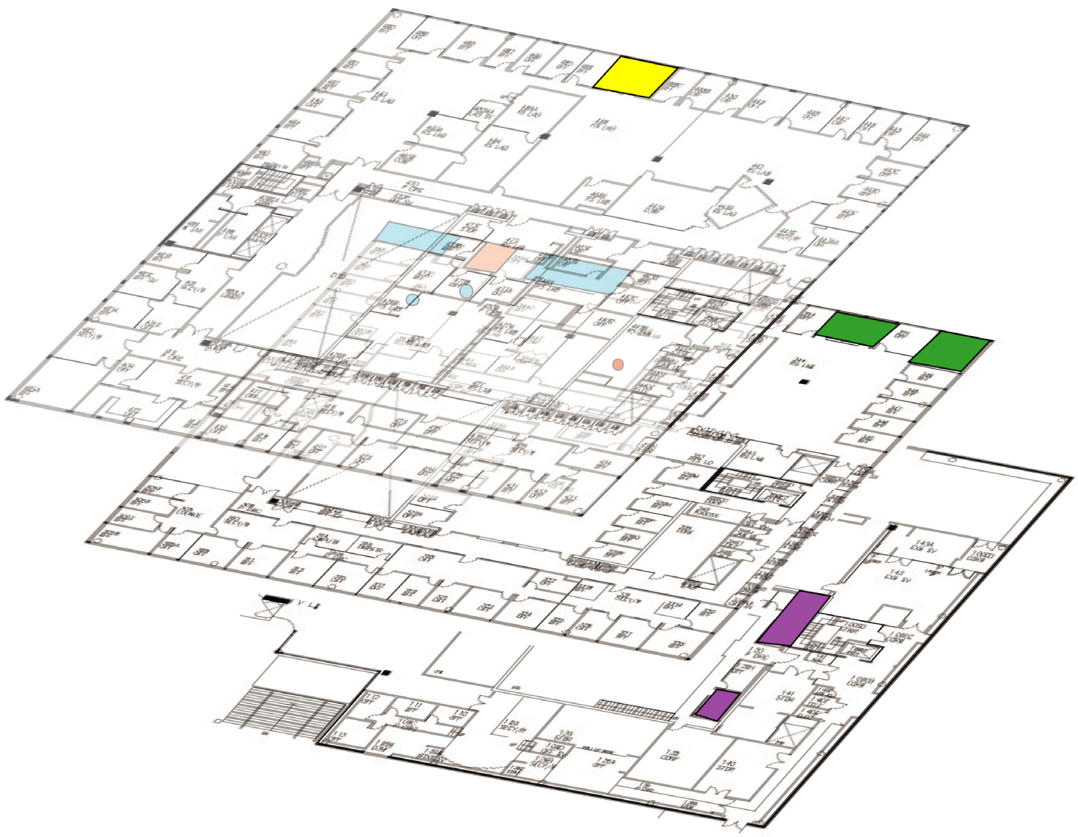

Figure 2-6: Distribution of people participating in the experiment. Different colors represent different research group

Over the course of the experiment, 25 different users wore the device and most of them were satisfied with the comfortable and aesthetic design of the device. Despite the comfort and convenience of wearing a sociometer, we are aware that subject's privacy is a concern in 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, spectral features, 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. For evaluation purposes in this thesis, we had four subject label their own data. Each subject labeled two days worth of data each. Another option maybe to garble the audio such that the content is unintelligible but the identity and pitch of the speaker is preserved [41]. In future versions of the sociometer we will store encrypted audio instead of the audio, which can also prevent unauthorized access to the data.

We take the following measures to address the privacy concerns of our subjects – (i) the identity of the subjects aren't revealed at any point during or after the study (ii) no transcription or speech recognition is done on the data (iii) each individual could withdraw from the study at anytime and the data collected from them would be destroyed (iv) the data collected from the subject will be used for this study only (v) the subjects were given an explanation of the experimental procedures and the consent form will included details about the nature of data that is being collected (vi) the data is stored in a secure computer only accessible by the principal and associate investigators and (vii) the ground-truth data from four different participants were labeled by the subject themselves and not the experimenter. The project is approved by MIT's Committee on the Use of Humans as Experimental Subject (COUHES) application # 2889.

2.4 User Reaction and Acceptance

At the end of the data collection phase we conducted a survey to gauge the acceptance of the sociometer among the users. **Table 2-1** summarizes the results from this survey results. 21 out of the 23 the users completed the survey. The overall reaction of the users is positive – the main complaint about the design is that the Velcro backing on the sociometer can sometimes snag on clothing. Some users commented that sometime they felt the device would fall of as it was not securely attached or fastened to the body.

Some researchers, other than those who participated in the experiment, have also expressed interest in using the device because of the functionality, form factor and unobtrusive design. Four units are currently being used by Professor Stacy Wolski in the Department of Communication at the University of Texas at Austin in her research, three units are being used by the House N project at MIT. However, we do feel in order to have much broader acceptability the sociometer needs to be made smaller, but two key attributes need to be maintained: (i) cannot be cumbersome to wear (ii) must be effortless for the wearer to use. We realize the users had a certain amount of trust in the experimenter and were convinced that the audio data will not be listened to or transcribed at any point. Future incarnations of the sociometer will be much more widely acceptable if features are processed on the fly, and the raw audio is never stored.

How much did the sociometer interfere with your normal interactions?	Never (7)	Few times (14)	Most of the time (0)	Always (0)
Would you wear an audio data collection device only if transcription is not done?	Don't care (6)	Yes (12)	Do not like wearing it, even if transcription is not done (3)	Would never wear any data collection device (0)
What did you think of the design of the sociometer?	Loved it! (4)	Liked it (13)	Did not like it (4)	Hated it! (0)
How comfortable was the device?	Did not notice it was there (5)	Comfortable (7)	Somewhat Uncomfortable (9)	Uncomfortable (0)

Table 2-1: Exit survey of the users

Figure 2-7: Various subjects wearing the sociometers during the two-week data collection period.

2.5 Summary

In this chapter we have presented the design criteria used to build the sociometer. We have described our experiment design and the steps we took to protect the privacy of the users. The sociometer enabled us to collect continuous data of people's daily interactions in their natural settings, opening up the possibility of collecting interaction data without relying on human observers or requiring additional effort from individuals to fill out surveys or questionnaires. Finally we presented how the subjects felt about the design of the sociometer and the idea of wearing a audio data collection device.

Chapter 3: Features and Models

One of the goals of this work is to identify face-to-face vocalic message exchanges, regardless of what is being said or what meanings are assigned to the messages. Although, the content of the messages is no doubt an important part of the interaction, knowing whom we communicate with and how the flavor of the interaction changes based on the identity of the person can provide strong cues about the connections/relationships we have with others in our community.

In order to achieve this goal we need to obtain the following information from our data: (i) when are two people in face-to-face proximity (ii) when is someone speaking (iii) who is the conversational partner and (iv) how often and for how long do a pair interact (v) what are the dynamics of the interaction. In this chapter, we describe the features and models we use to answer the above questions.

3.1 Proximity Features

The data from the IR receiver can be used to detect proximity to 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 ensuring that the proximity relation is symmetric (if receiver #1 detects the presence of tag id #2, receiver #2 should also receive tag id #1). This pre-processing helps to maintain consistency between different information channels.

3.2 Audio Features

In order to identify if two people are interacting we first need to reliably identify regions of speech from all other ambient noise. Because of the close placement of the microphone with respect to the speaker's mouth we can use a simple energy-based threshold to segment the speech of the user from the speech of others in close proximity.

The recent thesis work of Sumit Basu [42] identifies a set of audio features that can be used to detect voiced regions (speech regions that have pitch) within an audio stream very reliably. These features are robust to ambient noise and do not require the use of a close talking microphone.

In this section we will briefly describe these features and why they are capable of detecting voiced speech in regular work environment. For more details about the feature selection process and experiments done to test the robustness of the features in varying amounts noise please refer to [42].

Human speech can be separated into two broad categories – (i) voiced and (ii) unvoiced speech. The voiced region has more energy in the low frequency range and carries well over distance. **Figure 3-1**, shows the spectogram of a speech segment, notice the banded/periodic structure of voiced speech. Because voiced regions are periodic (with pitch period) they produce structured set of peaks in the FFT magnitude domain. Basu selects features that can capture these periodic properties of voiced speech – (i) non-initial maximum of autocorrelation peaks (ii) number of autocorrelation peaks and (iii) spectral entropy.

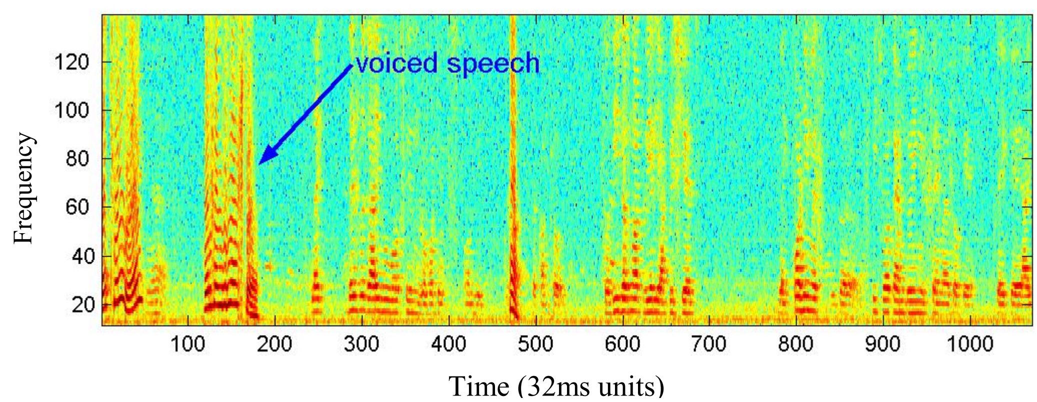

Figure 3-1: Spectrogram of 17s of audio data from the sociometer

Autocorrelation

The autocorrelation A of a signal y[n] is calculated as follows:

$$A[\eta] = \frac{\sum_{n=\eta}^{N} y[n]y[n-\eta]}{(\sum_{n=0}^{N-\eta} y[n]^2)^{1/2}(\sum_{n=\eta}^{N} y[n]^2)^{1/2}}$$

Autocorrelation is calculated per audio frame, where an audio frame in our case is 256 samples of audio (audio is recorded at 8 KHz). The autocorrelation results from a voiced frame (frame #150) and an unvoiced frame (frame #750) are shown in **Figure 3-2**.

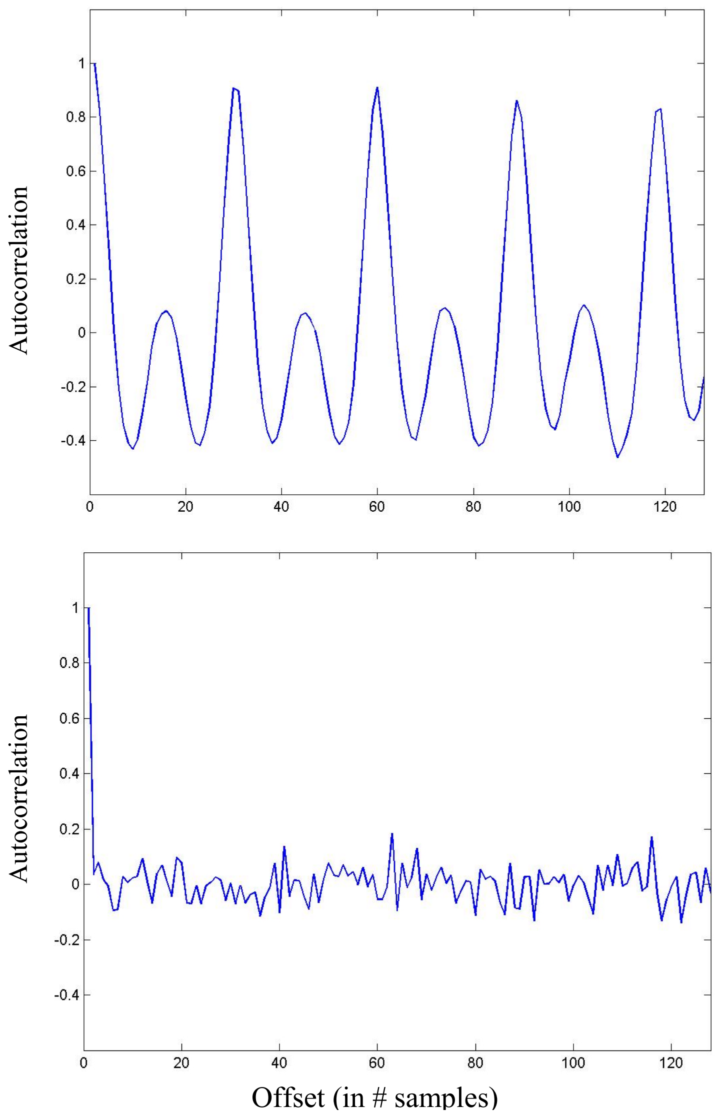

Figure 3-2: Autocorrelation results (a) voiced frame (b) unvoiced frame

Relative Spectral Entropy

Spectra of voiced regions will have strong peaks in FFT magnitudes arising from pitch period, whereas unvoiced regions will have a more noisy spectrum. The relative spectral entropy is calculated as follows:

$$H_{spec} = -\sum_{\omega} Y[\omega] \log \frac{Y[\omega]}{M[\omega]}, \text{ where } Y[\omega] = \frac{y[\omega]}{\sum y[\omega]}$$

Relative spectral entropy (KL divergence between the current spectrum and mean spectrum of the neighboring regions) is more robust in outdoor situations.

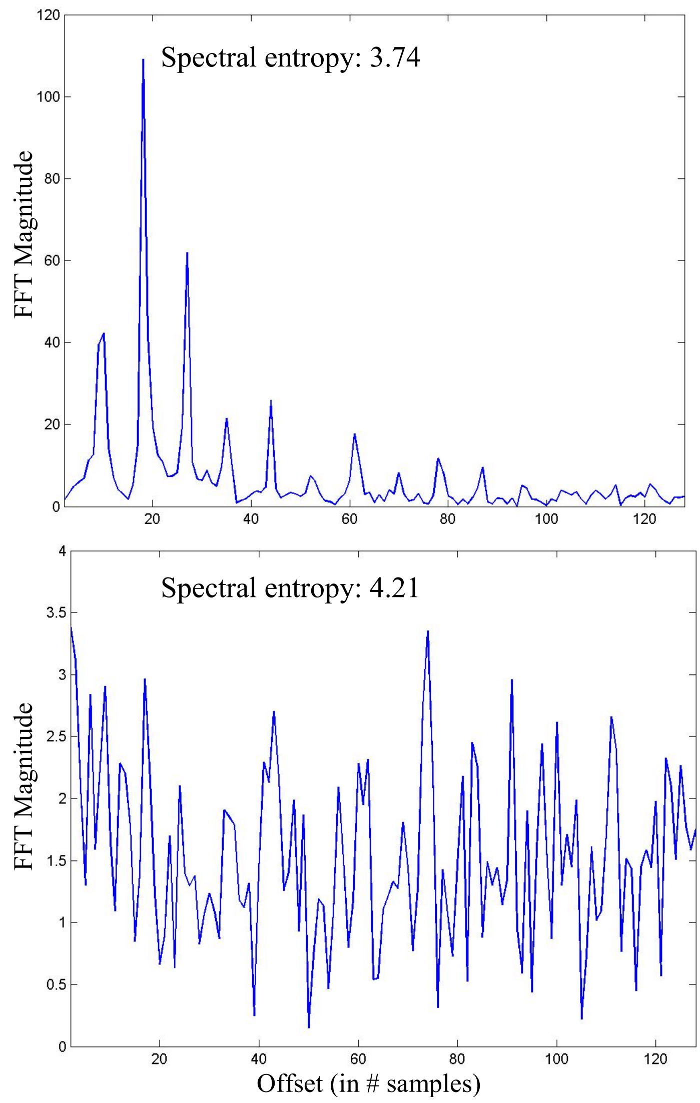

Figure 3-3: FFT magnitude of (a)voiced frame (#150) and (b)unvoiced frame (#750)

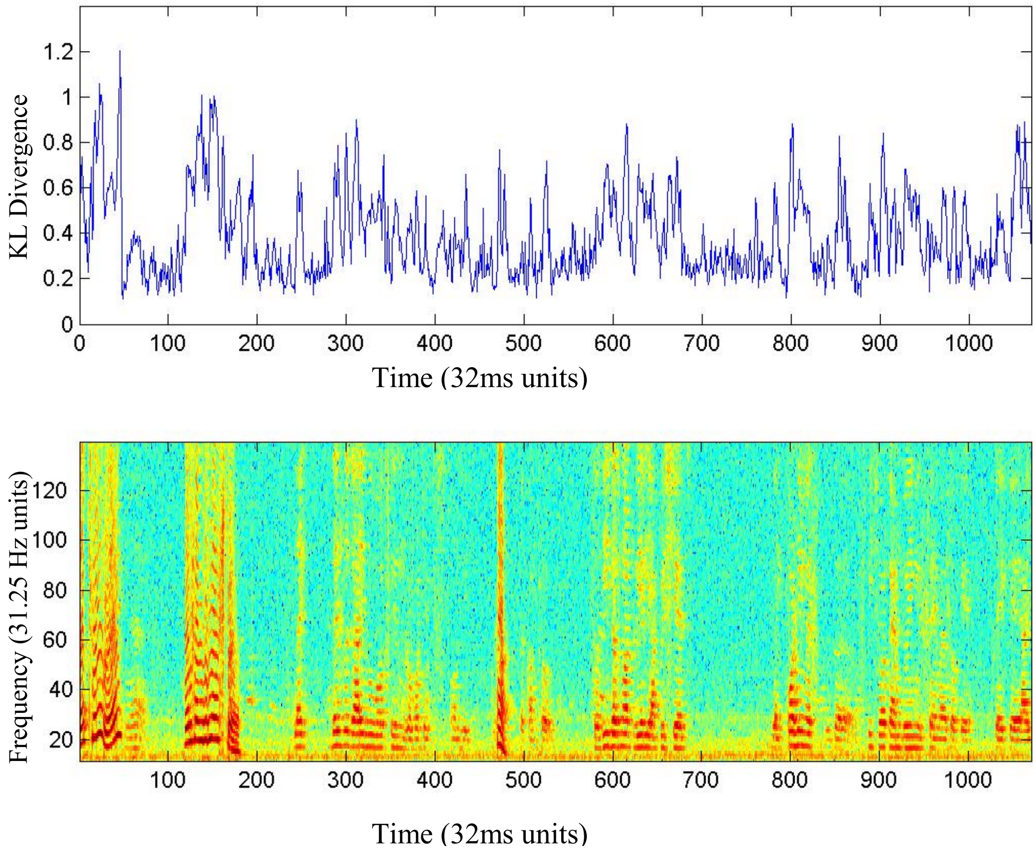

Figure 3-4: (a) KL divergence between the current spectrum and the mean spectrum. Notice the distance is greater in voiced regions. Relative entropy of frame #150 is 0.85 and frame # 750 is 0.24 (b) The spectrogram of the audio segment

Energy

The energy E of a signal y[n] is calculated as follows:

$$E[i] = \sqrt{\sum_k y_i[k]^2}, \text{ for ith frame of the signal.}$$

Because of the placement of the sociometer, the energy level of the wearer will be significantly higher than other voiced speech in the environment. We will use this feature to distinguish the wearer's speech.

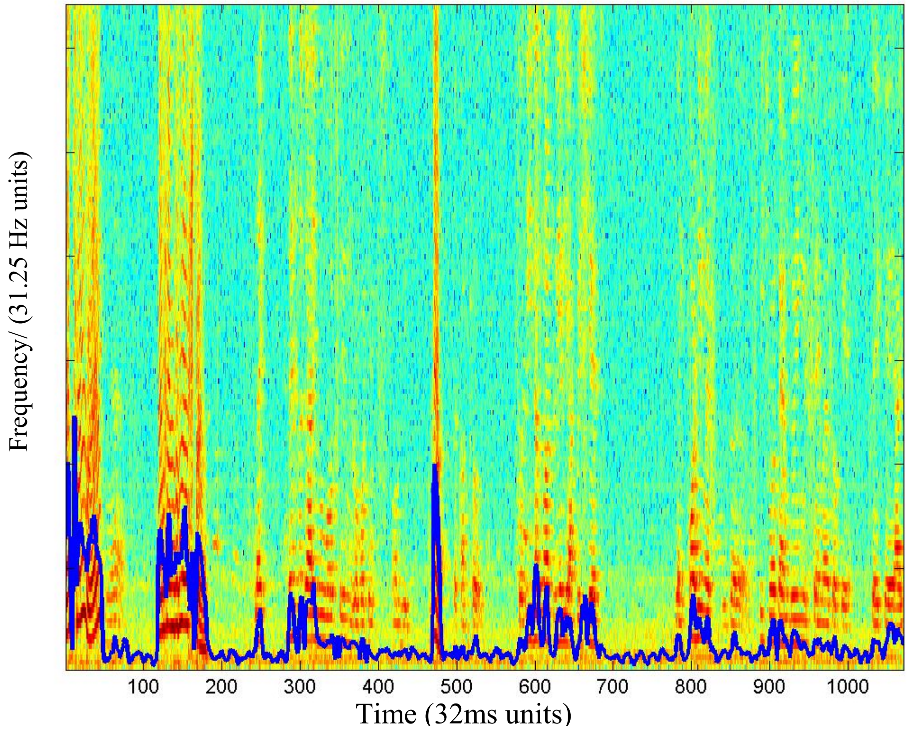

Figure 3-5: Energy from one of the participants in the conversation overlaid on the spectogram. The energy values have been scaled for viewing.

3.3 Computational Modeling Techniques

"Graphical models are a marriage between probability theory and graph theory. They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering -- uncertainty and complexity -- and in particular they are playing an increasingly important role in the design and analysis of machine learning algorithms. Fundamental to the idea of a graphical model is the notion of modularity -- a complex system is built by combining simpler parts. Probability theory provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data. The graph theoretic side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.

…. Moreover, the graphical model formalism provides a natural framework for the design of new systems." --- Michael Jordan, 1998.

In this thesis we mainly use two types of graphical models as our main modeling tools: (i) Hidden Markov Models (HMM) and (ii) the Influence Model. HMMs have been widely and successfully used in modeling various time-series data, e.g. in speech recognition, gesture/activity recognition, etc. The Influence model originally developed by Asavathiratham [43], captures the interaction between two or more interacting time-series data. In this thesis, we cast the Influence model in a graphical model framework. We generalize the model with hidden states/observations and describe an algorithm for learning the parameters of this model from data.

We will describe the Influence Model in significant depth and detail. However, we will only briefly describe the HMM and make reference to relevant literature for readers who seek to get more detail. For those interested in reading more about graphical models - the general framework for representation, learning and inference, we strongly recommend the upcoming book by Jordan and Bishop [44].

3.3.1 Hidden Markov Models

Hidden Markov models are stochastic finite state machines and one of the simplest forms of time-series models. It is a first order Markov model, where the state sequence is not directly observable. The observations are dependent on the state of the model. An HMM is specified by its first order Markov state transition probability, initial state probabilities and observation/emission distribution. For more details about HMMs and how to learn the parameters of an HMM from observations please refer to [45]. **Figure 3-6** shows a graphic description of an HMM – (a)-(c) drawn based on Dan Ellis' class-notes [46].

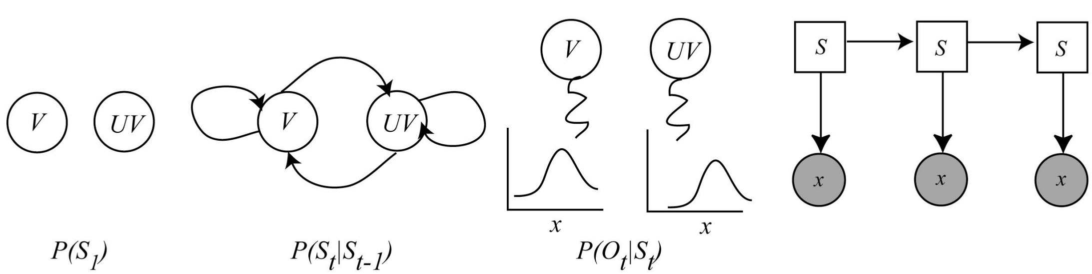

Figure 3-6: HMM is specified by (a) initial state probability (b) state transition matrix and (c) observation probability (d) graphical model representation of an HMM

3.3.2 *The Influence Model*

Now picture two or more Markov processes that have their own dynamics but also interact with each other and effect each other's dynamics. We can think of building a model by stacking a series of HMMs and connecting their state transitions, e.g. a coupled HMM (CHMM). However, the learnability and interpretability of a model greatly depends on its parameterization. The "Influence Model," developed as a generative model by Chalee Asavathiratham in his PhD dissertation[43], is a tractable means of representing the influences a number of Markov chains have on each other. The model describes the connections between many Markov chains with a simple parameterization in terms of the "influence" each chain has on the others. We later found a similar model also proposed by Saul and Jordan[47].

Asavathiratham showed how complex phenomena involving interactions between large numbers of chains could be simulated through this simplified model, such as the up/down time for power stations across the US power grid. The Influence model is a tractable framework for understanding the global system and its steady state behavior. The representation of the model makes the analysis of global behavior possible, which otherwise would become intractable with increasing number of individuals or agents.

In Asavathiratham's description all states were observed. He did not develop a mechanism for learning the parameters of the model – he assumed that they were known apriori. Learning the model parameters from observation in an important requirement in our case. We generalize the model with hidden states/observations and develop an algorithm for learning the parameters of this model from data. We then show the performance of this algorithm and describe the characteristics of the solution space using synthetic data.

The graphical model for the influence model is identical to that of the generalized *N*-chain coupled HMM[48], but there is one very important simplification. Instead of keeping the entire $P(S_t^i \mid S_{t-1}^1, \ldots, S_{t-1}^N)$, we only keep $P(S_t^i \mid S_{t-1}^j)$ and approximate the former with:

$$P(S_t^i \mid S_{t-1}^1, \ldots, S_{t-1}^N) = \sum_j \alpha_{ij} P(S_t^i \mid S_{t-1}^j)$$

In other words, we form our probability for the next state by taking a convex combination of the pair-wise conditional probabilities for our next state given our previous state and the neighbors' previous state. As a result, we only have N QxQ tables and N α parameters per chain, resulting in a total of $NQ^2 + N^2$ transition parameters, where N is the number of chains and Q is the number of states per chain. This is far fewer parameters than the CHMM where the total number of parameters is of NQ^{N+1}. The real question, of course, is whether we have retained enough modeling power to determine the interactions between the participants. Asavathiratham refers to the α's as "influences," because they are constant factors that tell us how much the state transitions of a given chain depend on a given neighbor. It is important to realize the ramifications of these factors being constant: intuitively, it means that *how much* we are influenced by a neighbor is constant, but *how* we are influenced by it depends on its state. This simplification seems reasonable for the domain of human interactions and potentially for many other domains. Furthermore, it gives us a small set of interpretable parameters, the α values, which summarize the interactions between the chains. By estimating these parameters, we can gain an understanding of how much the chains influence each other.

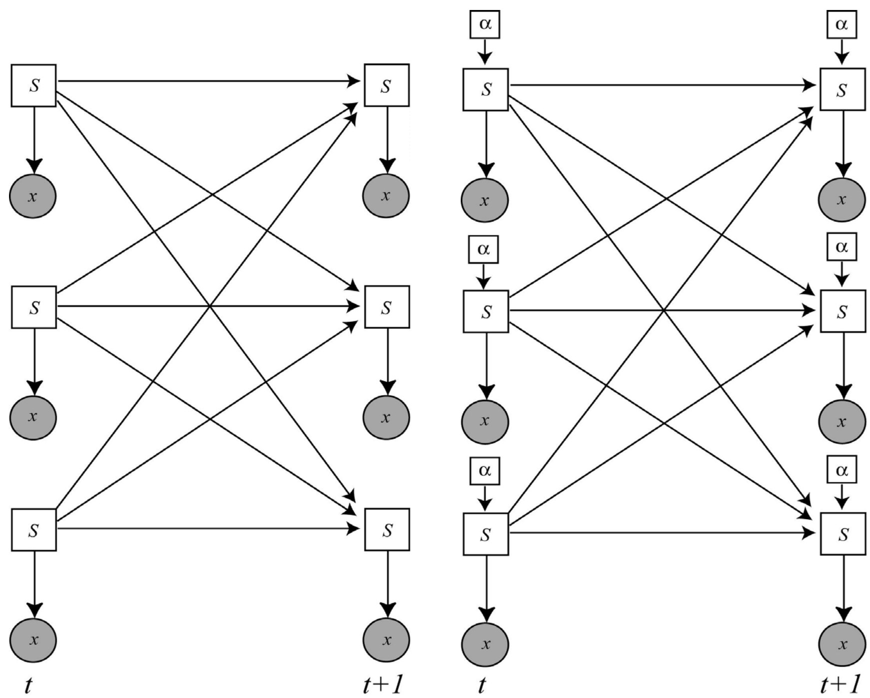

Figure 3-7 Graphs for (a) a generalized coupled HMM and (b) an Influence Model. Shaded nodes are observed and the others are hidden nodes

3.3.2.1 Learning for the Influence Model

The problem of estimating the Influence Model from data can be stated as follows. We are given sequences of observations, $\{x_t^i\}$, from each chain, i. The goal is to estimate the amount of influence, α_{ij}, that chain *j* has on chain *i,* along with the pair-wise conditional probability distributions that describe this inter-chain influence, $P(S_t^i \mid S_{t-1}^j)$. In this section we develop methods for doing this and illustrate them with synthetic data.

Expectation-Maximization Method

In Figure 3-7 we show the graphical model for the most general form of the Influence Model with hidden states and continuous observations. Fitting this model to data requires us to maximize the likelihood of Influence Model over its free parameters. The likelihood function can be readily written as:

$$P(S,X) = \left(\prod_i P(S_0^i) P(x_0^i \mid S_0^i) \right) \prod_i \prod_t P(x_t^i \mid S_t^i) \sum_j \alpha_{ij} P(S_t^i \mid S_{t-1}^j)$$

One possibility for estimating the parameters of this model is Expectation-Maximization. The E-step requires us to calculate $P(S \mid X)$ which in most cases amounts to applying the Junction Tree algorithm (exact inference) or other approximate inference algorithms. We will discuss the possibilities for doing inference on this model later. The M-step is specific to this model and requires maximizing the lower bound obtained in the E-step. Examining this expression we can see that the M-step for all the parameters except the α_{ij}'s is only trivially different from the HMM[49]. However, we can readily write down the update equations for the α_{ij}'s by noticing that they are mixture weights for *N* conditional probability tables analogous to a mixture of Gaussians. The α_{ij} update equations are obtained by following the derivation of the M-step for a Gaussian mixture (i.e. introduce a hidden state to represent the "active" mixture component and then take an expectation over its sufficient statistics):

$$\alpha_{ij}^{new} = \frac{\sum_t \sum_k \sum_l P(c_t^i = j, S_t^i = k, S_{t-1}^j = l \mid X)}{\sum_t \sum_k \sum_l P(S_t^i = k, S_{t-1}^j = l \mid X)}$$

The "$c_t^i = j$" event means that at time t chain i was influenced by chain j, and the "$S_t^i = k$" event means that chain i was in state k during time t.

Unfortunately, exact inference of the Influence model is computationally intractable because of the densely connected hidden variables [50]. Variational methods or approximate inference techniques may be alternate tractable methods for learning the full model.

The Constrained Gradient Descent Method

Imagine a slightly different situation where we already have the state transition probabilities of the influence model (e.g. we have voicing/unvoicing states for N different individual participating in a conversation). Now our states are observed but the influence values are still hidden, we called this model the observed Influence Model (**Figure 3-8**). During learning we need to estimate the inter-chain influences α_{ij} s.

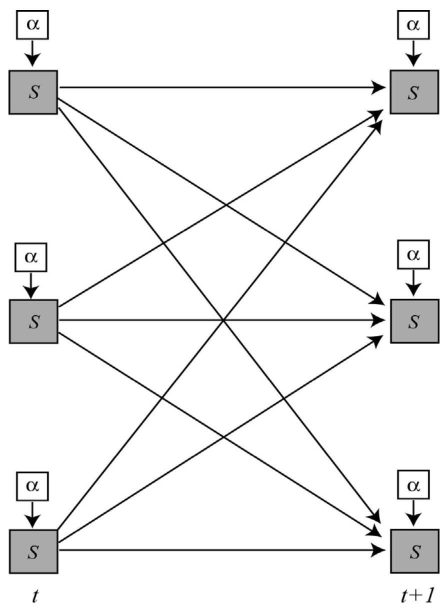

Figure 3-8: The observed influence model. Shaded nodes are observed and the others are hidden nodes

The observed Influence model is an unusual DBN where the observed nodes are strongly interconnected and the hidden states are not. This presents problems for inference because marginalizing out the observed state nodes causes all the hidden states to become fully connected across all time and all chains. Unless we apply an approximation that can successfully decouple these nodes, a maximization procedure such as EM will not be tractable. However, there is a simpler way to estimate the α_{ij} values in our observed scenario. Let us first examine how the likelihood function simplifies for the observed Influence Model:

$$P(S \mid \{\alpha_{ij}\}) = \left(\prod_i P(S_0^i)\right) \prod_i \prod_t \sum_j \alpha_{ij} P(S_t^i \mid S_{t-1}^j)$$

Converting this expression to log likelihood and removing terms that are not relevant to maximization over α_{ij} yields:

$$\alpha_{ij}^* = \arg\max_{\alpha_{ij}} \left[\sum_i \sum_t \log \sum_j \alpha_{ij} P(S_t^i \mid S_{t-1}^j) \right]$$

We can further simplify this expression by keeping terms relevant to chain i:

$$\alpha_{ij}^* = \arg\max_{\alpha_{ij}} \left[\sum_t \log \sum_j \alpha_{ij} P(S_t^i \mid S_{t-1}^j) \right]$$

This *per chain* likelihood is concave in α_{ij}, which can be easily shown as follows: Let

$$\alpha = \begin{bmatrix} \alpha_{i0} \\ \vdots \\ \alpha_{iN} \end{bmatrix} \quad B_t^i = \begin{bmatrix} P(S_t^i \mid S_{t-1}^0) \\ \vdots \\ P(S_t^i \mid S_{t-1}^N) \end{bmatrix}$$

then the *per chain* likelihood becomes: $f_i(\alpha) = \sum_t \log \left\langle \alpha, B_t^i \right\rangle$.

This is concave since for any $0 < w \leq 1$ and α_0, α_1:

$$
\begin{aligned}
f((1-w)\alpha_0 + w\alpha_1) &= \sum_t \log\left\langle (1-w)\alpha_0 + w\alpha_1, B_t^i \right\rangle \\
&= \sum_t \log\left[(1-w)\left\langle \alpha_0, B_t^i \right\rangle + w\left\langle \alpha_1, B_t^i \right\rangle \right] \quad \text{(using Jensen's inequality)} \\
&\geq \sum_t (1-w)\log\left\langle \alpha_0, B_t^i \right\rangle + w\log\left\langle \alpha_1, B_t^i \right\rangle \\
&= (1-w) f(\alpha_0) + w f(\alpha_1)
\end{aligned}
$$

Now take the derivative w.r.t. α_{ij}:

$$
\frac{\partial}{\partial \alpha_{ij}}(.) = \sum_t \frac{P(S_t^i \mid S_{t-1}^j)}{\sum_k \alpha_{ik} P(S_t^i \mid S_{t-1}^k)} = \sum_t \frac{P(S_t^i \mid S_{t-1}^j)}{\left\langle \alpha_i, B_t^i \right\rangle}
$$

Re-parameterize for the normality constraint:

$$
\begin{aligned}
\beta_{ij} &= \alpha_{ij} \\
\beta_{iN} &= 1 - \sum_j \beta_{ij}
\end{aligned}
$$

$$
\beta_{ij}^* = \arg\max_{\beta_{ij}} \left[\sum_t \log \sum_j \beta_{ij} P(S_t^i \mid S_{t-1}^j) + (1 - \sum_j \beta_{ij}) P(S_t^i \mid S_{t-1}^N) \right]
$$

Now take the derivative w.r.t. β_{ij}:

$$
\frac{\partial}{\partial \beta_{ij}}(.) = \sum_t \frac{P(S_t^i \mid S_{t-1}^j) - P(S_t^i \mid S_{t-1}^N)}{\sum_k \beta_{ik} P(S_t^i \mid S_{t-1}^k) + (1 - \sum_k \beta_{ik}) P(S_t^i \mid S_{t-1}^N)}
$$

We need to maintain the following constraints on β_{ij} to fulfill the normality constraints on α_{ij}:

$$
\begin{aligned}
\beta_{ij} &> 0 \\
1 - \sum_j \beta_{ij} &> 0
\end{aligned}
$$

These constraints can be enforced with the following log boundary functions:

$$
\beta_{ij}^* = \arg\max_{\beta_{ij}} \left[\sum_t \log \sum_j \beta_{ij} P(S_t^i \mid S_{t-1}^j) + (1 - \sum_j \beta_{ij}) P(S_t^i \mid S_{t-1}^N) + A \log \sum_j 1 - \beta_{ij} + B \sum_j \log \beta_{ij} \right]
$$

We update the derivative to reflect these log boundary penalties:

$$\frac{\partial}{\partial \beta_{ij}}(.) = \sum_t \frac{P(S_t^i \mid S_{t-1}^j) - P(S_t^i \mid S_{t-1}^N)}{\sum_k \beta_{ik} P(S_t^i \mid S_{t-1}^k) + (1 - \sum_k \beta_{ik}) P(S_t^i \mid S_{t-1}^N)} - \frac{A\beta_{ij}}{1 - \sum_k \beta_{ik}} + \frac{B}{\beta_{ij}}$$

The gradient and the per-chain likelihood expression above are inexpensive to compute with appropriate rearranging of the conditional probability tables to form the B_t^i vectors. This along with the facts that the per-chain likelihood is concave and the space of feasible α_{ij}'s is convex means that this optimization problem is a case of constrained gradient ascent with full 1-D search (see [51]). Furthermore, in all examples, 20 iterations were sufficient to ensure convergence.

3.3.2.2 Performance of the Learning Algorithms

To evaluate the effectiveness of our learning algorithm we show results on synthetic data. The data is generated by an Influence Model with 3 chains in lock step: one leader which was evolving randomly (i.e., flat transition tables) and 2 followers who meticulously followed the leader (i.e., an influence of 1 by chain 2 and a self-influence of 0). We sampled this model to obtain a training sequence of 50 timesteps for each chain. These state sequences were then used to train another randomly initialized Influence Model. For this learned model, the $P(S_t^i \mid S_{t-1}^j)$ were estimated by counting and the α_{ij}'s by maximizing the likelihood with gradient ascent as described above. The resulting influence graph is shown along with a typical sample sequence in **Figure 3-9**. Note how the "following" behavior is learned exactly by this model – chains 1 and 3 follow chain 2 perfectly.

The alpha matrix captures the strength of chain 2's dynamics on chains 1 and 3 very well. The learned alpha matrix is:

$$\begin{bmatrix} 0.0020 & 0.9964 & 0.0017 \\ 0.2329 & 0.4529 & 0.3143 \\ 0.0020 & 0.9969 & 0.0011 \end{bmatrix}$$

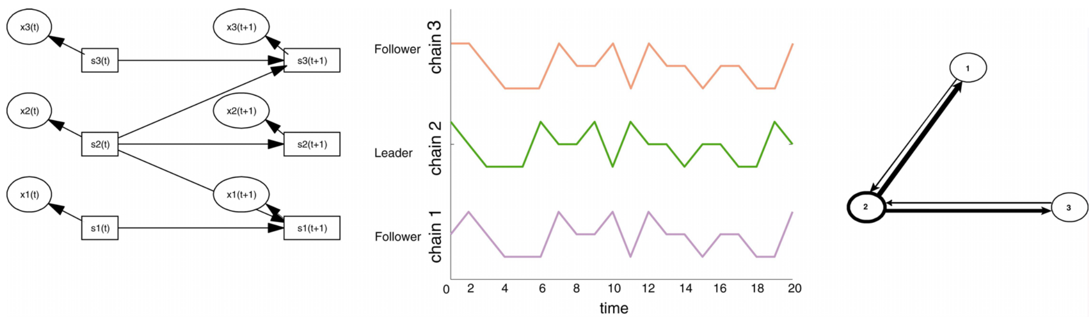

Figure 3-9 The evaluation pipeline for testing the Influence Model on the lockstep synthetic data: (a) the graph for the generating model at time t and t+1 (b) the training sequence (c) the learned influences (α's) – the thickness of the lines corresponds to the magnitude of the influence. Note that the strong influence of chain 2 on 1 and 3 was correctly learned.

We also evaluated the Generalized Coupled HMM (i.e. full state transition tables instead of the mixtures of pair-wise tables) on this data using EM, using the Junction Tree Algorithm for inference[44]. Again we sampled from the lock step model and trained a randomly initialized model. In this case, the learned model performed reasonably well, but was unable to learn the "following" behavior perfectly due to the larger number of parameters it had to estimate ($P(S_t^i \mid S_{t-1}^1,\ldots,S_{t-1}^N)$ vs. $P(S_t^i \mid S_{t-1}^j)$).

In order to measure the effect of training data size on the quality of the model, we test how well the learned model predicts the next state of the follower chains given the current state of all the chains for different size training set. **Table 3-1** shows the prediction results for both the Influence Model and the Generalized Coupled HMM (GCHMM). Clearly, the Influence Model requires a lot less training data to model the dynamics accurately.

Training Data Size	Influence Model Chain #		GCHMM Chain #	
	1	3	1	3
10	100%	99.5%	67%	50.5%
20	99.5%	100%	66%	90.5%
50	100%	100%	100%	100%
100	100%	100%	100%	100%

Table 3-1: Prediction Results for the follower chain in the synthetic dataset

3.4 Summary

In this chapter, we have described the main features and computational models that we will be using on our sensor data collected. Now we are ready to move forward and learn the social network structure and interaction dynamics among the network of people who were equipped with the sociometers.

Chapter 4: Learning the Social Network

With a large dataset of natural interactions from a social group, a set of features and computational tools in hand we are ready to analyze various aspects of group interactions and map the social network. We start at the simplest level by looking at the proximity map, determined from the infrared (IR) sensors. These sensors can detect when people are facing each other within a range of approximately six feet, which is the spatial arrangement in which people would typically be if they were having a conversation. Even this very simple IR measure can provide a reliable estimate of the average statistics of people's social networks. However, proximity alone does not tell us enough about whether two individuals interact. In order to get a better picture of the communication flow we need to be able to detect when people are talking, who they are talking to and how often they talk to each other. We present in this chapter our analysis of vocalic interactions, and network maps based on such interactions. During the data collection phase we also collected a daily self-report from each subject about the interactions they had during the time they were wearing the sociometer. We compare our automatic methods against the survey data and also hand-labeled data.

4.1 Proximity Map

The IR data can tell us who is within face-to-face proximity of whom. However, as we mentioned earlier, 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. We need to detect the contiguous time chunks (an episode) during which people are in proximity, starting from a bursty IR signal. A hidden Markov model (HMM) 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 the binary state 1 when the IDs received come from the same episode, and 0 when they are from different episodes.

In order to train the model, 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 of two people being within face-to-face proximity. **Figure 4-1** shows five days of one person's proximity information. Each gray bar in the sub-image identifies a person with whom the wearer is in close proximity, and the length of the bar shows the duration of contact. Note that we are also able to detect when multiple people are in close proximity at the same time.

The IR data provides no information about whether two people are actually having a conversation. They may just have been sitting face-to-face during a meeting. Also, there are many instances where we have a conversation with someone with being face-to face, e.g. walking along the corridor and having a chat, sitting at a 90˚ angle to someone, or having a conversation with someone whose at your office door while you are sitting at your desk oriented towards the monitor. In general the IR data did not prove to be a reliable indicator of the fine-structure of actual interactions. In data labeled by four subjects IR tags identified 9 out of the 23 pairs of interactions that were labeled. The performance further dropped when we calculated accuracy numbers for each five minute chunks of labeled data. However, if we just want the an average picture of the interaction over the course of the experiment the tag data performs well, but we loose the finer grain breakdown of the network interaction patterns that we initially set out to obtain. **Table 4-1** summarizes the performance of the IR tags in identifying interactions at different time granularities. When analyzing the survey data, we found that only 54% of the time both individuals in a conversation pair acknowledged having a conversation and only 29% acknowledged having the same number of conversations with each other. Such inconsistencies are not unique to this experiment but are known drawbacks of personal recall based method [5]. This also highlights the need for automatic techniques that aren't susceptible to the fallibility of human memory.

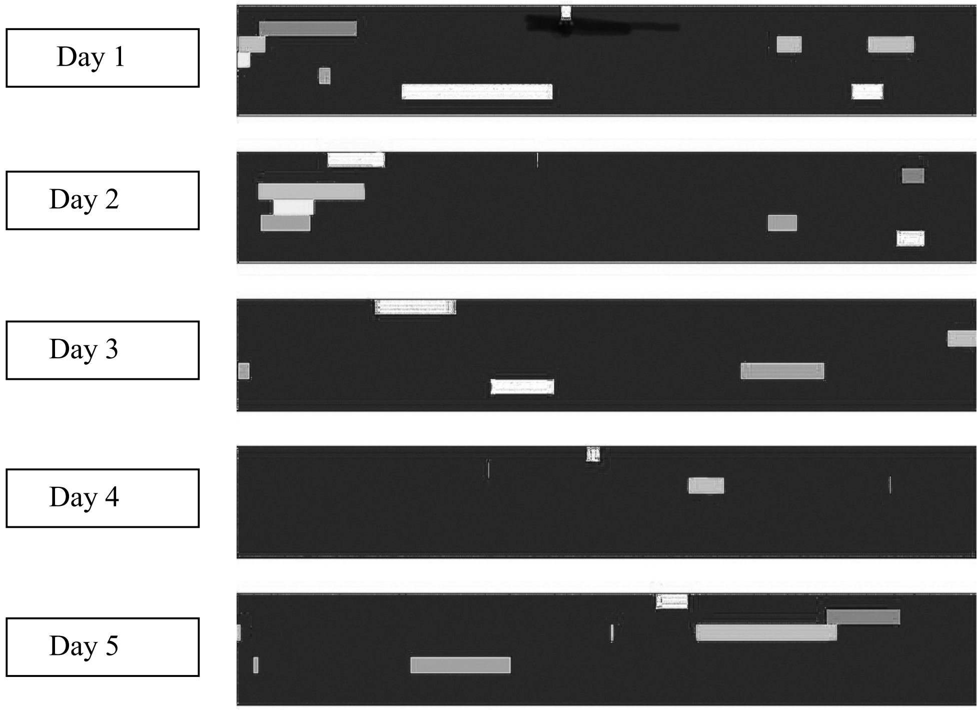

Figure 4-1 - Proximity information for one individual. Each sub-image shows one day's information. Each row within the sub-image corresponds to a different person. Bursty raw sensor data transformed into the contiguous time chunks using HMM.

	Accurate detection
Per 5 minute chunk using hand labeled data	19%
Per conversational pair using hand labeled data	39%
Over all days – per conversation pair using survey data	86% (10% false rejection)

Table 4-1: Performance of the IR tags

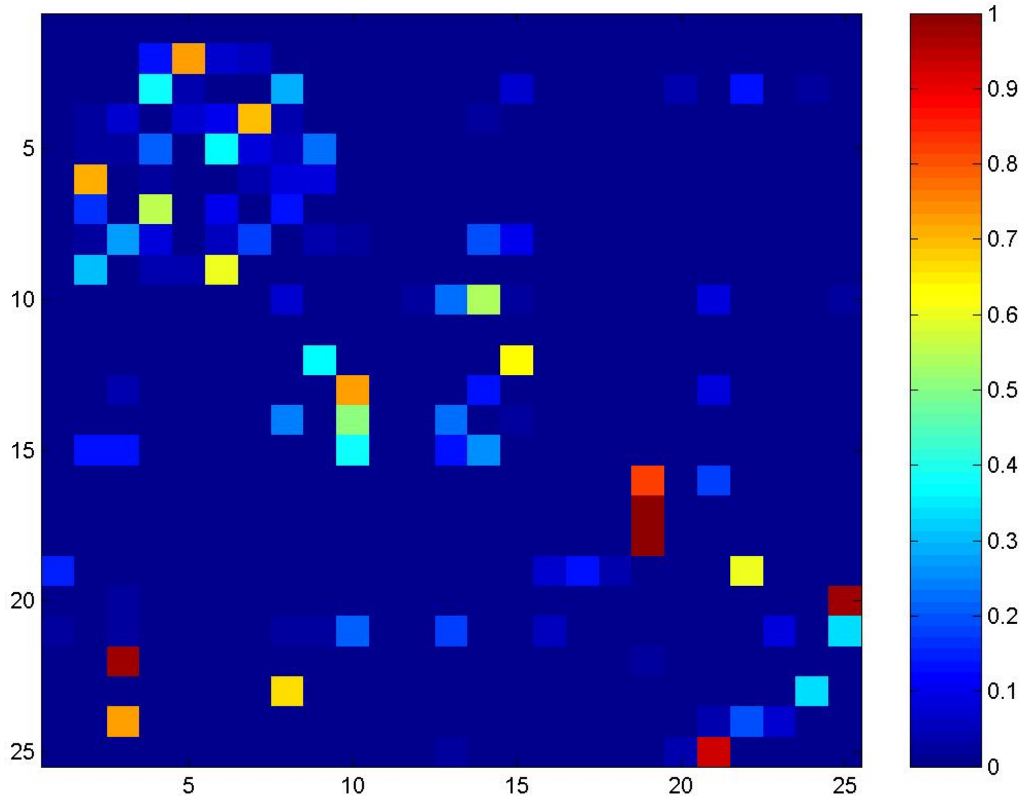

Figure 4-2: Network map based on proximity information. x-axis and y-axis values represent subject ID number, and the color value indicates, for each subject, the proportion of their total interactions that they have with each of the other subjects.

4.2 Identifying Interactions

Face-to-face proximity data can give us an average picture of the network. It does not say much about whether there was real interaction and nothing about the dynamics of the interaction between a conversational pair. In order to do that, we need to start looking at speech information. In this section we describe how we segment speech from other ambient sounds, identify conversations, and segment speakers within conversation chunks.

4.2.1 Segmenting Speech Regions

In Chapter 3, we described some discriminatory features for detecting voiced speech that have been shown to be robust to ambient noise[52]. We use these features to train a two state HMM that can segment voiced speech for other sounds in the environment. The observation variables are the autocorrelation and spectral entropy features, which were described in the previous chapter. We model the emission probability $P(O_t|S_t)$ of the HMM with a single Gaussian. The HMM is trained on a two-minute segment of hand labeled data. Because the hidden nodes of the HMM were labeled (voiced/unvoiced), training was trivial in our case. **Figure 4-3** and **Figure 4-5** show the features and segmentation output of the HMM for a sample audio chunk.

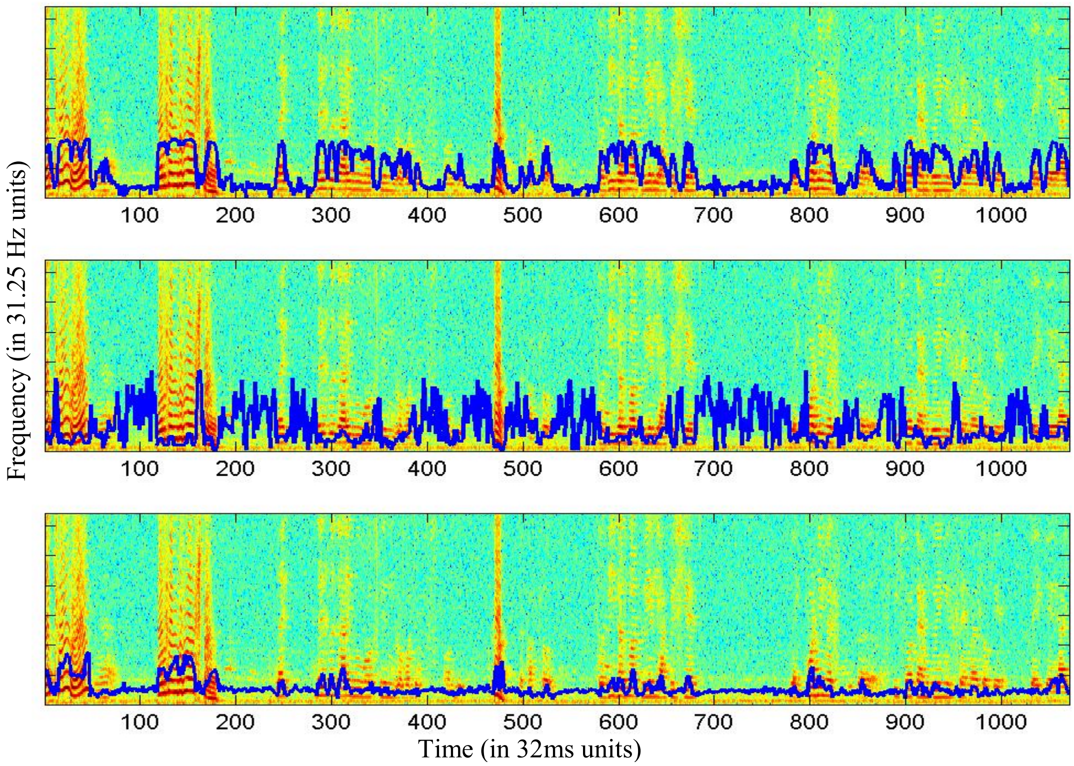

Figure 4-3: Speech features overlaid on a spectogram: (a) auto-correlation maximum (b) number of autocorrelation peaks (c) relative spectral entropy. Note that measures (a) and (c) indicate that voiced speech is present when they are high, whereas measure (b) indicates speech when its value is low.

In order to test the performance of the HMM we took the data that was labeled by four different subjects, who each labeled two days worth of data. Each five minute chunk was labeled as containing speech or just environmental background noise. We tested whether our algorithm found voiced regions in all the segments containing speech and its performance on segments containing no speech. The false alarm rate was less than 0.1% on over nine hours of data labeled as containing no speech. It detected voicing in all the five minute chunks labeled as containing speech, however we do not have a finer grain performance number on correct identification rate. However, Sumit Basu, in [42], showed less than 2% error in detection on telephone speech at 10dB noise. He was using a two-layer HMM that was doing segmentation of voicing and speech chunks, so the accuracy number may not be exactly comparable.

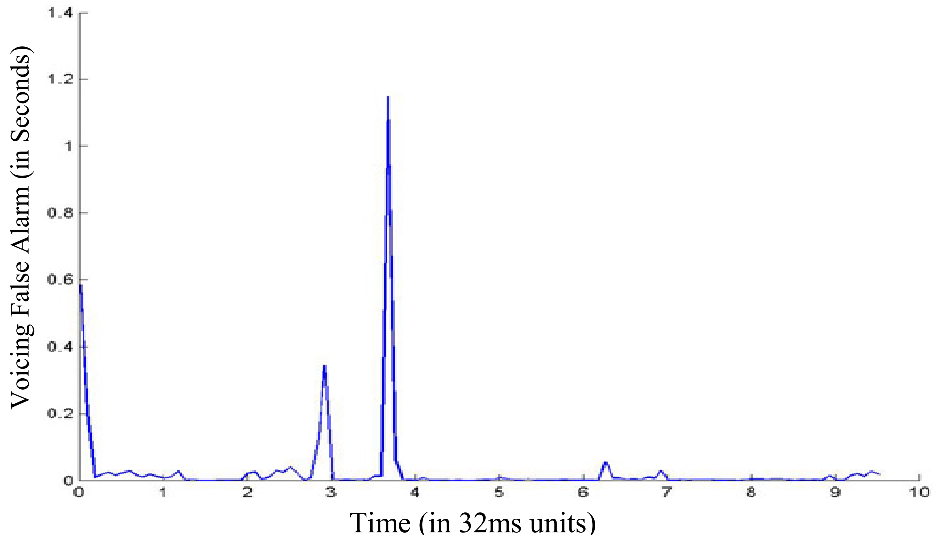

Figure 4-4: Voicing false alarm per five minute chunks of audio

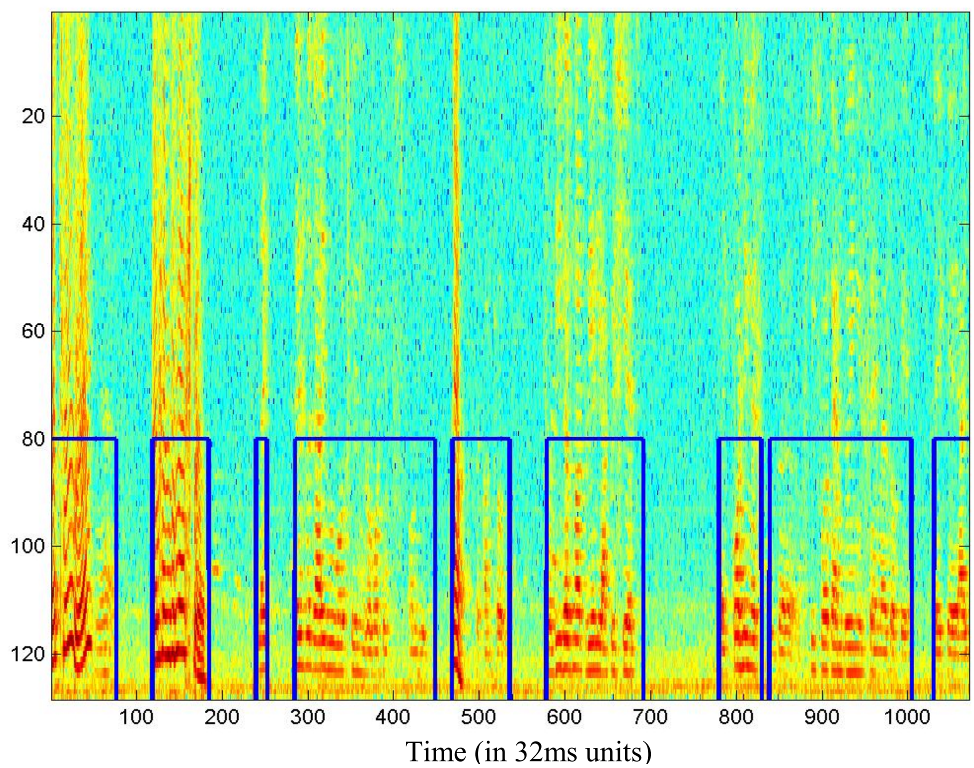

Figure 4-5: Segmentation of voiced region

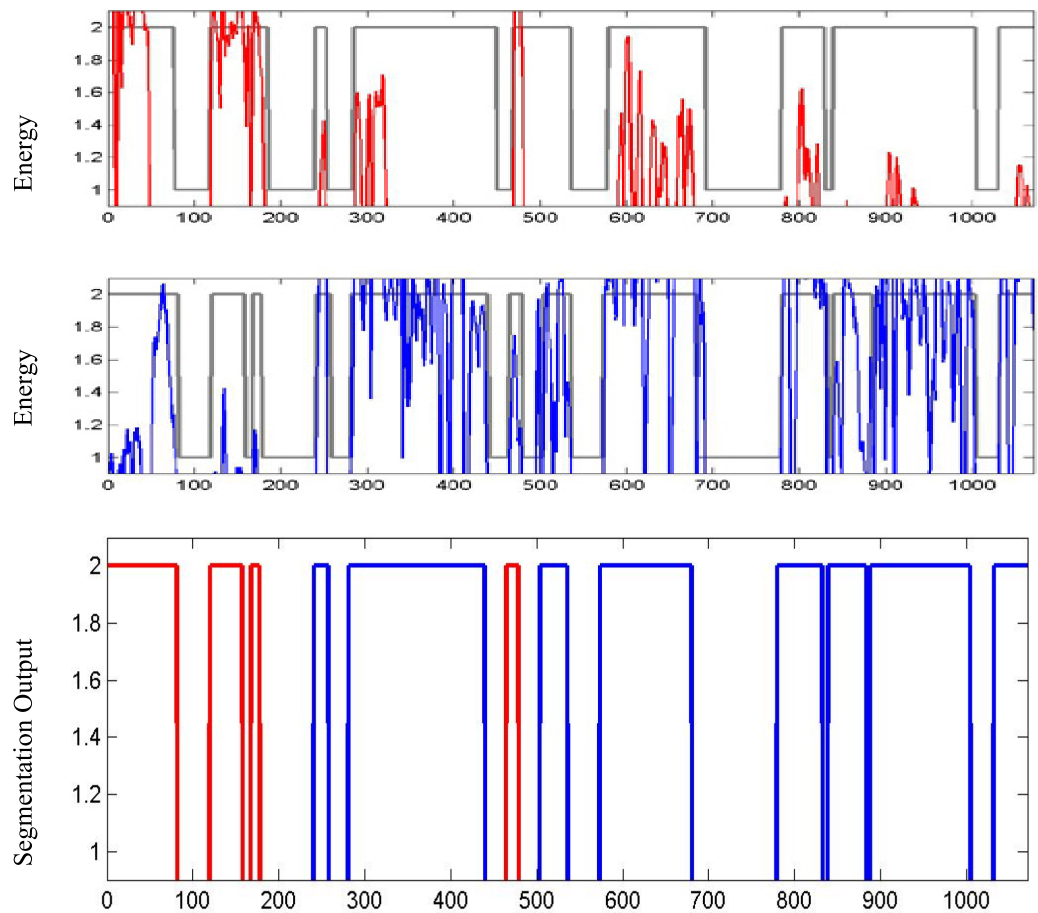

Figure 4-6: Energy-based segmentation of wearer's speech (a) energy signal of speaker A, (b) energy signal of speaker B, (c) segmentation of the speakers

4.2.2 Segmenting Speakers

The voicing segmentation gives us regions that contain human speech. However, many of these regions do not contain any speech from the wearer, but instead contain speech from other people who were talking in the vicinity. Because the sociometer microphone is placed quite close to the mouth of the wearer (about 6"-8" from the mouth), we can distinguish the wearer's speech from all other speech by just using simple energy features. We do the energy based segmentation in two ways: (i) we train a simple Gaussian classifier on the energy features within the voiced region or (ii) if we already knew that two people are having a conversation (we will see how we do that in the next section), we can use the energy ratio between the two sociometer audio signal to segment the wearer's speech.

4.2.3 Finding Interactions

So far we have shown how to reliably identify speech regions and speakers, but not how to find who is talking to whom. We already have the features necessary to find pair-wise interactions. If two people are in a conversation with each other, we expect to see some correlation between the audio features of the two participants. If we have the voicing segmentation of each of the audio data stream, the mutual information between the pair tells us how predictable a person's voicing state is, given the voicing state labels of another. Using mutual information measure to find conversations from multiple streams of audio was first proposed in [42].

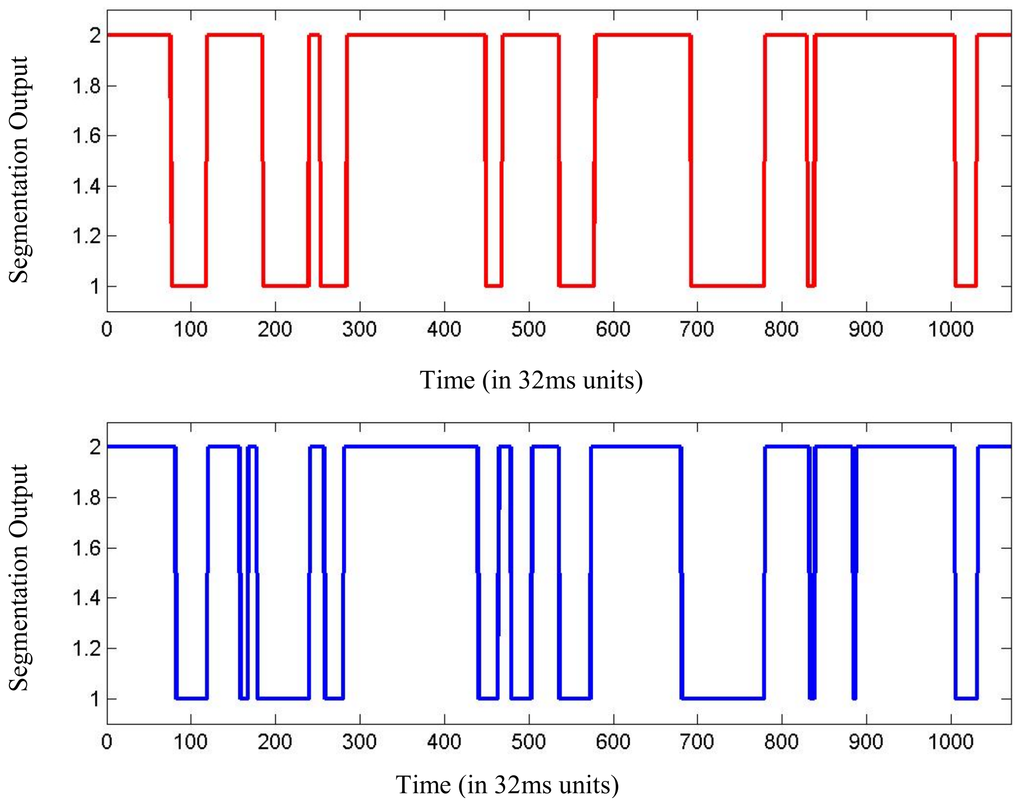

Figure 4-7: The voicing labels of two subjects having a conversation. Note how highly correlated their voicing labels are. This is because each person's microphone picks up the voicing both from the wearer and from their conversation partner.

If we compute the mutual information between the voicing labels of two audio streams that are having a conversation versus any other pair we will observe a sharp peak in the mutual information value when we have the correct pair at the correct time. The conversation mutual information measure is calculated as follows:

$$a_{ij}[k] = I(v_i[t], v_j[t-k])$$
$$= \sum_{m,n} p(v_i[t]=m, v_j[t-k]=n) \log \frac{p(v_i[t]=m, v_j[t-k]=n)}{p(v_i[t]=m)p(v_j[t-k]=n)}$$

where v_i and v_j are two voicing streams from subject *i* and subject *j* and m and n are the states (we have two states: 0 for unvoiced and 1 for voiced frames). Our data streams are all time-stamped, so we can calculate the mutual information between all cross-pairs. Although the clocks of the device were synchronized every other day, there are slight drifts. This affects the mutual information score as it relies on the tight synchrony between the two channels of data. In order to tackle this problem we compute the mutual information value for each one minute segment for all pairs of subjects over $\pm$30s time window (the drift was always observed to be less than 30s). For each one-minute segment, the maximum value within the sliding window is selected as the final score. The choice of using one minute chunks of conversation does put a lower limit on the duration of interaction we can detect. However, if we use a time window that is shorter than one minute then this significantly lowers the accuracy – computing mutual information over a short duration can give rise to many false peaks. Intuitively this makes sense, the mutual information score over a longer duration will result in a sharp peak only if the two channels are synchronized. For short segments it is not inconceivable that we will see random spurious synchronization between two data streams. This will make the mutual information score less reliable by giving rise to many false positives.

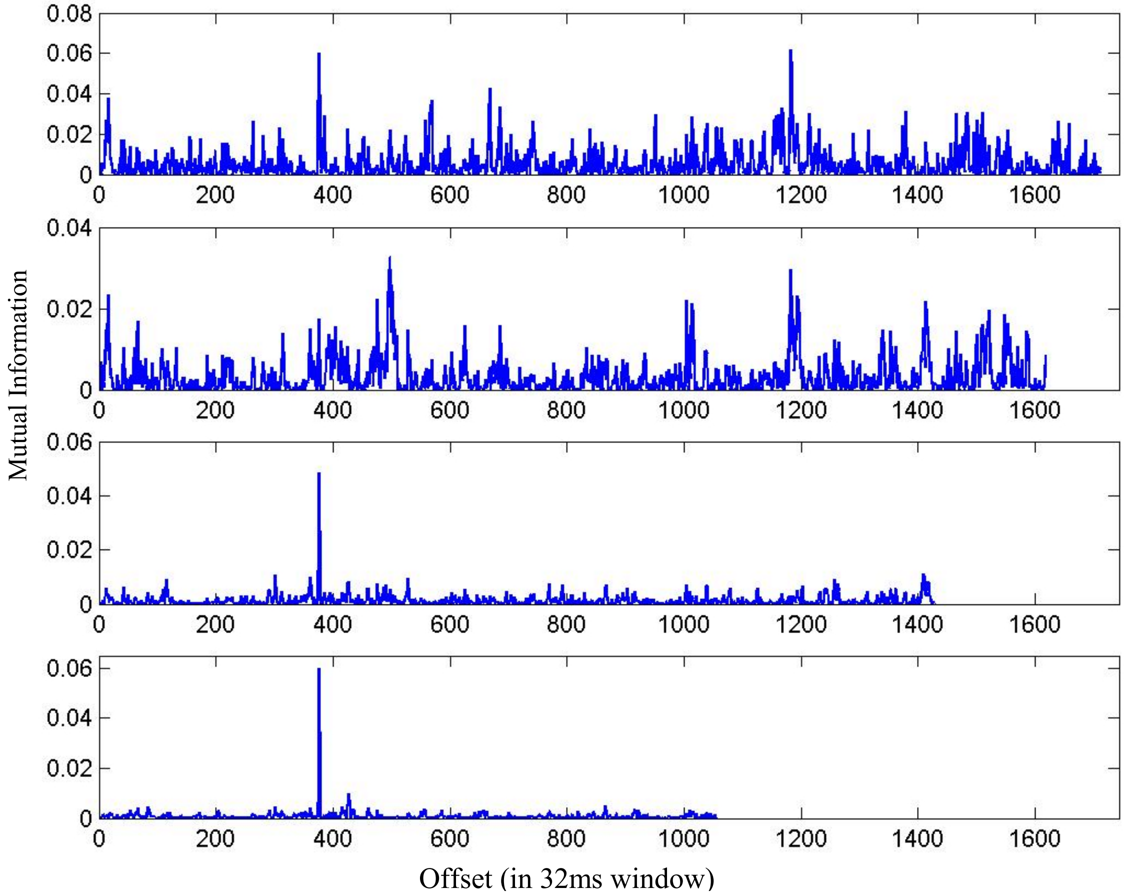

Figure 4-8: Mutual information score with varying window length (a) 15 seconds (b) 30 seconds (1) 1 minute (d) 2 minutes

Figure 4-8 shows the effect we mention above. We compute the mutual information score for 15 seconds, 30 seconds, 1 minute and 2 minutes chunks of data. For each case we calculate the score for the correct pair (i.e. when the data streams were synchronized and the pair were having a conversation) and over multiple other incorrect pairs (pairs that weren't having a conversation). The x-axis is the sample pairs (#376 is the correct pair) and the y-axis is the mutual information score. Notice when we use shorter time windows (15s or 30s) the mutual information is not very informative (lots of false positives) and cannot be used to find out whether a pair is interacting. However, the reliability improves significantly when we use a one minute or two minute chunk of data. For our experiments we calculate mutual information for all pairs (23 choose 2 = 253) for one minute chunks. The consequence of doing this is if a pair does not interact for almost one minute we aren't able to detect that conversation. We feel that this is not a terrible price to pay, as most conversations other than casual greetings are usually longer than one minute.

We tested how well the mutual information score performed in detecting the interactions in our hand-labeled data set. As mentioned in the previous section, we had four different subjects label two days worth of data each. We had a total of 48 hours of data that was labeled. Each five minute chunk was given a label that contained the name of the person whom the subject was interacting with (if the subject was also participating in the experiment) or was labeled as 'conversation with other', 'background speech' or environmental 'background noise'. As there was only one label per five minute chunk, we did not know whether a conversation was less or more than a minute in duration. In order to get separate accuracy numbers for segments greater than a minute, the author labeled her data on a one minute scale. **Table 4-2** gives the accuracy numbers for mutual information based interaction detection.

	Accuracy
All segments (estimated from 48 hours of data from 4 subjects)	63.5%
Segments greater than one minute (estimated from 12 hours of data from 1 subject)	87.5%

Table 4-2: Accuracy in identifying interactions based on mutual information score

4.3 Interaction Network

Now that we are able detect interactions between people, we can analyze the actual communication patterns that occur within the community. We use the voicing labels computed from the participants audio to compute pair-wise mutual information. These labels are obtained prior to speaker segmentation and thus some chunks may not contain any voiced speech from the actual wearer. Consequently, a high mutual information score between wearer A and wearer B can arise from any of the following scenarios: (i) wearer A is participating in a conversation with wearer B (ii) wearer A or B is participating in a conversation with someone else nearby but not to each other or (iii) neither A or B are participating in a conversation but their microphone is picking up another conversation happening nearby. We want to use the mutual information between wearer A and wearer B to identify conversations between them and not the other false positive scenarios described above. We achieve this making sure that both A and B have voiced regions contributed by them (using energy-based speaker segmentation classifier) within the time window that they have a high mutual score.

Figure 4-9(a) shows the mutual information matrix for one of the subjects (ID 8) for a given day. Each row of the matrix is the mutual information between that subject and another experiment participant. The x-axis is time and mutual information is calculated for every one minute segment. Notice the subject has high mutual information score with ID 2, 3 and 20 for about an hour (between minutes 68-122) for that day, however in this case they were all listening to a talk by another individual. Hence, the voicing energy for ID 8 (**Figure 4-9**(c)) for that time window is minimal. Filtering the mutual information score with this energy measure results in **Figure 4-9**(b) which eliminates the false positives that arise from multiple subjects picking up conversations nearby simultaneously. **Figure 4-10**(a-c) shows an example of when there was actual conversation going on between ID 13 and ID 21. We use this method to identify who is talking to whom, when and for how long. As a result, we can again look at the overall network matrix, this time based on conversations rather than proximity. In order to get exact accuracy measures for this matrix, we would need to have all the interactions labeled over all subjects. However, **Table 4-2** presents the accuracy of detecting conversations on a subset of the dataset. As this subset was collected under the same condition as the rest of the dataset, we expect the performance will be similar for the entire the dataset.

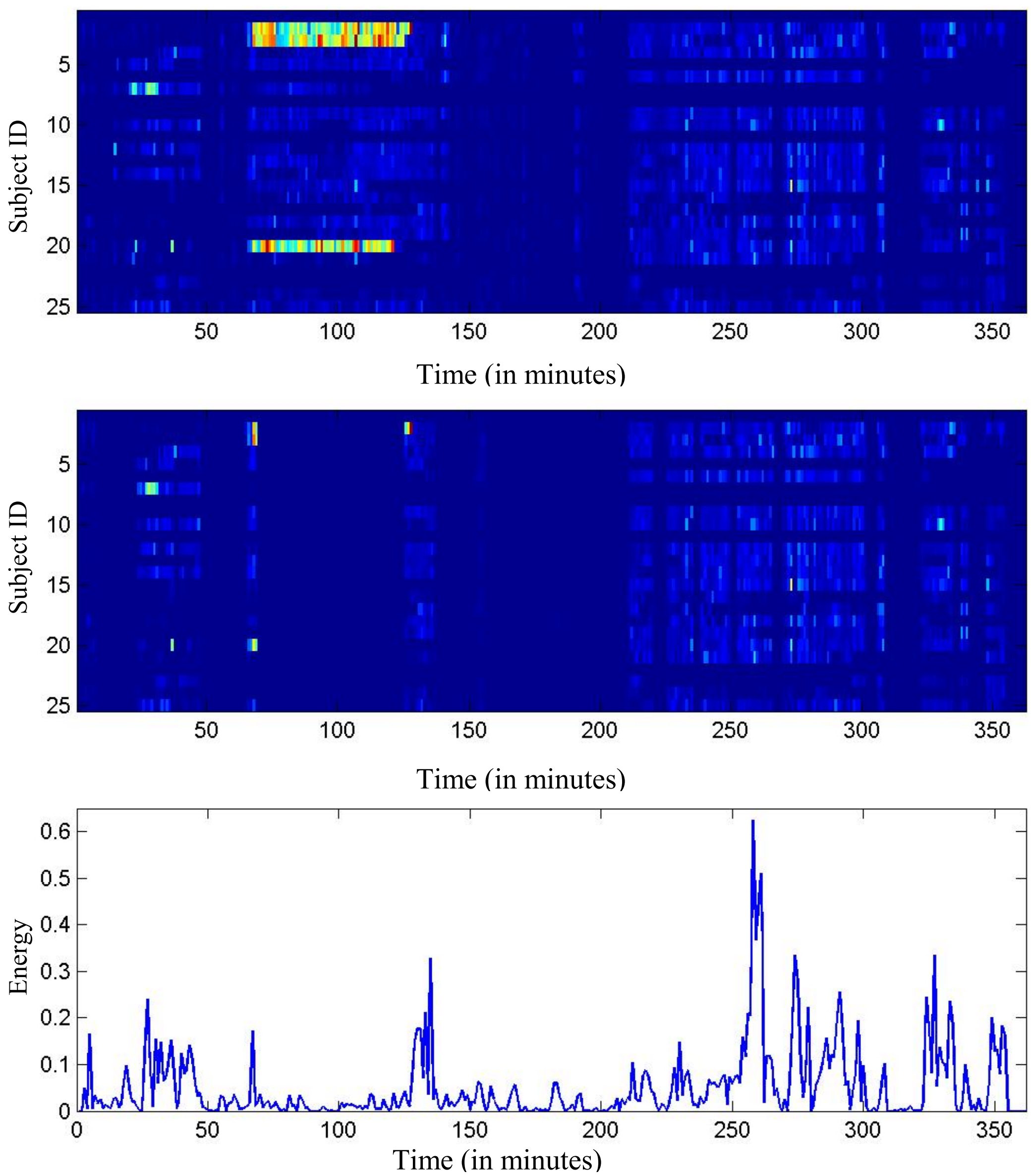

Figure 4-9: (a) mutual information score between ID 8 and all other cross IDs. (b) mutual information score filtered by the voicing energy of ID 8 (c) The voicing energy for ID 8. The x-axis is time in minutes (6hours) and y-axis is the ID numbers.

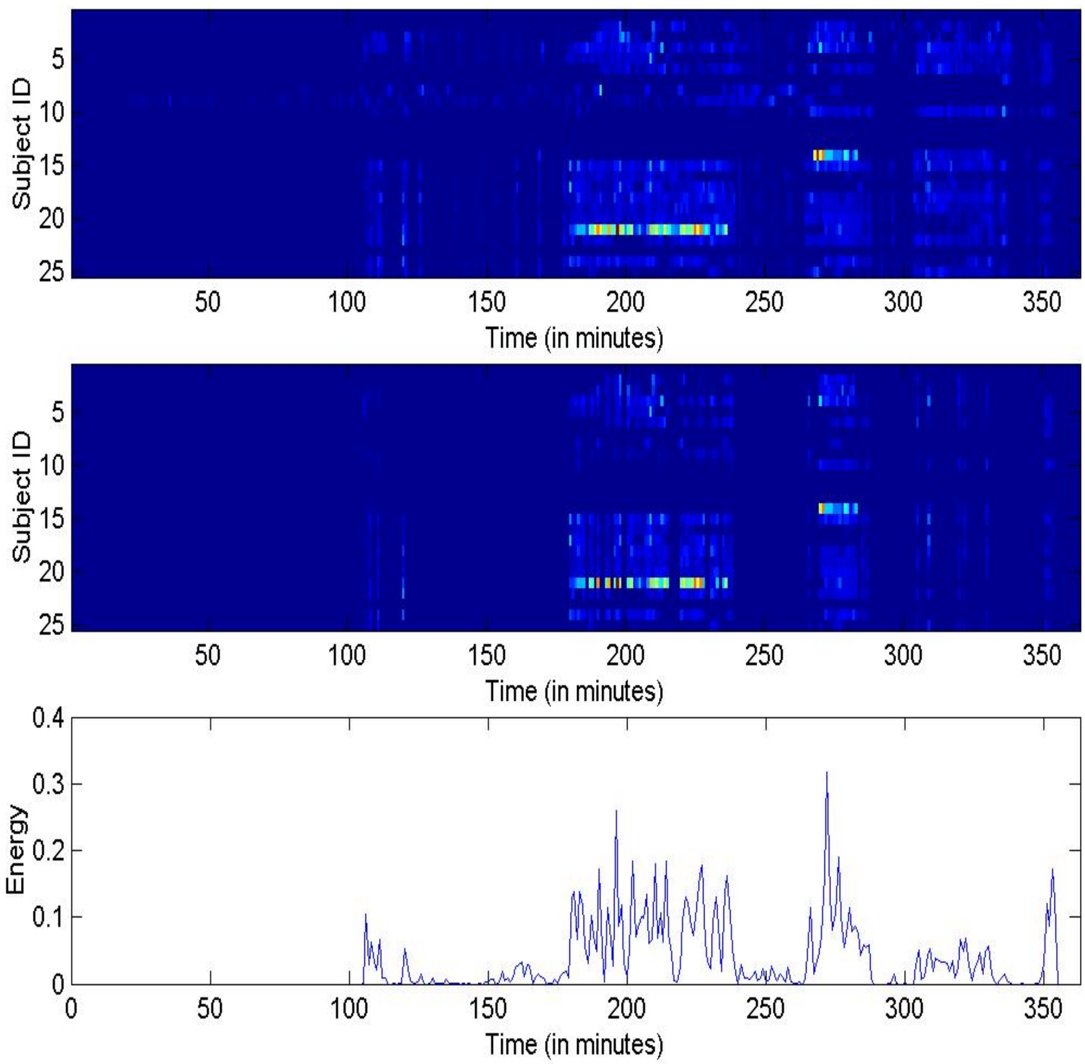

Figure 4-10: (a) mutual information score between ID 13 and all other cross IDs. (b) mutual information score filtered by the voicing energy of ID 13 (c) The voicing energy for ID 13. The x-axis is time in minutes (6hours) and y-axis is the ID numbers.

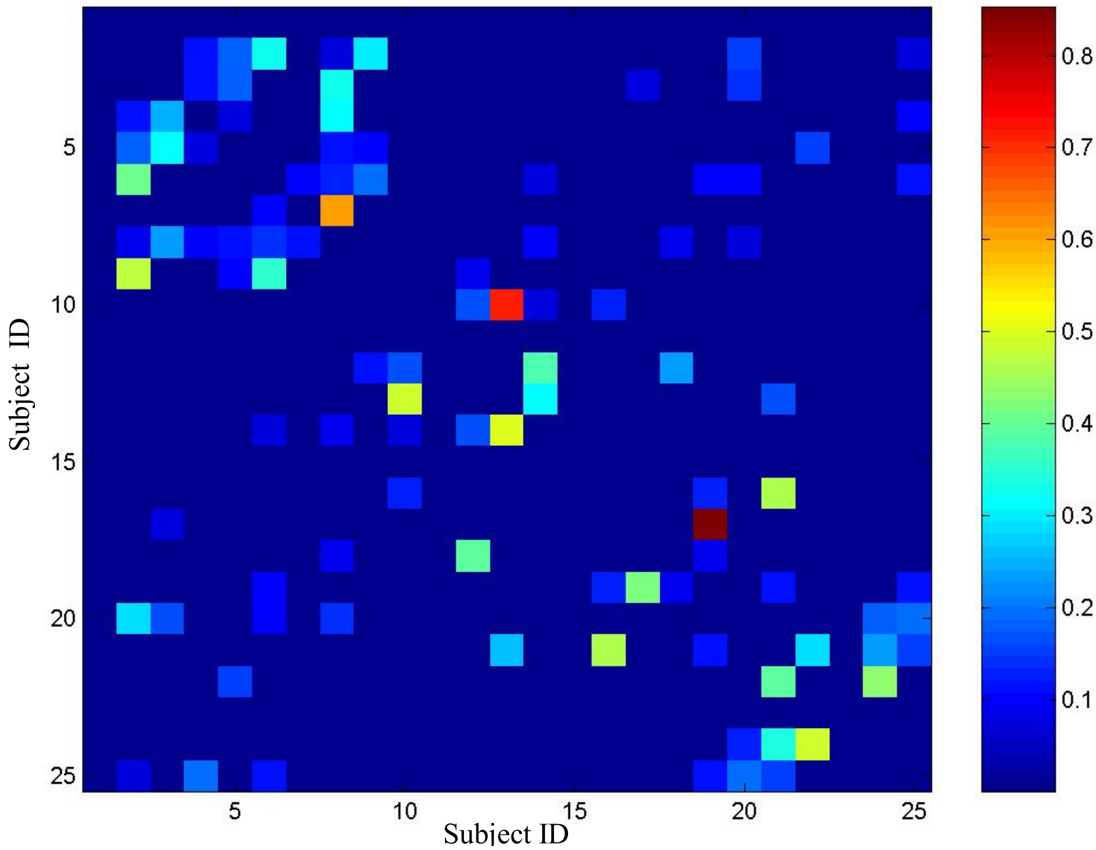

Figure 4-11: Interaction matrix. Each row corresponds to a different person. ID 1 and ID 11 were not assigned. The values are proportional to fraction of total interaction. The microphone of ID 15 failed most of the days we conducted the experiment and ID 23 did not have the sociometer on most of the time. So we have excluded their data in future analysis.

Figure 4-11 is the network matrix based on interaction. Each row shows the interactions of one person with other people in the network. The value of each entry (row *i* column *j*) is equal to person i's interaction with person j as a fraction of person i's total interaction. Subject IDs 2-8 belong to group 1, IDs 9,10,12-14, and 18 belong to group 2, IDs 15-17 and 19 to group 3 and 21-24 to group 4, IDs 20 and 25 were physically co-located with groups 1&2 (no one was assigned ID# 1 or 11). **Figure 4-12** shows each person's ego-network as a bar graph. Note, that there are a few individuals who have broad connections across groups - this type of individuals usually can have an effect on the information flow within the community. To get a more intuitive picture of the interaction pattern within the group of people who were equipped with sociometer, we visualize the network diagram by performing MDS (multi-dimensional scaling) on the geodesic distances. This type of visualization is commonly used in social network analysis [7].

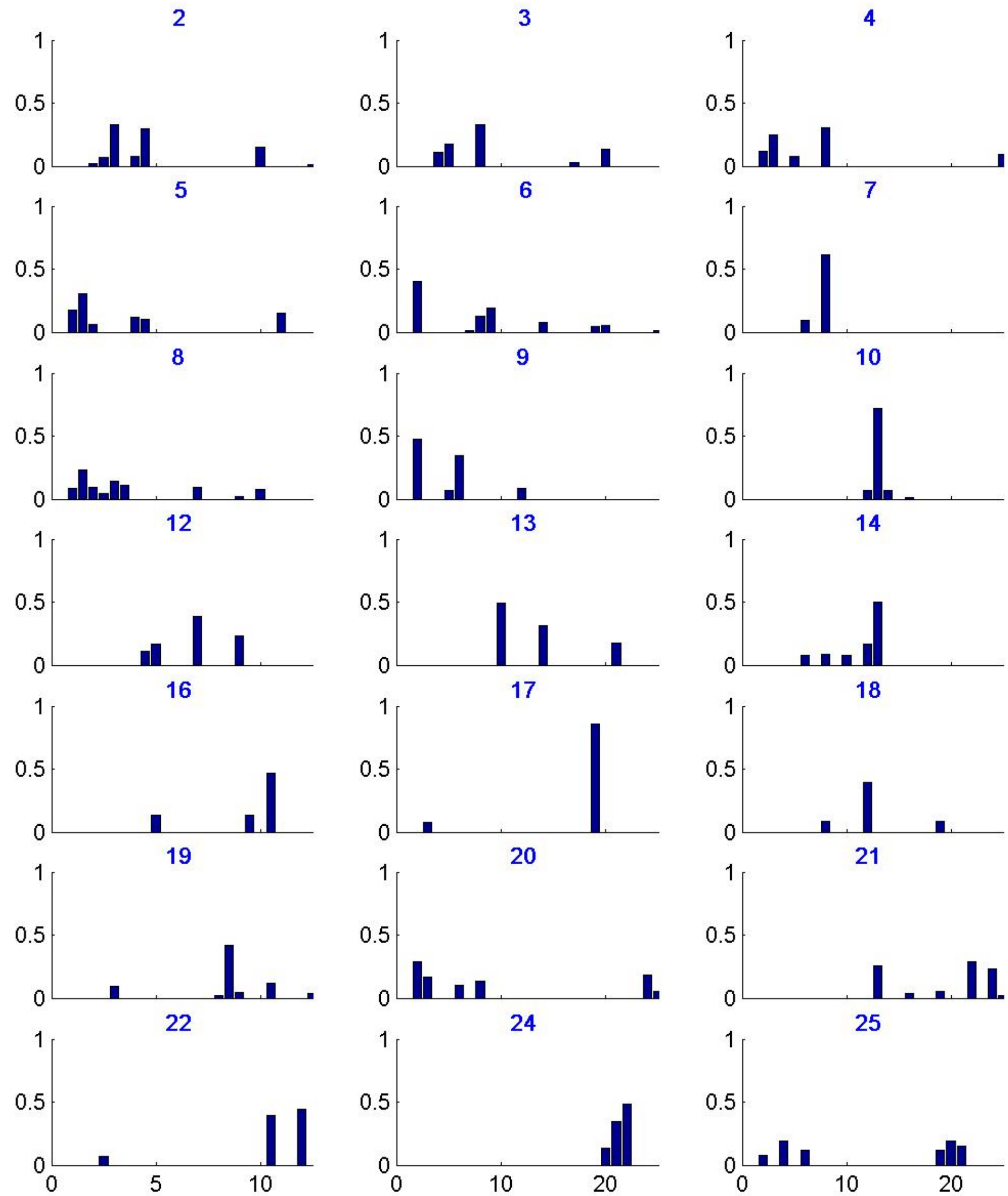

Figure 4-12: Distribution of interaction. Each subplot depicts a subject's interaction with other experiment participants – x-axis shows the subject ID number and y-axis is the fraction of interaction. Note how some people interact with a much smaller subgroup than others.

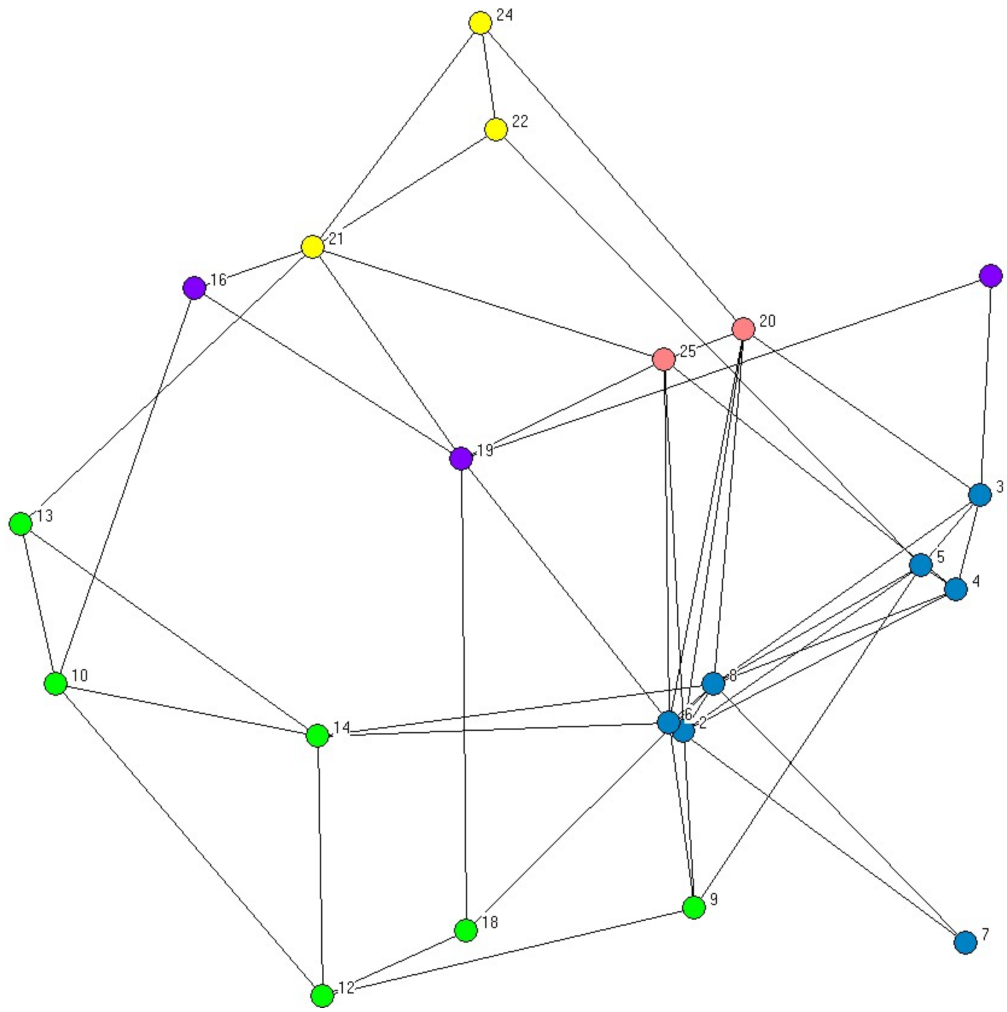

Figure 4-13: Network diagram based on multi-dimensional scaling of geodesic distances. Node numbers represent the subject IDs.

The link structure for the nodes is calculated by thresholding or binarizing the interaction matrix, and the distances between a pair of nodes is the length of the shortest path connecting the two nodes. Multi-dimensional scaling provides a visual representation of the pattern of proximities among the set of people based on some distance measure [53], we use the geodesic distance as our distance metric. MDS method projects points from a higher dimensional space to a lower dimensional space (2D in our case) such that the distances in the projected space is as close as possible to distances in the original space. **Figure 4-13** shows the network visualization obtained via MDS. The nodes are colored according to physical closeness of office locations (using the same colors that were assigned to subjects in Figure 2-5, where the location of the participants were overlaid on the building map). People whose offices are in the same general space seem to be close in the communication space as well. In the next section we show the effect of distance on the overall communication pattern.

4.3.1 Effects of Distance on Face-to-face Interaction

The architecture or the structural layout is known to affect the communication within an organization or community[2, 54, 55]. We measured how the probability of communication changes as the physical distance between the subject increase. **Figure 4-14** shows the probability of communication as a function distance between offices. We grouped distances into six different categories – (i) office mates (x-axis 0) (ii) 1-2 offices away (x-axis 1) (iii) 3-5 offices away (x-axis 2) (iv) office on the same floor (x-axis 3) (v) offices separated by a floor (x-axis 4) (vi) office separated by two floor (x-axis 5).

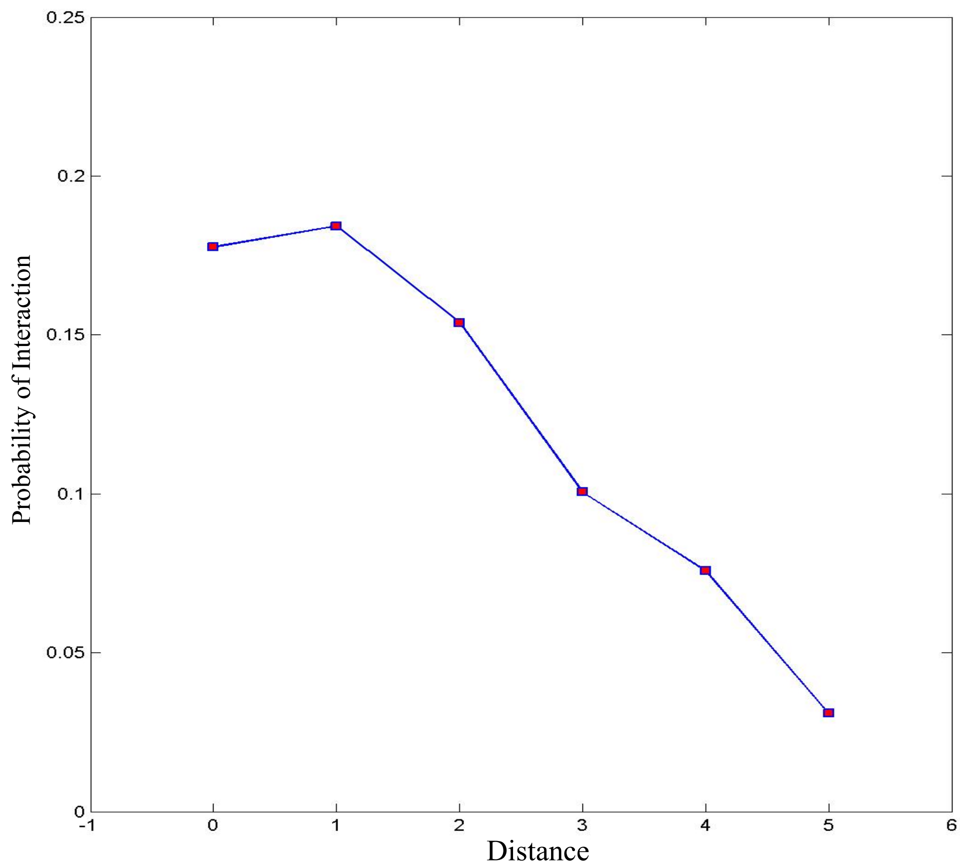

Figure 4-14: Probability of communication as a function of office distance. The distance is grouped into six different categories.

We measured what proportions of a group's interactions are within the group, as opposed to with other groups members. Overall 73% of the interactions were within group and 27% across group. **Figure 4-15** shows the per group breakdown – group 3 however has almost equal amounts of within and across group interactions. This group in general had very few interactions, and was the quietest group in our experiment. This is also consistent with their self reports.

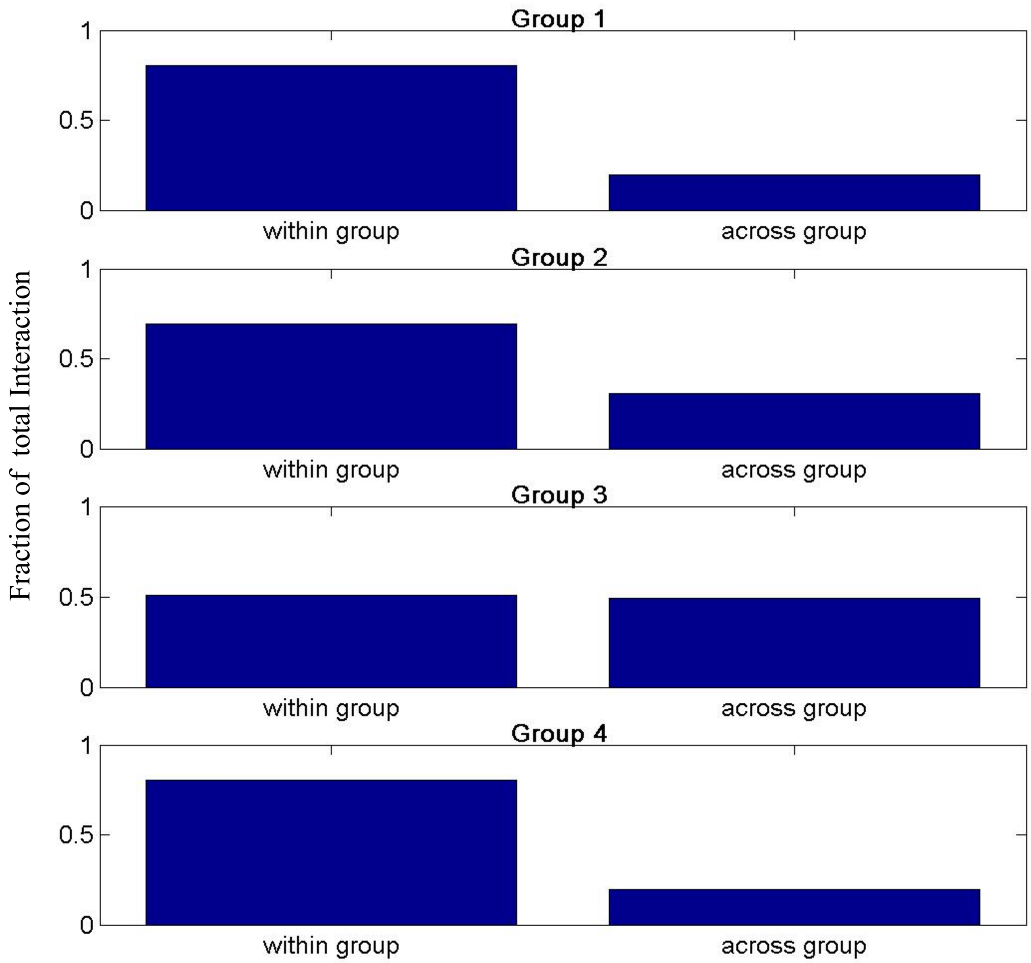

Figure 4-15: Distribution of within group and across group interactions for each of the participating groups

4.3.2 A Few Other Observations

In this section we present some analysis of individual speaking characteristic and how speaking activity changes through the course of the day. These analyses were done using the segmented voicing labels calculated from each subject, i.e. based on the amount of talking time for each person.

Figure 4-16 shows the average minutes of voicing energy above threshold detected per hour for each of the subjects. The least 'talkative' subjects are ID 16 and 17, which also validates our observation that group 3 (ID 16,17 and 19) had the lowest amount of interactions with other participants. Among our subjects, those who were faculty members were more talkative than students.

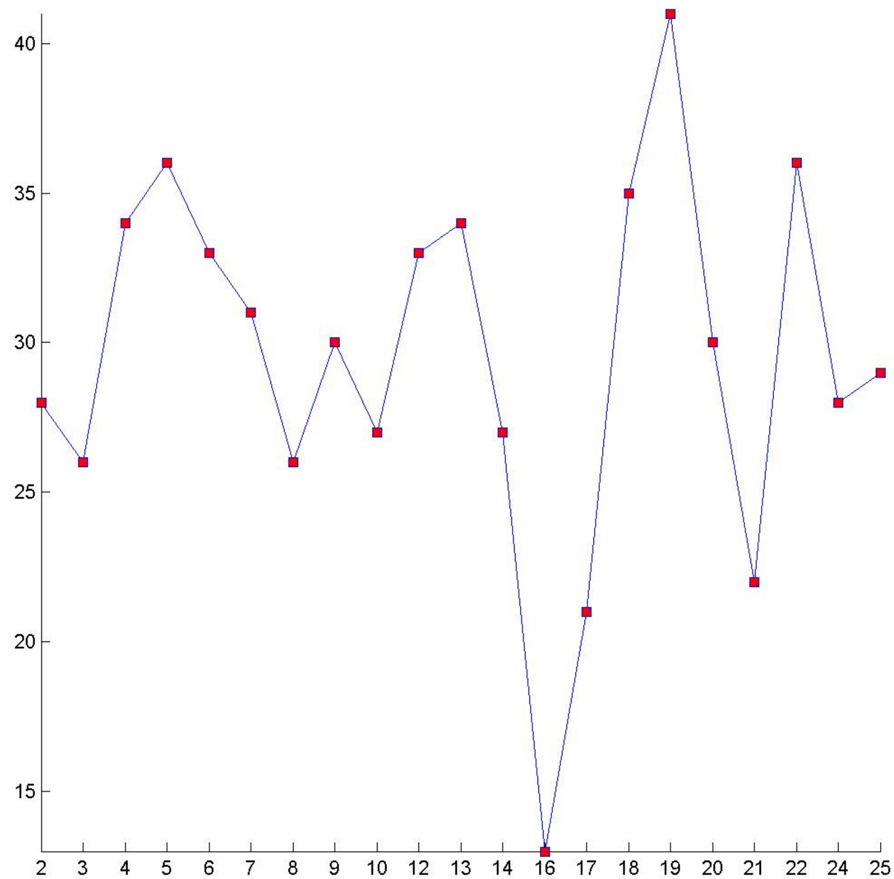

Figure 4-16: Average minutes of voicing per hour for the subjects

The average talking pattern throughout the day was calculated based on the fraction of time that speech was detected from a wearer's device (for every one-minute unit of time). **Figure 4-17** shows the daily pattern averaged over all the subjects and over nine days. This result is quite intuitive, as talking peaks during lunch time and also in the late afternoon when students often take breaks (the weekly Media Lab student tea is also held in the afternoon).

These types of analysis of network behavior are much harder to do using surveys or self-reports, whereas they are easily extracted from the analysis of the sensor data.

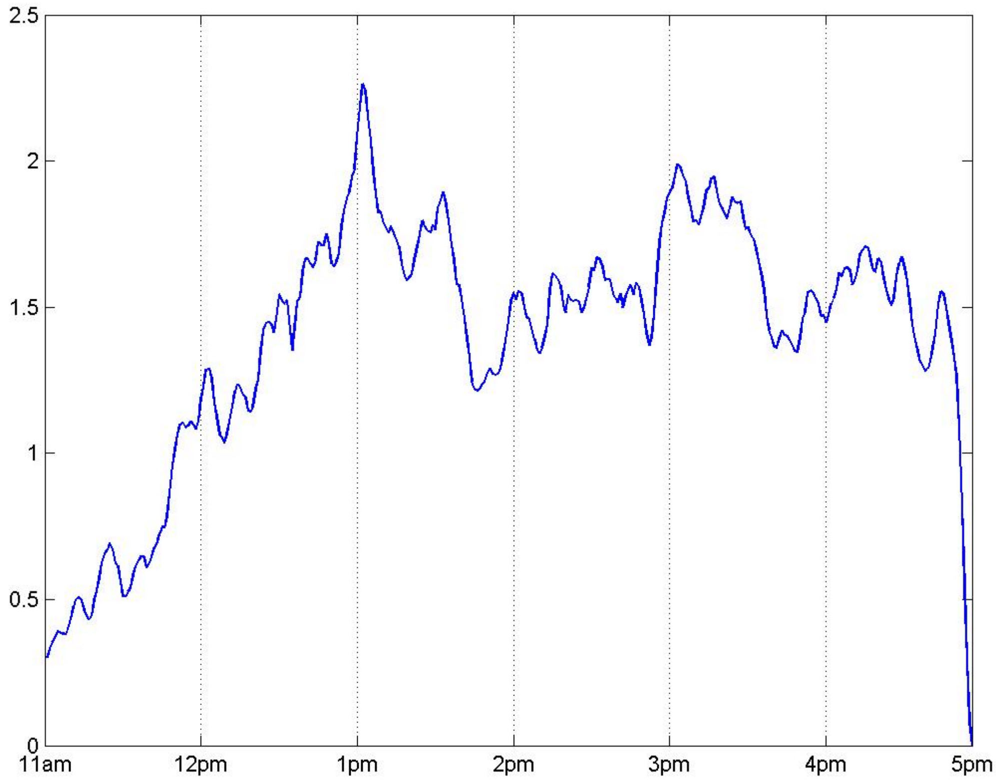

Figure 4-17: Speech activity over the course of the day averaged over all subjects

4.3.3 Centrality Measure

Centrality measures are extensively used in social network analysis to understand an individual's involvement within a community. These measures are primarily non-directional based only on the presence or absence of link between nodes/people and not the directionality of the links. The aim of these measures is to identify individuals who have the most control over or are most central to the network [56]. There are various centrality measures based on the number of links, as well as the importance of the links in maintaining the connectedness of the network.

One particular measure is the '*betweenness centrality*', which tries to measure how much control an individual has over the interaction of other individuals who are not directly connected [57]. The idea behind betweenness centrality is that if an individual lies between the shortest path of two other individuals then he/she has control over their interaction. Shimbel [58], who was one of the first people to recognize the importance of betweenness stated the following:

> *Suppose that in order for (actor) i to contact (actor) j, (actor) k must be used as an intermediate station. (Actor) k in such a network has a certain "responsibility" to (actors) i and j. If we count all of the minimum paths which pass through (actor) k, then we have a measure of the "stress" which (actor) k must undergo during he activity of the network. A vector giving this (count of minimum paths) for each (actor) of the network would give us a good idea of the stress conditions throughout the system"*

If g_{jk} are the number of geodesics linking two actors and all the geodesics are equally likely to be chosen. And if individual i is involved in $g_{jk}(n_i)$ geodesics between j and k. The betweenness for i is calculated as follows:

$$C_B(n_i) = \sum_{j<k} g_{jk}(n_i) / g_{jk}$$

Individuals with high betweenness play a role in keeping the community connected and removing someone who has high betweenness can result in isolated subgroups [14, 57].

We estimated the betweenness centrality of the communication network as shown in **Figure 4-11**, however the link values are converted to binary 0/1 values. Some notable aspects of the betweenness measures shown in **Figure 4-18** are the following: (i) ID 8 was assigned to the author who communicated with the subjects to coordinate data collection and hence the high betweenness measure may be biased (ii) ID 7 was an undergraduate student working with the author, note ID 7 mainly interacts with ID 8 and has a betweenness of 0 (iii) ID 16 and 17 are our least communicative subjects, and have low betweenness scores (iv) ID 4, although very tightly integrated within group 1, is mostly isolated from other groups and consequently has a low betweenness score.

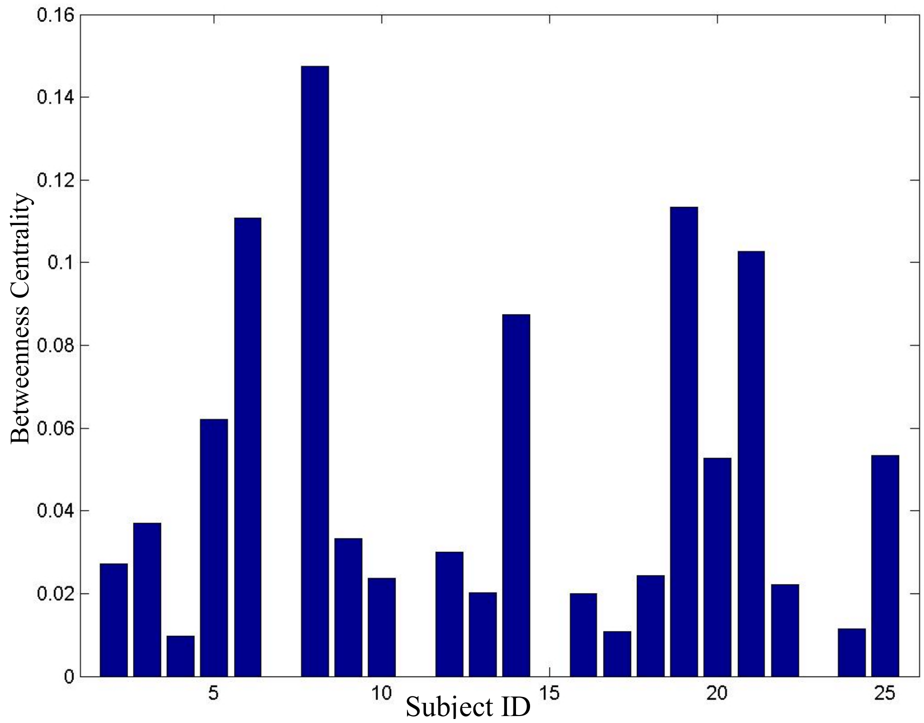

Figure 4-18: Betweenness centrality of the interaction network

4.4 Summary

In this chapter, we started by analyzing the face-to-face proximity network for the IR data, quickly realizing that proximity information alone leaves out much of the interesting details of the social network. In order to get those details, we need to know who talks to whom within the community. We describe how we model and segment speech, speaker and interactions using the audio features described in Chapter 3. Then we analyze the social networks based on real interactions and measure various properties of the network, including the effects of different individuals and physical distances on the communication patterns within the network. We also compute the centrality score of the individual in our network, a measure that is often used in social network analysis to identify the importance of various people in maintaining the connectivity of the overall network. We compare the performance of various algorithms, using either survey data or hand-labeled data from the subjects.

Chapter 5: The Dynamics of Interaction

We mentioned as one of the goals of this thesis that we want to model the dynamics of a person's interaction and how one person's dynamics affects another's style of interaction. So far we presented the overall characteristics of interactions observed among the individuals in our network. But with the necessary features and tools at hand we are ready to embark on our final task – modeling the dynamics of interaction and how people influence each others dynamics. In this chapter, we will describe how we go about quantitatively modeling a person's dynamics during interactions and how this dynamics changes with changing conversational partner. How do different individuals 'influence' our interaction style? What can we say about those who tend to influence other people? We also look at how we can use the make use of the network structure to construct a dynamic model that could capture how messages/information may propagate within that community.

5.1 Turn-taking patterns

The main aspect of the interaction dynamics we will focus on is turn-taking. We want to measure the turn-taking patterns of individuals and how they differ from each other. We will use these individual dynamics to later estimate how much an individual's overall pattern changes during her interaction with another specific individual.

We start by defining a "turn". For each unit of time we estimate how much time each of the participants speaks, the participants who has the highest fraction of speaking time is considered to hold the "turn" for that time unit. For a given interaction, we can easily estimate how a pair participating in the conversation transitions between turns. In Chapter 4, we described how we segment speech, speaker and interactions. We use the speaker segmentation output within conversations to estimate the turn-taking transition probability. Because most of the conversations in the dataset are between pairs, we transition between two states: speaker A's turn and speaker B's turn. We selected eighty conversations which were on average of 5 minutes long to compute the individual turn-taking dynamics. In selecting the conversation we made sure that we had at least four different conversation partners for each individual and multiple conversation instances for the same conversational pair.

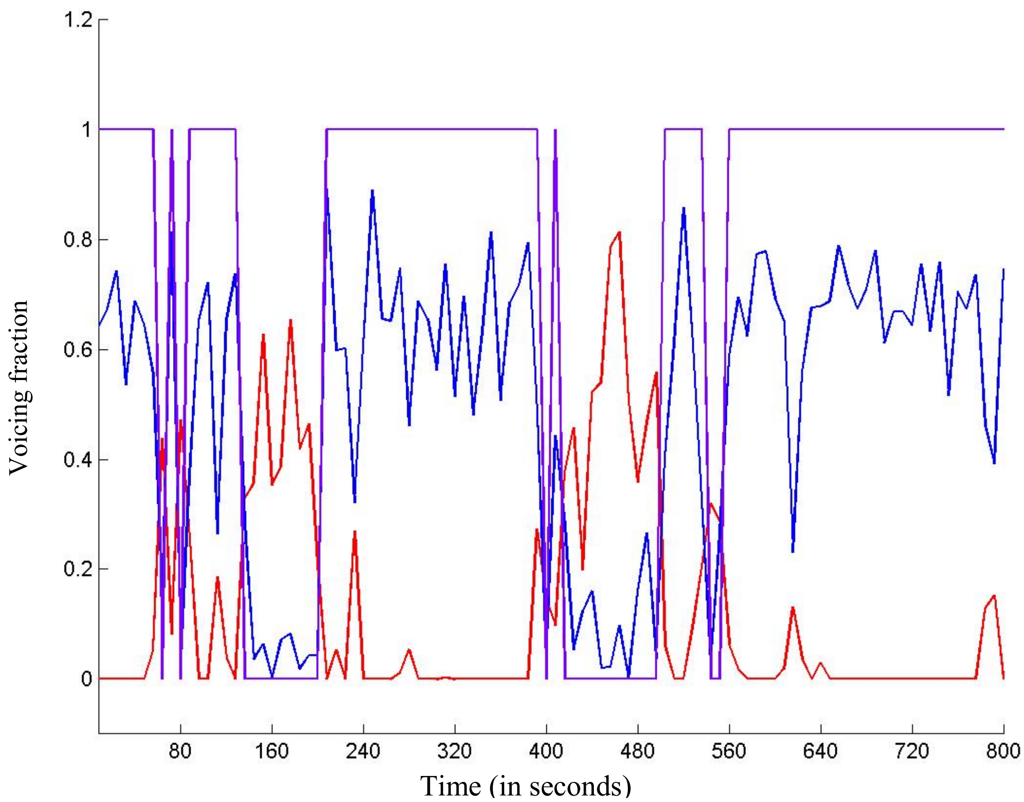

Figure 5-1: Example of turn-taking pattern between a pair. The purple line identifies who is holding the turn. The red and blue line the fraction of speaking time for the two people.

Once we have estimated the turn-taking transition probabilities for the individuals we can measure how similar or dissimilar they are from each other. **Figure 5-2** shows the output of multidimensional scaling of the transition probabilities using a Euclidean distance metric (which is equivalent to projecting onto the first two principal components). The x and y axes here are the projection coefficients for the two principal components – the x-axis mainly varies based on individual's overall talk-time and y-axis varies based on the duration an individual holds her turn. So, a point in the upper left corresponds to an individual who talks little and does not hold her turn for a long time, and a point in the lower right represents a talkative person who holds her turn for longer. From **Figure 5-2**, we observe that a number of people are quite similar to each other in their turn-taking behavior and are clustered in the center; these individuals tend to talk as much as they listen, e.g. IDs 19, 12, 20 etc. Some of the people on the edges have more skewed turn-taking and turn-holding patterns, e.g. IDs 9, 6, 17, and 22. The ellipses around each individual's average style show the variance in that speaker's turn-taking over different conversations. **Figure 5-2** demonstrates that individuals have distinctive turn-taking styles and that these turn-taking patterns are not just a noisy variation of the same average style. Certain individuals vary their styles more than others and later in this chapter we will show how the extent of variability is related to the person's role within the social network obtained from betweenness centrality. In this figure, we show the turn-taking styles of seventeen individuals out of the twenty-one that we have been analyzing so far, because these seventeen are the ones who were participants in the eighty conversations selected. Later we will use ten of these subjects to do further analysis (IDs 2-10 and 14). As we mentioned earlier in this chapter, we want each subject to have at least four different conversation partners, in order to estimate their mean behavior and how they change or are influenced by other people's interaction behavior. Calculating influence measures for a subject who has only one conversation partner would be pointless, as there would not be any difference between the person's overall turn-taking behavior versus her behavior with a specific individual.

In this thesis, we have recorded and analyzed turn-taking styles at a finer time granularity than had been done before, and also over multiple conversations in natural settings. Such analysis would be extremely difficult or almost impossible to do using human-coders. Being able to automatically estimate people's interaction style allows us to study more concretely the differences between individual turn-taking styles. We explore the connection between the role of individuals within their social network and how they communicate with others. Furthermore, automated data analysis opens the possibility to study whether one's gender or formal position in the network has any influence on one's interaction style.

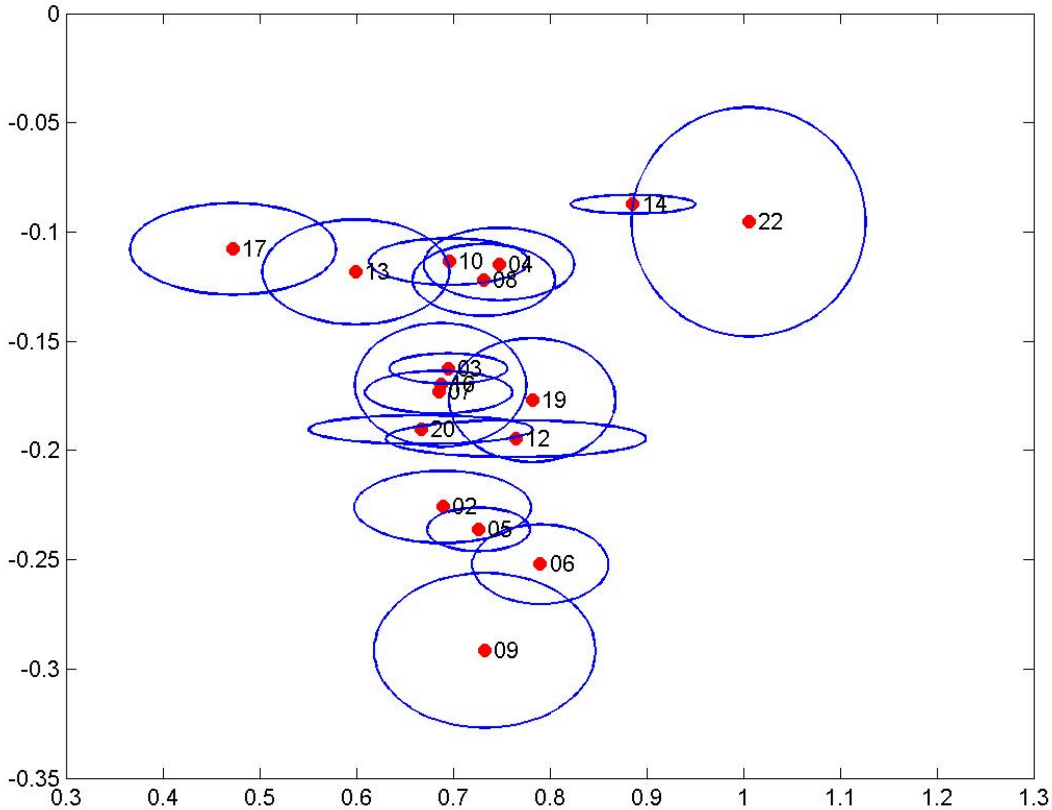

Figure 5-2: Multidimensional scaling of the average turn-taking transition tables estimated from eighty conversations averaging five minutes each. Each individual's mean transition is given by the red circle and the ellipse around shows the variance in speaker's style over different conversations.

5.2 Estimating Influences from Turn-taking Dynamics

When we talk to other people we are influenced by their style of interaction. Sometimes this influence is strong and sometimes insignificant – we are interested in finding a way to quantify this effect. We probably all know people who have a strong effect on our natural interaction style when we talk to them, causing us to change our natural style as a result. For example, consider someone who never seems to stop talking once it is their turn to talk, she may end up imposing her style on us, and we may consequently end up not having enough chance to talk, whereas in most other circumstances we tend to be an active and equal participant.

How would we go about quantifying these type of influences? When we calculate the average transition probabilities for individuals over all the conversations they participate in, we learn how much individuals hold their turn and how much they transition back and forth with their conversation partner. So the average transition table also describes the turn-taking behavior of the individual's average partner. Consequently, if we rotate the transition table by 180º we will get the turn-taking dynamics of the average partner (see **Figure 5-3**) – the matrix entries will remap as follows: (1,1) → (2,2), (1,2)→(2,1), (2,1)→(1,2) and (2,2)→(1,1)

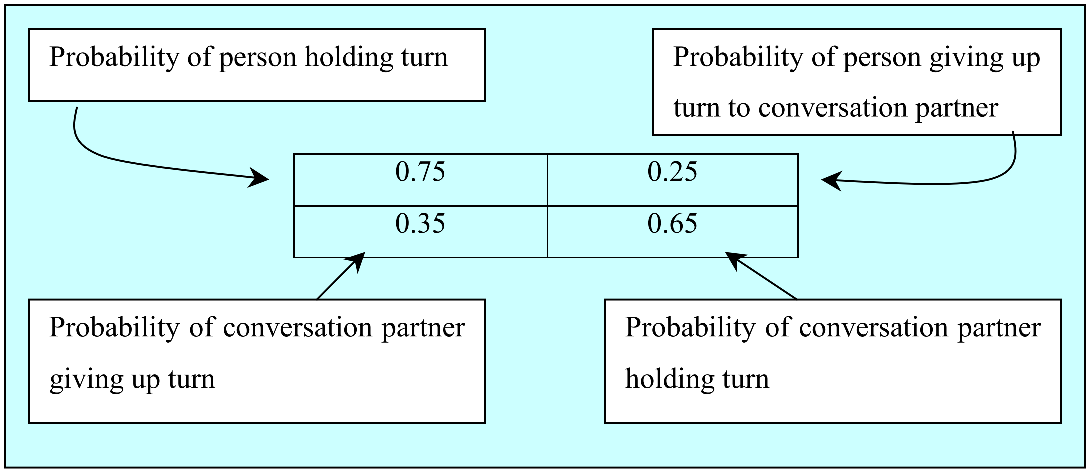

Figure 5-3: An example average transition table for a person's turn-taking behavior. Rotating the transition matrix by 180 degrees gives us the turn-taking behavior of the 'average' conversation partner for that person

When two people are interacting it is plausible their average turn-taking dynamics will affect each other and the resulting turn-taking behavior for that interaction will be a blend of the two transition matrices. If someone affects our average pattern a lot we may adapt to the behavior of that person's 'average conversation partner', if we are not affected at all we will probably maintain our average dynamics completely, or the resulting interaction behavior may be somewhere in between the two extremes. How would we model and learn the affect of the interaction patterns on each other?

Let us recall the Influence Model from chapter 3, where we described a computational framework to model effects like the one we describe above. In the influence model the transition probability of interacting dynamic processes is formed by taking a convex combination of the pair-wise transition probabilities. For a two chain (two person) model the transition probabilities will be –

$$P(S_t^1 \mid S_{t-1}^1, S_{t-1}^2) = \alpha_{11} P(S_t^1 \mid S_{t-1}^1) + \alpha_{12} P(S_t^1 \mid S_{t-1}^2)$$
$$P(S_t^2 \mid S_{t-1}^1, S_{t-1}^2) = \alpha_{21} P(S_t^2 \mid S_{t-1}^1) + \alpha_{22} P(S_t^2 \mid S_{t-1}^2)$$

we will call the two person interaction model the blended interaction model.

The cross transition matrix in this case $P(S_t^1 \mid S_{t-1}^2)$ and $P(S_t^2 \mid S_{t-1}^1)$ are the 180º rotated transition matrix of the conversation partner. Now by learning the α_{ij} values we can measure the effect of influence on one another's turn-taking behavior.

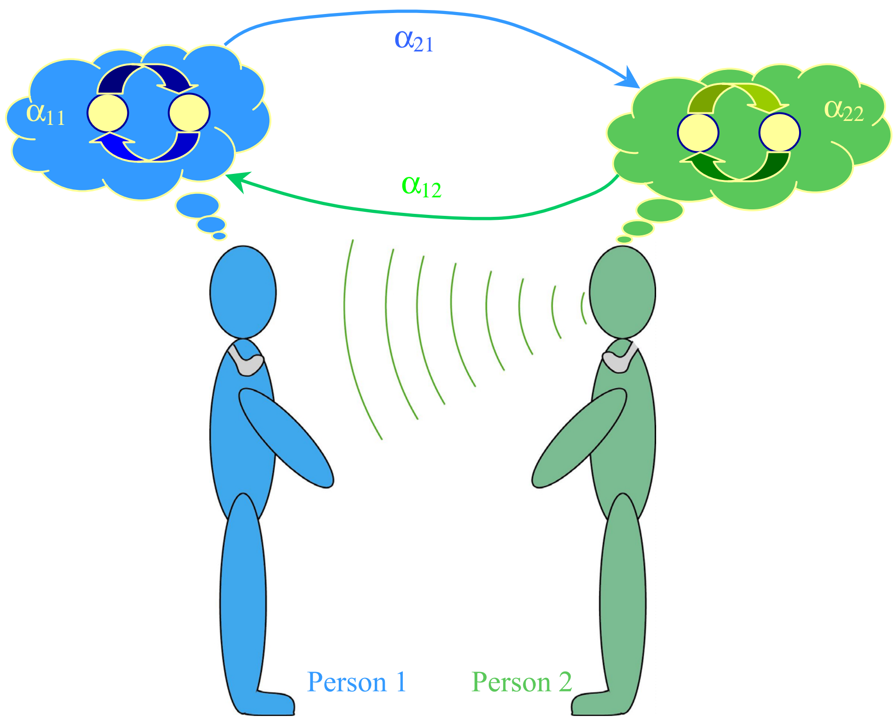

Figure 5-4: A cartoon representation of how two people influence each other's style of interaction

We learn the influence values for all the pair-wise conversations we selected. In order to evaluate whether this model buys us any additional benefit over than just using the personal model, we estimate the reduction in KL divergence by using the blended interaction vs. using the personal interaction model. By using the blended model we were able to reduce the KL divergence by 32%. Since in the blended model we add extra degrees of freedom we tested if the better fit was statistically significant by using the F-test – the p-values was less than 0.01 implying the blended model is a significantly better fit to the data. In calculating the F-test score, we used the sum of square difference between the transition tables instead of the KL divergence score. To make sense of all the pair-wise influences we aggregated all the influences an individual has on all other conversation partners she interacts with. This would identify individuals who overall tend to affect the others' dynamics to a greater degree than average. **Figure 5-5** show this aggregated influence value for the subset of people participating in the set of eighty conversations analyzed.

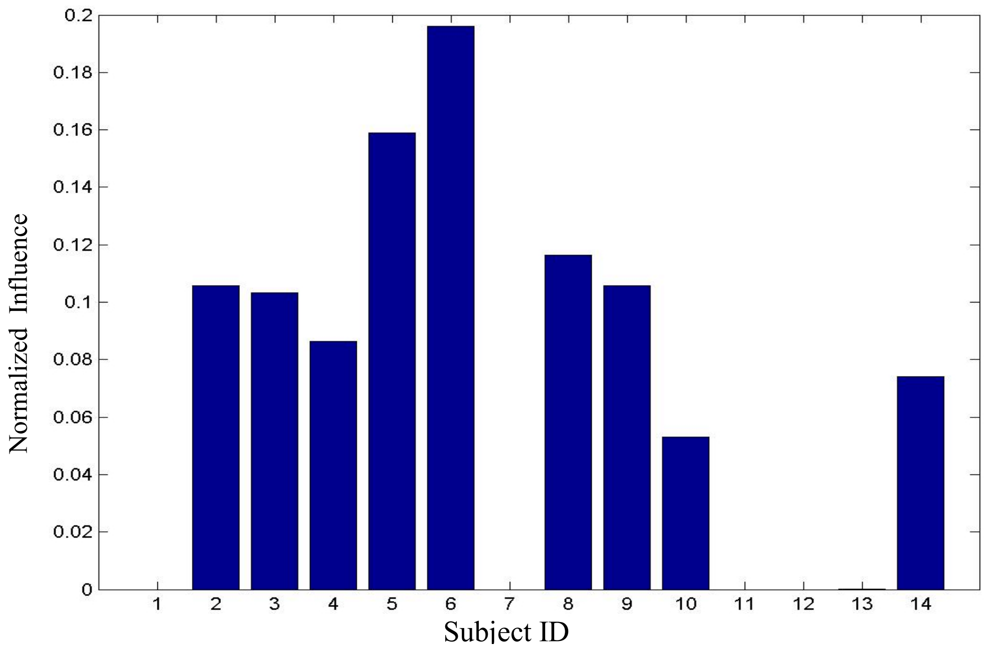

Figure 5-5: Aggregate influences scores of a subset of the subjects

One aspect of the influence values we observed was that if two interaction matrices are very similar (KL divergence close to zero) the influence scores often tend to be very high for the conversation partner and low for the subject. This influence is not very meaning however, as both individuals behave similarly and the resulting blended is almost the same individual model. So we re-estimated the aggregate influences by weighting the influence values by the KL divergence between the transition matrices used in estimating them.

5.2.1 Correlating Influence Values with Centrality Scores

In Chapter 4 we calculated the centrality scores for the subjects, which is a measure of an individual's importance in maintaining the connectedness of the network. We discovered a suprising and highly statistically significant correlation between a person's influence score and their centrality. **Figure 5-6** shows the weighted influence values along with the centrality scores. It seems like a person's interaction style is indicative of their role within the community based on centrality measure. Note that ID 8 (the author) is somewhat of an outlier – a plausible explanation for this can be that during the data collection ID 8 went and talked to many of the subjects (which is not her usual behavior). This resulted in the author having high centrality (based on link structure) but not high influence based on her interaction style.

We computed the correlation between the influence values and the centrality scores, both including and excluding the outlier subject ID 8. The correlation excluding ID 8 was 0.90 (p-value < 0.0004, rank correlation 0.92) and including ID 8 it was 0.48 (p-value <0.07, rank correlation 0.65). The two measures, namely influence and centrality, are highly correlated, and this correlation is statistically significant when we exclude ID 8, who was the present author, and whose centrality was therefore artifactually large due to being the coordinator of the sociometer project.

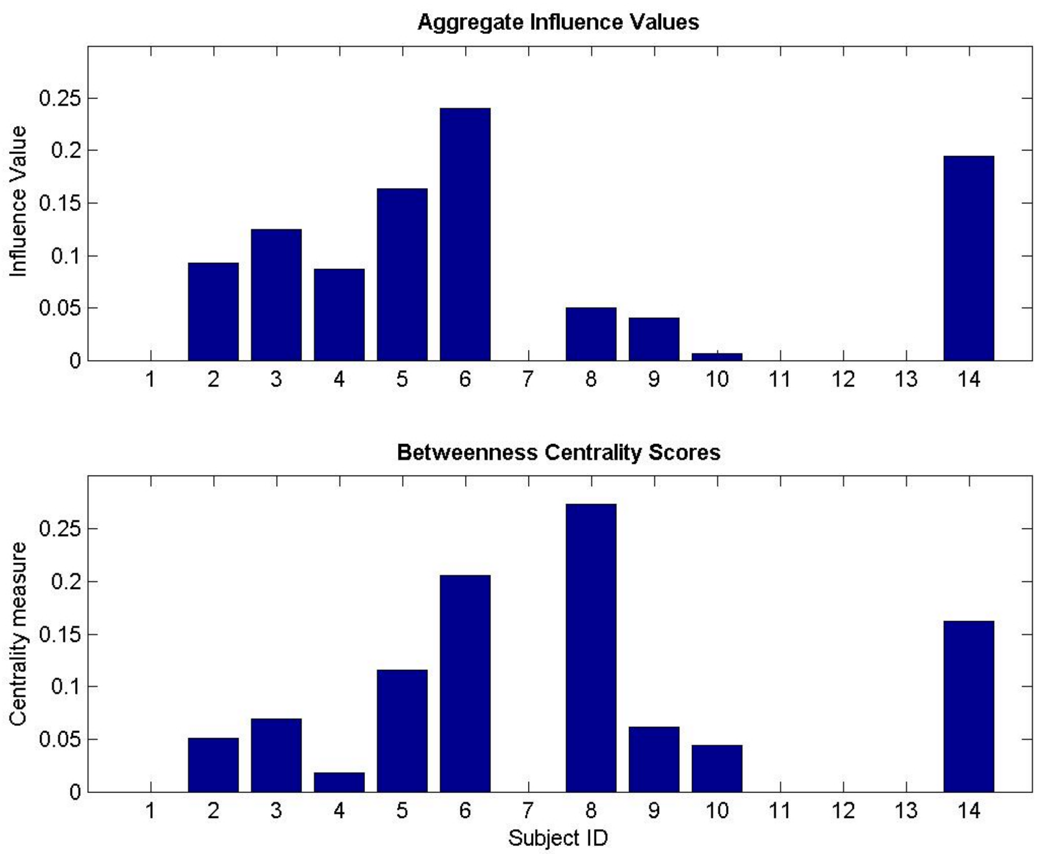

Figure 5-6: Weighted influence value and centrality measure for a subset of the subjects

5.3 Propagation of Information

An important property of networks is how information spreads. The pattern of people's communication regulates how information propagates within the community. We have already calculated the communication link structure of the people within our experimental community in Chapter 4. We used this link structure to calculate the centrality score of the people. However, these links were binary, based on whether a person had interactions with someone else or not. But as we saw from the interaction matrix, also in chapter 4, the amount of communication that people have with others varies significantly. We can use the interaction amount to weight the links with others in the network. A weighted link structure can be used to build a dynamic model of interacting objects (or people), as proposed in the influence model. Using the weights in the link matrix as the influence values, we can build a model of propagation of information.

In this section, we build a simple model of information flow using the influence model framework. All the analysis presented in this section is obtained via a set of simulations, as during the data collection phase we did not have any experiments that would test and record how a real piece of information propagated. We assume the influence a person has on another is the fraction of her total interaction with that person. We also assume, once a person has some information she has it forever – thus other people can only effect the person's states if she doesn't have the information already, i.e. information can not be taken away. We do not make distinction between probability of communication and probability of exchanging information. If two people are likely to interact then they are just as likely to share information. Although quite a simplistic approach, it seems intuitively plausible and simple threshold based models have been proposed for modeling diffusion of information [8, 59].

We use this model to estimate how long it would take for a message or piece of information to reach the entire network for different initial messengers. If someone is not embedded within the community we can imagine it may take a lot longer for the message to diffuse through the network. We sample from the network using different initial messengers and estimate the time it takes for the message to spread throughout the community. For each subject, we sampled 20 sequences in order to estimate the average spreading time. **Figure 5-7** shows the average time needed for a message to spread given different initial messenger. For each subject the figure shows the fraction of population reached for each incremental unit of time. If message originated at ID 16 it takes the longest amount of time to spread through the network – ID 16 also was the least talkative with very low centrality.

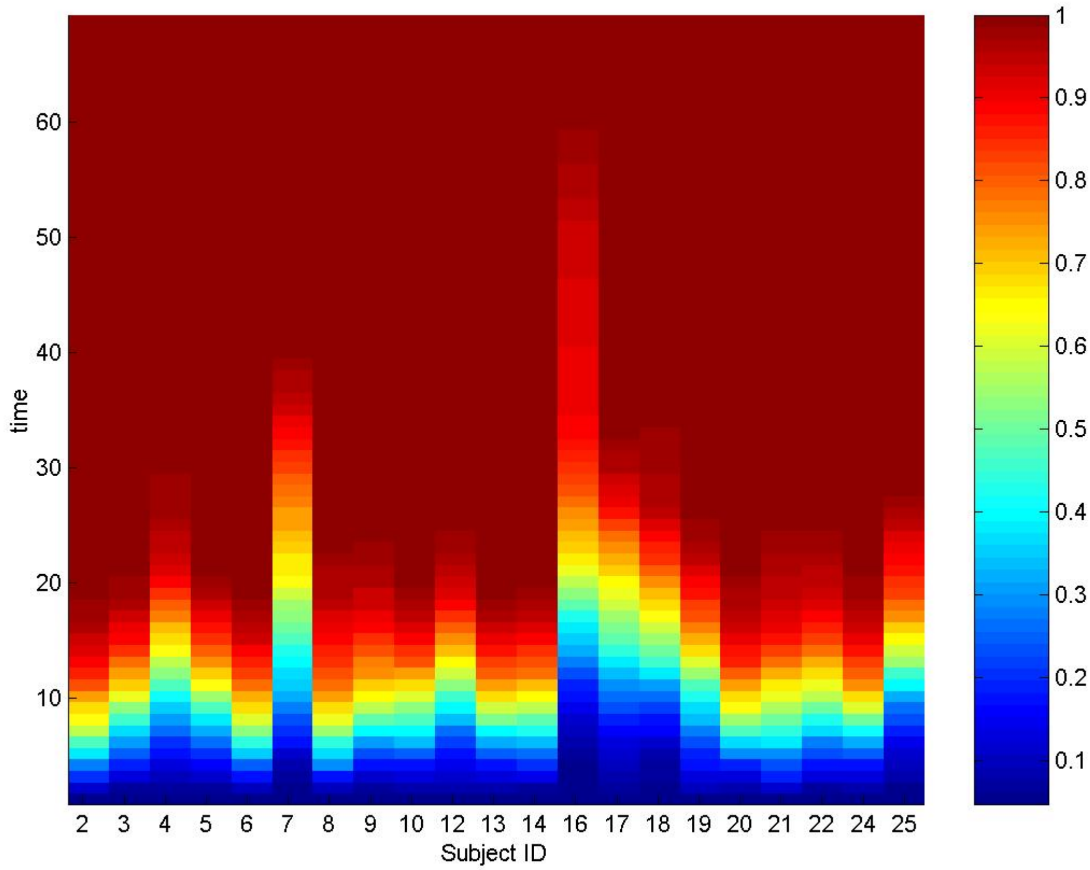

Figure 5-7: The time it takes for a piece of information to spread for different initial messenger

In the next two figures, we show a sample simulation of how message might propagate for two different initial messengers. In the first case, the subject has low aggregate influence, low centrality, and in the second case the subject has high aggregate influence and high centrality values. Consequently, the time to spread is also significantly different for the two subjects.

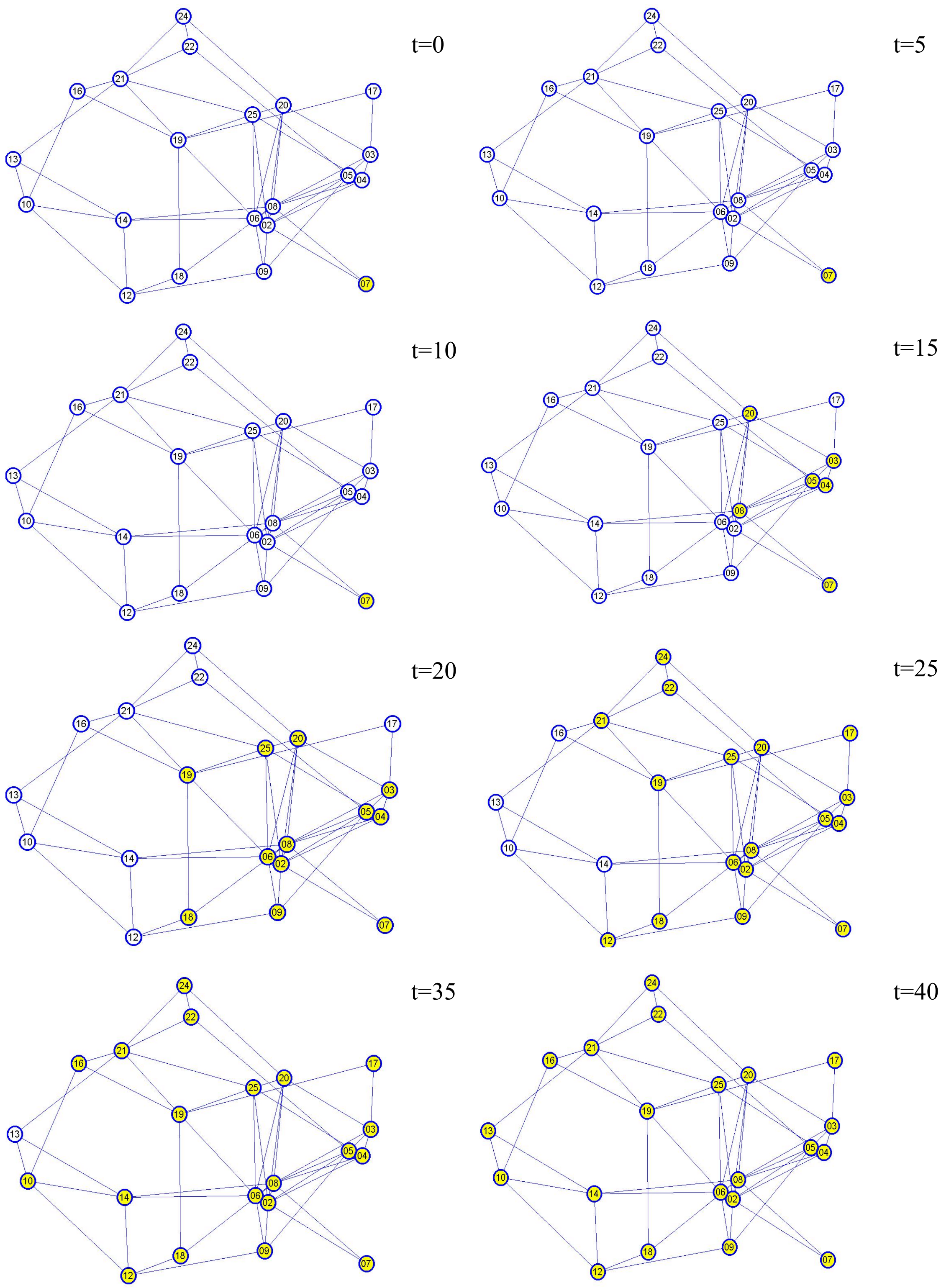

Figure 5-8: Message spreading when initial messenger is ID 7

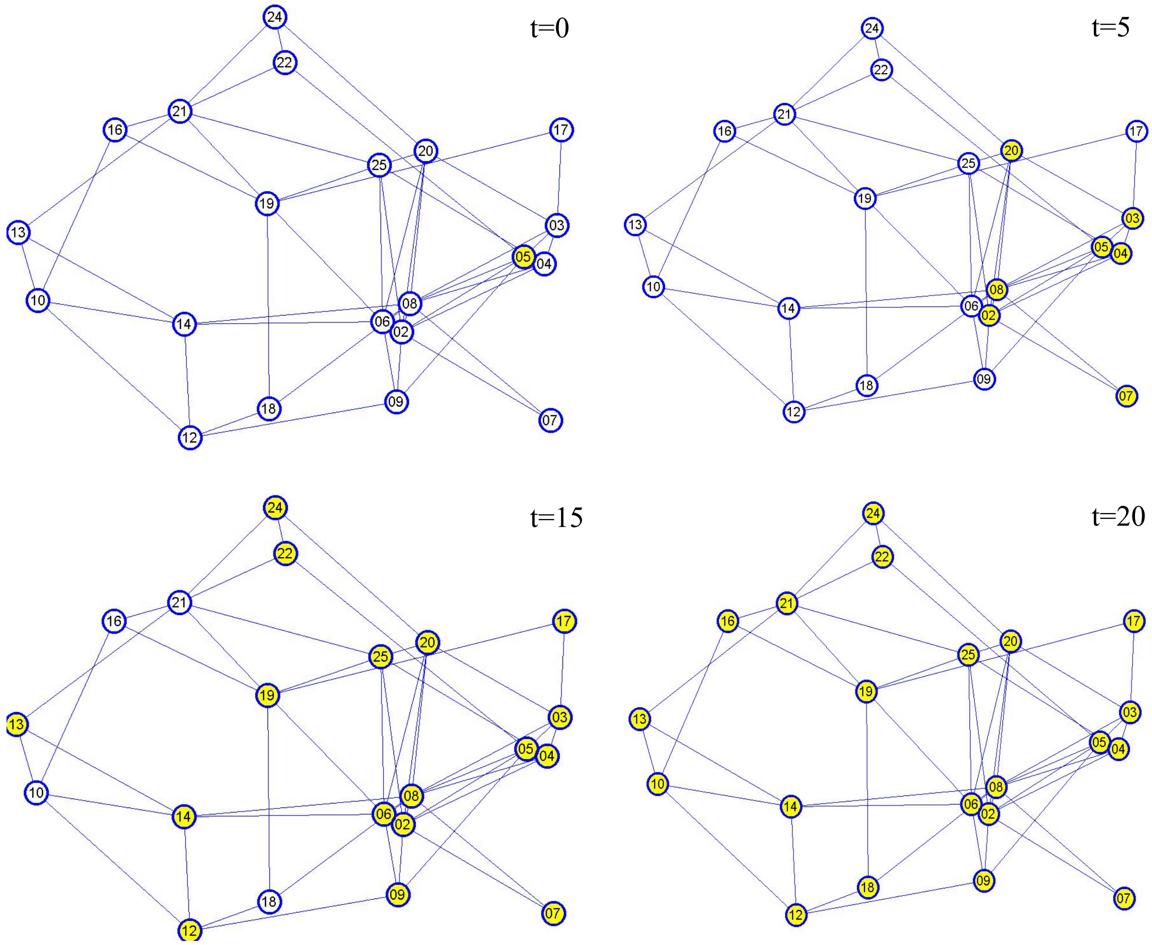

Figure 5-9:Message spreading when initial messenger is ID 5

5.4 Summary

In this chapter we moved from "who" to "how", exploring the dynamics of people's conversations. We used "turn-taking" as a quantitative measure of someone's conversational style. When a person is talking, they may either continue to talk, or pause to listen to their conversation partner. Similarly, when they are listening, they may either remain silent, or may start to talk. Each individual's overall average pattern of conversation is summarized by the probabilities of their making each of these types of transition. Having derived this quantitative measure of "overall conversational style", we were then able to measure, for each individual, the degree to which her conversation partners tend to conform to her overall style. This measure is the “influence value” for that person, and is computed using the Influence Model that was presented in Chapter 3.

We then explored the question of whether these influence values, derived from each individual’s conversational dynamics, correlated with people’s centrality scores, which were derived in Chapter 4 from their connectivity as nodes in the overall social network. We found a strikingly high correlation between these two separate measures, indicating that if a person’s one-on-one conversations tend to go “their way”, then the overall flow of communication in the social network tends to go “their way” as well. An interesting direction for possible future studies will be to explore whether this correlation between pair-wise conversation and network-graph measures will continue to hold in larger sample sizes.

Finally, we carried out simulations to explore possible ways in which information might propagate within the social network structure. We found that the rate at which information flows can vary greatly, depending on the connectedness of the initial messenger.

Chapter 6: Conclusions

"I do not fear computers. I fear the lack of them."
Isaac Asimov

6.1 Contributions

The main contribution of this thesis is to have demonstrated the feasibility of learning social interactions from raw sensory data. In this thesis we have presented a framework for automatic modeling of face-to-face interactions, starting from the data collection methods and working up to computational methods for learning the structure and dynamics of social networks. Specific contributions of this thesis include:

- The first successful experiment in learning the communication patterns by equipping a group of people within a community with wearable sensing devices. Carefully considering functionality, design and the privacy concerns of users, we built the sociometer as a solution to the problem of how to unobtrusively collect natural interaction data. A big hurdle in using perceptual sensing to model human behavior was that of collecting and being able to use real-world data from natural scenarios. We solved the problem of collecting the data by building an appealing and unobtrusive device, and the noisy real-world data was made relatable to social-network level phenomena by a variety of preprocessing steps.
- The thesis work generated a unique dataset of interaction data from twenty three individuals. This dataset recorded six hours of data daily for a period of two-weeks. Such continuous sampling of interaction behavior would not be feasible using standard social network sampling methods (i.e. surveys, human

observers, personal diaries etc). Previous sensor-based recordings of large groups of people have been limited to collecting only proximity information from infra-red and radio-frequency transmitters. This thesis is the first to move beyond proximity, in this case by measuring conversational interactions in large groups.

- Having collected and preprocessed the social interaction data, the thesis then explored methods for computationally modeling the underlying social network structure. We have integrated methods from speech processing, conversational scene analysis and machine learning to demonstrate that it is possible to extract information about people's patterns of communication without imposing any restriction on the user's interactions, or the environment in which the user usually functions.
- We have analyzed social network data at much finer time granularity than had been done before.
- Investigating conversational dynamics: analyzing not only whom we talk to, but also how we talk to them. We have presented results that demonstrates distinctive and consistent turn-taking styles for individuals.
- A new interaction model for quantitatively capturing the effects of one person's turn-taking behavior on another. We capture how much our interactions deviate from our average turn-taking pattern for certain conversation partners. We have presented new results on how some individuals tend to influence people's dynamics more than others.
- A new result that shows strong statistically significant correlation between a person's aggregate influence value and her centrality score. Indicating the possibility that a person's style during one-on-one conversations may be indicative of the person's overall role in the connectivity of the network.

Although during the data collection stage of the project we recorded the audio of the subject, none of the analysis and results on this thesis rely on the recognizing the spoken words that formed the semantic content of the interaction. This thesis shows that it is possible to get rich information about people's interaction with others and how they are embedded in their social network, without being too intrusive of their privacy.

6.2 Future Directions

This thesis is the first step in trying to answer some of the traditional questions in social network analysis and organizational behavior, using new sets of sensing and analysis tools. It attempts to explore new questions that were not addressable using traditional techniques. The resulting work has opened up a number of possible avenues for future experiments and research. In this section we list a few of these directions that we think are the most interesting:

- **Influence and centrality – a sustainable correlation?** We discovered a strong correlation between people's talking style and the role they play in connecting the network. This connection needs to be explored further, with a greater number of people and in different environments. Is this correlation an artifact of the specific community that we were looking at, or is it a more general effect which holds for different communities? Do communities with very different centrality distributions also have very different influence distributions?
- **Comparing face-to-face vs. online behavior** – How similar or dissimilar are people's patterns of online social behavior, compared to their face-to-face interactions? Are the people who are the strong influencers during face-to-face interaction also the ones with strong influence values during email or chat communications, or is this measure dependent on the mode of communication?
- **Additional sensing modalities** – For the analyses done in the thesis, we used only the audio data from people. However, it has been shown that people also affect each other's body language[1, 25]. Other sensory channels, such as accelerometers, EKG, or galvanic skin response, are additional modalities that could be included for richer analysis of people's face-to-face communication behavior.
- **Analyzing the interaction style among three or more people** – In this thesis we looked at turn-taking behavior in conversations that had two participants.

An interesting question to explore would be whether interactions within a triad could be modeled by combining dyad models.

- **Measuring information flow** – An experiment that tests the real path of information flow would be important in understanding the value of the links between people. Can we improve the flow of information by adding new links to the individuals through whom the greatest amount of information flows? Or can we add alternate paths in the network, to reduce the reliance on people who are currently over-used within the network.
- **Evolution of the network over time** – One of the major advantages in real-time sensing and automatic modeling is that we can now potentially analyze the long term trends of the network. A fruitful line of exploration would be to study networks over a long period (months, or even years) and analyze how they evolve over time. Can we detect the significant transition points where a lot of changes are taking place – e.g. during layoffs, or for a person's own network during relocation?
- **Adding value for the users** – At present the people wearing the sensing devices do not obtain any direct benefit from participating in the study. In order for such sensing technology to be widely acceptable they need to provide some additional value to the user, over and above the broader but impersonal value obtained from understanding the network behavior. One incentive may be to provide a detailed dairy to the user of her communication patterns with different people, or even simple measures such as how much people talk over the course of the day. However, more economically viable incentives probably exist, and we are interested in exploring such possibilities.
- **Beyond the Sociometer** – The participants in the current experiment were already far more familiar than average with the concept of wearable computing, and hence much more accepting of it. However, in order for a more diverse community of people to accept wearing sensors for a significant part of the day, we need to improve upon the current form factor of the sociometer. What will the next generation of the sociometers look like? Will they be incorporated into a watch or be part of a PDA or cell phone, or will

they be something completely different? The form factor and wearability issues need to be considered seriously. Even in the unusually accepting community of the MIT Media Lab, the broad user base in the present study was achievable only as a result of extensive efforts to make the sociometer easy to use, and much less cumbersome than is the norm amongst existing wearable devices.

6.3 Concluding Remarks

Insights into how humans communicate and function as a group can enable us to build better and smarter context aware systems, and also tools that support collaboration. By giving computers the ability to move beyond focusing on the actions or behavior of a single individual, and by trying to understand and model the networked community, we can extend the capabilities of current intelligent systems.

Ever since Stanley Milgram's famous "small world" experiment that showed that any individual is connected to any other through a short chain of social ties (consisting of six people, on average)[60], researchers have been trying to understand the mysteries of human networks. Although many have proposed different topologies that explain the short path length property of human networks, few have addressed the question of "why" – are some people's mannerism and style of interacting indicative of having a hub or a connector status within the network. In human communication networks, understanding the micro-level dynamics may hold answers to understanding macro-level behavior. We believe that discovering ties between individuals' behavior patterns and their roles within the network may provide insights into understanding networks, beyond just topology and path length. Social psychologists have tried for a long time to understand how people change their behavior patterns as they interact with others, but have been limited by the problem of how to record and analyze large groups of people. We are now beginning to have the data acquisition devices and the computational tools what will make these problems tractable for the first time, allowing for richer and more scalable explorations than could previously be imagined.

Chapter 7: Bibliography

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, WP Number 165-97.
3. Huberman, B. and Hogg, T., *Communities of Practice: Performance and Evolution.* Computational and Mathematical Organization Theory, 1995. **1**: p. 73-95.
4. Allen, T., Organizational Structure for Product Development, 2000, Sloan School of Management, MIT, WP Number 166-97.
5. Bernard, H.R., Kilworth, P.D., Kronenfeld, D., and Sailer, L., *The problem of informant accuracy: the validity of retrospective data.* Annual Review of Anthropology, 1984. **13**: p. pp. 495-517.
6. Watts, D., *Six Degress: The Science of a Connected Age*. 2003: W. W. Norton & Company.
7. Wasserman, S. and Faust, K., *Social Network Analysis Methods and Applications*. 1994: Cambridge University Press.
8. Valente, T.W., *Social network thresholds in the diffusion of innovations.* Social Networks, 1996. **18**: p. 69-89.
9. Solomon, H., *Mathematical Thinking in the Measurement of Behavior: Small Groups, Utility, Factor Analysis. A Study of the Behavior Models Project*. 1960, Glencoe, IL: Free Press.
10. Granovetter, M., *The strength of weak ties.* American Journal of Sociology, 1973. **78**(6): p. 1360-1380.

11. Allen, T., *Managing the Flow of Technology*. 1977, Cambridge MA: MIT Press.

12. Wasserman, S. and Pattison, P., *Logit models and logistic regressions for social networks: an introduction to Markov Graphs and p*.* Psychometrika, 1996. **60**: p. 401-426.

13. Liljeros, F., Edling, C., Amaral, L., Stanley, H., and Aberg, Y., *The web of human sexual contacts.* Nature, 2001. **411**: p. 907-908.

14. Tyler, J., Wilkinson, D., and Huberman, B. *Email as spectroscopy: Automated discovery of community structure within organizations*. In *International Conference on Communities and Technologies*. 2003. Amsterdam, The Netherlands.

15. Adamic, L.A., Buyukkokten, O., and Adar, E., *A social network caught in the Web.* First Monday, 2003. **8**(6).

16. Erdos, P. and Renyi, A., *On the evolution of random graphs*. Publications of the Mathematical institute of the Hungarian Academy of Sciences. 1960. 17-61.

17. Watts, D. and Strogatz, S.H., *Collective dynamics of 'small world' networks.* Nature, 1998. **393**: p. 440-442.

18. Barabasi, A. and Albert, R., *Emergence of scaling in random networks.* Science, 1999. **286**: p. 509-512.

19. Huberman, B. and Adamic, L.A., *Evolutionary dynamics of the world wide web.* Nature, 1999. **401**: p. 131.

20. Adamic, L.A., Lukose, R., and Huberman, B., *Local search in unstructured networks*, in *Handbook of graphs and netwroks: From the genome to the internet*, S. Bornholdt and H. Schuster, Editors. 2002, Wiley.

21. Kleinberg, J., The small-world phenomenon: An algorithmic perspective, 1999, Cornell Computer Science Technical Report 99-1776.

22. Watts, D., Dodds, P., and Newman, M., *Identity and Search in Social Networks.* Science, 2002. **296**: p. 1302-1305.

23. Hofmann, T. *Learning Probabilistic Models of the Web*. In *Proceedings of the 23rd International Conference on Research and Development in Information Retrieval (ACM SIGIR'00)*. 2000.

24. Kleinberg, J. *Authoritative sources in a hyperlinked environment*. In *ACM-SIAM Symposium on Discrete Algorithms*. 1998.

25. Wells, G., Petty, R., *The effects of overt head movements on persuasion.* Basic and Applied Social Psychology, 1980. **1**(3): p. 219-230.

26. Morris, R.J. and Hogg, D., *Statistical models of object interaction.* International Journal of Computer Vision, 2000. **37**(2): p. 209-215.

27. Pentland, A., *Smart Rooms.* Scientific American, 1996. **274**(4): p. 68-76.

28. Pentland, A. and Choudhury, T., *Face recognition for smart environments.* IEEE Computer, 2000. **33**(2): p. 50-55.

29. Greene, T., *Feds use biometrics against Super Bowl fans*. 2001. http://www.theregister.co.uk/content/archive/16561.html.

30. Bregler, C. *Learning and recognizing human dynamics in video sequences*. In *IEEE Conference on Computer Vision and Pattern Recognition*. 1997.

31. Stauffer, C. and Grimson, W.E.L., *Learning patterns of activity using real-time tracking.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000. **22**(8): p. 747-757.

32. Addlesee, M.D., Curwen, R., Hodges, S., Newman, J., Steggles, P., Ward, A., and Hopper, A., *Implementing a sentient computing system.* IEEE Computer, 2001. **34**(8): p. 50-56.

33. Hofer, B., Blank, R., and Meier, R., *Spotme: An Autonomous Short-Range Communication System*. 2002. http://www.spotme.ch/.

34. Borovoy, R., *Folk Computing: Designing technology to support face-to-face community building*. 2001.

35. 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.

36. Adar, E., Lukose, R., Sengupta, C., Tyler, J., and Good, N., Shock: Aggregating Information while Preserving Privacy, 2002, HP Laboratories: Information Dynamics Lab.

37. Gibson, D., Kleinberg, J., and Raghavan, P. *Inferring Web communities from link topology*. In *Proc. 9th ACM Conference on Hypertext and Hypermedia*. 1998.

38. Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., and Martin, R., *Design for Wearability*. 1998, Institute for Complex Engineered Systems, Carnegie Mellon University. http://www.ices.cmu.edu/design/wearability/files/Wearability.pdf.

39. Gerasimov, V., Selker, T., and Bender, W., *Sensing and Effecting Environment with Extremity Computing Devices.* Motorola Offspring, 2002. **1**(1).

40. DeVaul, R. and Weaver, J., *MIT Wearable Computing Group*. 2002. http://www.media.mit.edu/wearables/.

41. Marti, S., Sawhney, N., Jacknis, M., and Schmandt, C., *Garble Phone: Auditory Lurking*. 2001. http://www.media.mit.edu/speech/projects/garblephone.html.

42. Basu, S., *Conversational Scene Analysis*, in *Electrical Engineering*. Ph.D. 2002, MIT: Cambridge.

43. Asavathiratham, C., *The Influence Model: A Tractable Representation for the Dynamics of Networked Markov Chains*, in *Dept. of EECS*. 2000, MIT: Cambridge. p. 188.

44. Jordan, M. and Bishop, C., *An Introduction to Graphical Models*. In press: MIT Press.

45. Rabiner, L., *A tutorial on Hidden Markov Models and selected applications in speech recognition.* Proceedings of IEEE, 1989. **77**(2): p. 257-286.

46. Ellis, D., *Speech and audio processing and recognition*. 2003, Columbia University. http://www.ctr.columbia.edu/~dpwe/e6820/.

47. Saul, L.K. and Jordan, M., *Mixed memory Markov models.* Machine Learning, 1999. **37**: p. 75-85.

48. Brand, M., Coupled hidden Markov models for modeling interacting processes, 1996, MIT Media Lab Vision & Modeling, TR #405.

49. Blimes, J., A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, 1997, University of California, Berkely, ICSI-TR-97-021.

50. Lauritzen, S.L. and Spiegelhalter, D.J., *Local computations with probabilities on graphical structures and their application to expert systems.* Journal of the Royal Statistical Society, 1988. **B50**: p. 157-224.

51. Bertsekas, D.P., *Nonlinear Programming*. 1995, Belmont: Athena Scientific.

52. Basu, S., *Conversation Scene Analysis*, in *Dept. of Electrical Engineering and Computer science*. Doctoral. 2002, MIT. p. 1-109.

53. Kruskal, J. and Wish, M., *Multidimensional Sclaing*. 1978: Sage.

54. Bertodo, R., *Evolution of an engineering organization.* International Journal of Technology Management, 1990. **3**(6): p. 693-710.

55. Menzel, H., *Review of studies in the flow of information among scientists.* Columbia University Bureau of Applied Social Research, 1960.

56. Bavelas, A., *A mathematical model for group structure.* Human Organizations, 1948. **7**: p. 16-30.

57. Freeman, L.C., *A set of measures of centrality based on betweenness.* Sociometry, 1977. **40**: p. 35-41.

58. Shimbel, A., *Structural parameters of communication networks.* Bulletin of Mathematical Biophysics, 1953. **15**: p. 501-507.

59. Bikhchandani, S., Hirshleifer, D., and Welch, I., *A theory of fads, fashion, custom and cultural change as information cascade.* Journal of Political Economy, 1992. **100**(5): p. 992-1026.

60. Milgram, S., *The small world problem.* Psychology Today, 1967. **2**: p. 60-67.