Social Motion:
Mobile Networking through Sensing Human Behavior

by
Jonathan Peter Gips

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

Master of Science in Media Arts and Sciences at the

MASSCHUSETTS INSITTUTE OF TECHNOLOGY

September 2006

Author:____________________________________
Program in Media Arts and Science
MIT Media Laboratory August 11, 2006

Certified By:____________________________________
Alex (Sandy) Pentland
Toshiba Professor of Media Arts and Sciences
Thesis Supervisor

Accepted By:____________________________________
Andrew B. Lippman
Chair Department Committee on Graduate Students
Program in Media Arts and Sciences

Social Motion:
Mobile Networking through Sensing Human Behavior

By

Jonathan Peter Gips

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

Master of Science in Media Arts and Sciences at the

MASSCHUSETTS INSITTUTE OF TECHNOLOGY

September 2006

Abstract:

Low-level sensors can provide surprisingly high-level information about social interactions. The goal of this thesis is to define the components of a framework for sensing social context with mobile devices. We describe several sensing technologies – including infrared transceivers, radio frequency scanners, and accelerometers – that both capture social signals and meet the design constraints of mobile devices. Through the analysis of several large datasets, we identify features from these sensors that correlate well with the underlying social structure of interacting groups of people. We then detail the work that we have done creating infrastructure that integrates social sensors into social applications that run on mobile devices.

Thesis Supervisor: Alex (Sandy) Pentland
Title: Toshiba Professor of Media Arts and Sciences

Social Motion:
Mobile Networking through Sensing Human Behavior

Jonathan Peter Gips

Thesis Committee:

Advisor:______________________________________

Alex (Sandy) Pentland
Toshiba Professor of Media Arts and Sciences
Thesis Supervisor

Advisor:______________________________________

William J. Mitchell
Professor of Architecture and Media Arts and Sciences
Reader

Reader:______________________________________

Chris Csikszentmihályi
Associate Professor of Media Arts and Sciences
Reader

Acknowledgements

Firstly, I give my thanks to the entire Human Dynamics group, past and present, for working together with me on ambitious projects and making me feel like part of a team. This group has taught me that anything is possible with the right group of people, and I could not have asked for a better home at MIT. I owe a special thanks to my advisor, Sandy Pentland who has granted me tremendous intellectual freedom combined with the guidance to realize my goals. Over the past few years in this group, I have met some extraordinary people and made some close friendships. My thanks go out to Michael Sung for showing me how to balance work and play and for collaborating on both. I also thank Nathan Eagle for showing me how to think big and to seize the day. I thank Rich DeVaul for being both a mentor and a friend.

I thank Bill Mitchell and the Smart Cities group for being an extended MIT family. Through working with this group, I have learned new ways of thinking to which I would not have been otherwise exposed. I am indebted to Philip Liang for being a great friend and coconspirator. I also thank Federico Casalegno for involving me with his projects and showing me how to enjoy research. I am grateful for the support that this group has given to the Elens project, which would not have happened without it. Thanks also to the Elens team for all their hard work and dedication in putting together an exciting new platform.

I am very grateful for the hard work that Joe Paradiso, Mat Laibowitz, Ryan Aylward and the rest of the Responsive Environments group put into the UbER-Badge. I would also like to thank Lisa Lieberson for her invaluable work in making the badge events go off without a hitch.

Thanks to Highlands and Islands Enterprises for their generous support of my research.

Thanks to my readers – Sandy Pentland, Bill Mitchell, and Chris Csikszentmihályi – for their support of my work and their feedback in this process.

I thank my family for their endless support in everything that I do.

Finally, I thank Leah for keeping my life balanced and having the patience of a saint.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Today, over one and a half billion people are using mobile phones as they go about their daily lives (3G.co.uk 2005). These devices are a compelling platform by virtue of the range of uses to which they are put, from basic telephony to music, video, note-takers and navigators. Mobile phone and wireless carrier industries are currently looking at these devices to deliver location-based services (LBS) with the aim of changing how people shop, travel, and retrieve information. However, this use of positioning technologies represents only a one-to-many model where a set of content providers feed location-based information to users. When combined with many-to-many social software the likes of which drive Web 2.0 sites like Amazon, EBay, and Flickr, context-aware services have the potential to enable an exponentially more powerful set of search, chronicling, and collaborative filtering tools.

An intrinsic characteristic of social software is the ability for users to "link" to other users and their content with the end effect of emphasizing material of high quality as determined by their own experiences, opinions, and expertise. For Amazon, this primarily takes the form of user reviews on merchandise. EBay captures transactional evaluations by the buyer about the seller. Sites like Flickr and Del.icio.us allow people to create tags that help relate disparate works. MySpace is perhaps the best example of the power of social computing. With over eighty million users, this social networking site has nearly saturated the youth market in the United States by making the members themselves into the featured content. Social sites make it easy for members communicate a mediated identity to other members. In return, these sites have the potential to benefit hugely from the network effects that their members create.

Conventional web-based techniques for adding connections between members, such as searching for members and browsing buddy lists, face substantial hurdles in mobile computing. Mobile devices present constraints that are not present in desktop environments. User interaction is limited by the demands of the physical world where attention is at a premium in environments that are continuously changing. Additionally, the compact user interfaces on mobile devices further reduce the complexity of tasks that users can accomplish on the go. Interfaces that require extended periods of concentration are best left on the web.

While mobile devices disfavor user intensive operations, they open new avenues into sensing social interactions to augment the mobile user's ability to connect with people. Several projects have recognized this opportunity and implemented systems that use the sensing capabilities of mobile devices to create social metadata about how the phone is used. Proximity scanning of co-present devices, and more importantly their users, is one powerful and popular technique to measure the social context of a mobile device (Davis, King et al. 2004). This is a useful social measurement, but proximity alone does not equate to social interaction.

Proximity is only one of many behavioral signals that can be captured by mobile devices and used to analyze social interactions. Signals such as voice and conversation dynamics have been shown to predict the roles that people play in organizational contexts (Choudhury 2003). Other research has looked at patterns of location over time to infer high-level contexts such as "work" and "home" (Eagle 2005). In addition to proximity and audio, physical activity, as sensed by accelerometers on the body, can also function as such an indicator. By looking at the patterns of motion that people exhibit in groups, one can quantify the strength of the coupling between people in a social encounter.

Measuring face-to-face social interactions with mobile devices offers two opportunities to mobile, social computing. First, we can give mobile devices the ability to intelligently and proactively forge links with new peers as we interact with them. Whether we are attending conferences, concerts, bars, or taking a walk in the park, people go out of their homes to strengthen their interpersonal relationships and forge new social connections.

From exchanging business cards to writing numbers on napkins, cultural rituals abound for establishing the permanence of these connections. These mechanisms typically result in the addition of the strongest of these connections to our networks. Despite our best efforts and intentions, however, many meaningful and valuable interactions still slip through our digital fingers. Mobile devices equipped with social sensing capabilities create the unique opportunity to capture these interactions so that they can be utilized at a later date.

Second, we can begin to use the sensed social interactions to infer social context without depending on the user to continuously update the system. The idea here is that a social context is described by the social dynamics that occur within it. These dynamics include properties such as who tends to interact with whom and in what ways. If we learn these dynamics as sensed by mobile devices, then we should be able to identify the likelihood of a mobile user being in a particular context at a given time. By automatically keeping the social context of the user current, the system can avoid interrupting the user when she is engaged in an unrelated face-to-face interaction. The system can also prioritize social content that it presents to the user by considering its relevance to the user's social surroundings so as to maximize the limited interactions that can be expected from mobile users.

Bringing social software to bear on physical activities immediately raises privacy considerations, which are widely considered to be a major hurdle to ubiquitous computing (Weiser 1999). Economic research has shown that the possession of pertinent information by a subset of parties in a transaction causes a negative externality for the remaining parties. For example, a sales agent who had a real-time map displaying a competitor's position could formulate an optimal strategy that would impose an unaccountable cost on the competitor. Working from this theory, Jiang proposes the *Principal of Minimum Asymmetry*, which is a set of guidelines for assuring privacy in ubiquitous computing. The basic idea is to minimize the asymmetry of information between communicating parties in two ways. The first way is to decrease the flow of private information away from the person who owns it, and the second way is to notify the owner that someone else has accessed the private information so that he can make more informed decisions (Jiang, Hong et al. 2002).

Our proposed method of establishing links between people based on predictability learned from behavior satisfies Jiang's *Principal of Minimum Asymmetry.* Someone whose behavior predicts another person's behavior possesses information that is pertinent to the predicted person. That is, an imbalance already exists between the predictor and the predicted. By sensing this predictability and establishing a pathway for information to flow through, back to the predicted person, i.e. the link, the system can help to reduce the imbalance. For example, the leader of a sales team might predict her team's behavior as they work a trade show. When she actively engages with other attendees, her team may follow her lead and behave in a similar manner. Clearly, the leader's actions are relevant to the team members, and it makes sense that the agents would "link" to the team leader in this context. Perhaps more often, the predictability will be largely mutual, as may be the case with two good friends who are hanging out in a loosely bound group. In this case, the link is equally strong in both directions between friends, and the links to the weakly coupled group members is weaker.

Consider, for example, what could be possible for a group of six tourists visiting a foreign city. In deciding what to explore, each person must balance her own interests with those of the group. Mobile devices can already use interest profiles to make personalized recommendations, but the appropriateness of the suggestions will vary with the group's changing social structure. It may be impossible to convince a large group on the move to take a sightseeing detour, but notifying two like-minded people who are walking together about a Gaudi exhibit around the corner might be a welcomed side trip. A socially aware device could factor the group dynamics into the recommendation process.

Suppose the group went out to a popular bar where diverse sets of people pack into a tight space and mingle with each other. Over the course of the night, the tourists may have interactions with locals and other travelers that are worth remembering. However, exchanging numbers with each of these new acquaintances would be too much effort and possibly seem forced and unnatural. A socially aware mobile device might minimize the cost of exchanging information by logging these salient interactions automatically and making them available for review at a later time.

Imagine that the same group has returned home and shared all their camera phone photos on a website. A simple way to share would be to simply compile all the photos and give each person access to the collection. A more sophisticated way to arrange large collections would be to use metadata produced by the mobile device to arrange the photos. Photos where the viewer was directly interacting with the photographer might be placed most prominently. Photos taken with the viewer in proximity might receive the next highest weighting. In this way, an understanding of face-to-face interactions can inform how information is shared between people.

This thesis describes one attempt to build mobile systems that take into account social relationships of their users and use this information to streamline the users' interactions with the devices. Chapter 2 provides background and related work in several different fields. Chapter 3 presents the specific problems and challenges faced in creating socially aware devices along with a summary of the proposed solutions. Social sensors and their advantages and disadvantages are described in Chapter 4. In Chapter 5, we analyze how different features extracted from social sensors relate to the known social structure of several experimental data sets. Chapter 6 outlines foundational components for building mobile systems that incorporate social awareness. Chapter 7 concludes the thesis with future work and a discussion of the impact of trends in the mobile computing industry.

Chapter 2: Background

We can draw upon several domains in researching how to connect people on mobile devices by sensing their behavior. The Social Sciences are clearly relevant as we are analyzing human behavior in social interactions, albeit through largely quantitative means. Ubiquitous Computing (Ubicomp) is perhaps the field where this work fits most squarely. Ubicomp, an interdisciplinary field itself, incorporates sensor networks, distributed computing, mobile computing, and human-computer interaction. Wearable computing, which is philosophically differentiated from Ubicomp but often deals with similar technologies and practices, offers substantial work in on-body sensing and context awareness.

Social Sciences

The notion of social context employed in most of the aforementioned systems is quite basic: a list or even a count of proximate peers represents the social context of the user. The social sciences, including Sociology and Anthropology, have generated a large body of work that can serve to inform the concept of social context used in ubiquitous computing applications.

Sociometry

Sociometry is "the study and measurement of interpersonal relationships in a group of people." As a field of Sociology, Sociometry distinguishes itself through its emphasis on quantitative analysis that does not try to explain the structure of social encounters as much as it seeks to measure the interactions of subjects and extract patterns from the resulting data (Infield 1943).

The notion of a subject's "choice" of another subject at a particular "moment" has been at the core of Sociometry since its inception. Choice refers to one's inclination towards another person – ranging from negative to neutral to positive, and moment refers to an instantaneous measurement as opposed to a largely retrospective one. The standard way to measure choice is through questionnaires administered to the subjects as close to the moments of their interactions as possible. J.L. Moreno, the father of Sociometry, articulated a fundamental challenge of measuring choice as follows:

The problem is how to motivate men so that they all will give repeatedly and regularly, not only at one time or another, their maximum spontaneous participation. (Moreno 1937)

Moreno and others realized the difficulty in getting the amount of data they required from their subjects. Beyond achieving participation, they were well aware of the deleterious effects that repeatedly asking the question had on the quality of data collected. They also realized that the best measurements were made *in situ* instead of laboratory conditions, where the framing of the study could easily distort the data collected.

Moreno was also aware of the shortcomings of directly asking a subject for her ratings of the peers with whom she interacted. He notes that the subjects may not even be aware of their choices in a given interaction. He writes that "[a] person may not know to whom he is 'drawn'". Going forward, Moreno advised that Sociometry should branch out and invent new mechanisms to measure the many interrelations of society (Moreno 1937).

Proxemics

Proxemics is the "study of the nature, degree, and effect of the spatial separation individuals naturally maintain (as in various social and interpersonal situations) and of how this separation relates to environmental and cultural factors" (Dictionary 2006). Edward T. Hall, who coined the term Proxemics, conducted the founding work in this field and devised a notation system that allows anthropologists to record the "proxemes" of a social interaction much like a linguist would record "phonemes". Hall's Proxemics notation includes eight dimensions – postural, sociofugal-sociopetal orientation (SFP axis), kinesthetic factors, touch code, retinal combinations, thermal code, olfaction code, and voice loudness – that together functioned to determine the social distance between two people (Hall 1963). He identified four such social distances, each with a close and not close modifier: intimate, personal, social-consultive, and public. Through his observations of interactions across multiple cultures, which actually led to his research in this field, Hall concluded that different cultures have different boundaries for each of these social distances. For example, people from the United States tend to have larger distances than those from Arabic cultures, which he concluded leads Arabs to the impression that Americans are disingenuous (Hall 1968).

Recently, researchers have created models of motion in interacting groups to provide simulated data for ad hoc mobile networking. One particular effort has gone so far as to specify the social networks of the modeled agents in order to create more human-like motion (Musolesi, Hailes et al. 2004). By considering the connections between agents, the simulated data can more realistically model the effect that the presence of one agent has on another.

Ubiquitous Computing

Ubicomp and related fields such as Pervasive Computing provide a wealth of research into how computing is becoming integrated into our everyday lives away from the desktop computer. In the late 1980s, researchers at Xerox PARC proposed three new classes of devices – tabs, pads, and boards – that broke dramatically from the desktop metaphor. Of the three, Tabs have had the largest impact on today's trends in mobile computing. Tabs are the predecessors of personal digital assistants (PDAs) and smart phones. They are small devices that are kept on the person and remain continuously powered up. They are used to quickly enter and retrieve digital information via a touch sensitive screen (Weiser 1999).

Fundamental to the use of Tabs was the idea of communication and context. Weiser wrote that three types of context should inform interaction with these ubiquitous devices: location, proximate peers, and other environmental measurements. The idea was that by keeping devices connected to a network and to the context in which they were operating, interfaces could be more natural for the task at hand, and people would have to attend less to the device in order to accomplish what they wanted to do (Want, Schilit et al. 1995).

Around the same time that researchers were working on the first Ubicomp devices at Xerox PARC, the Active Badge system was undergoing trials in London. This system, consisting of infrared emitting badges enabled the intelligent routing of phone calls to the location of the dialed party. These badges were lightweight (about 50 g) and operated for over a week without recharging. The researchers noted that users initially had privacy concerns about being tracked by the system but that these concerns faded after extended use. However, they did not take the potential abuses of the system lightly and concluded that, in the case where location tracking systems are abused by society, "legislation must be drawn up to ensure a location system cannot be misused, while still allowing us to enjoy the benefits it brings." (Want, Hopper et al. 1992)

While the original goal of the Active Badge system – to route phone calls – has largely been made moot by mobile phones, location-tracking infrastructure has been put to new uses. Using inexpensive RFIDs with traditional conference badges, the Experience Ubicomp Project was able to link profiles describing many of the conference participants with their actual locations. When users would approach a tag reader and display, relevant "talking points" would appear on the screen. Other screens displayed "Neighborhood Windows" that gave nearby users a look at the aggregate interests that group members specified in their profiles (McCarthy, Nguyen et al. 2002).

An important aspect of location-tracking systems like the Experience Ubicomp project is that the location information is used largely to determine the social context of the users.

Related systems depend nearly entirely on social context acquired through other means. The Meme Tag is a wearable badge that uses infrared to register other users that come face-to-face with the wearer. The Meme Tag uses this information to match users on the basis of prerecorded questions. When users who were facing each other had similar answers to the questions, green LEDs would flash; if the answers were different then red LEDs would flash (Borovoy, Martin et al. 1998).

Sensing social context does not require fixed infrastructure. Several systems have employed periodic scans with radio transceivers in order to bring mobile social networking out into the world. Many of these systems are intended to support face-to-face collaboration by revealing the user's social context and promoting interaction. The Hummingbird is one such custom mobile RF device developed to alert people when they are in the same location in order to support collaboration and augment forms of traditional office communication mediums such as instant messaging and email. This interpersonal awareness device has been successfully tested at rock festivals and conferences where users found that the devices fostered a sense of connection in an unknown situation (Holmquist, Falk et al. 1999). Social Net is a project using RF-based devices to learn proximity patterns between people. When coupled with explicit information about a social network, the device is able to inform a mutual friend of two proximate people that an introduction may be appropriate (Terry, Mynatt et al. 2002). Jabberwocky is a mobile phone application that performs repeated Bluetooth scans to develop a sense of an urban landscape. It was designed not as an introduction system, but rather to promote a sense of urban community (Paulos and Goodman 2004). Serendipity is a mobile phone application that performs repeated Bluetooth scans in order to introduce people to each other. When a scan shows an unfamiliar person nearby, a query is sent to a central server containing profiles of participating individuals; these profiles are similar to those stored in other social software programs such as Friendster and

Match.com. When a match of interests is found, an introduction messages are sent (Eagle and Pentland 2005).

While social context has functioned as the primary feature for some successful devices, such as the Lovegety (Iwatani 1998), social context shows potentially much larger importance as a supporting piece of information for a variety of applications. Proximity scans have been used to generate metadata for images that enables sharing between dyads of proximal people on a per image basis (Davis, King et al. 2004). In the ContextContacts project, a measure of real-time social context for each user was added to their respective entries in a smart phone contact book. The research showed that social context could aid in several types of spontaneous communication and coordination tasks. While this study showed that user location is more useful than social context, the software showed only a very minimal measure of social context: a number representing the number of other users located around a user.

Wearable Computing

In the mid-1990s, a group of researchers at the MIT Media Laboratory formed the first active wearable computer users group. These "cyborgs" outfitted their bodies with biosensors, computing cores, networking gear, and chording keyboards to push the limits of personal computing (Starner, Mann et al. 1997). While Ubicomp emphasizes embedded computing that disappears into the objects of everyday life, wearable computing places computation directly on the person. This makes wearable computers natural tools for sensing social interactions.

Robert Hooke, discoverer of plant cells, anticipated a central goal of wearable computing back in 1665:

> The next care to be taken, in respect of the Senses, is a supplying of their infirmities with Instruments, and as it were, the adding of artificial Organs to the natural...and as Glasses have highly promoted our seeing, so 'tis not improbable, but that there may

be found many mechanical inventions to improve our other senses of hearing, smelling, tasting, and touching. (Hooke 1961)

Hooke had invented a complex microscope that made his discoveries possible. He understood the opportunity in enhancing other human sensing capacities with technology. Wearable computers promise precisely this; they should augment the abilities of the user to capture, process, and act on signals coming from both the physical and digital world.

Fully functional wearable computers date back to Ed Thorp and Claude Shannon's creation of a body-worn device that helped predict outcomes on a roulette wheel. By pressing a hidden trigger at precise times in the ball's path, the wearer could expect a 44% gain in payoffs by betting on the octet where the ball was predicted to land (Thorp 1998). This early example shows how sensing the world can be translated into probabilistic expectations that the wearer can then consider to her own benefit.

One goal of wearable computing is to minimize the amount of cognitive load imposed on the wearer while using the system. To this end, designers have employed sensors that are capable of automatically reading signals from the environment and the wearer in order to streamline the human-computer interaction of the system. The idea is that if the wearable can be aware of the user's context, the interface can reduce the amount of time the user spends interacting.

Many groups have pursued context-awareness in wearable computers. The Remembrance Agent is the one of the first wearable interfaces to support context-awareness. In this application, the context consists of the text being input into an Emacs text editor, as many of the first wearable computers worked with head mounted displays and text only editing environments (Rhodes 1997).

In order to address the ever-increasing number of context sources, the Context Toolkit aimed to provide reusable components that abstract away the underlying mechanisms that create context in context-aware application building. Written on top of common computing infrastructure like TCP/IP, HTTP, XML, and JAVA, this toolkit provided the basic primitives of context-aware systems: encapsulation of sensors, access through APIs, abstraction, sharing storage, and access control (Salber, Dey et al.).

Using multiple accelerometer data logging devices placed around the body, Bao and Intille were able to achieve approximately 80% accuracy in classifying twenty activities such as walking, running, eating, reading, and vacuuming. They also found that reducing the number of accelerometers to two – one placed on the upper body (wrist, preferably) and one on the lower body (waist or thigh) – still resulted in high recognition rates (Bao and Intille).

Blum adopted a two-accelerometer (wrist and waist) scheme and added in audio processing that was capable of detecting human voiced segments. This system, called LifeWear, is able to determine activity states such as standing, walking, running, bicycling, and typing as well as the conversation state of the user, e.g. "user speaking", "other speaker", and "loud crowd" (Blum 2005).

Choudhury used a wearable device called the Sociometer to measure and analyze voiced conversational dynamics among twenty-three coworkers. This device used infrared to identify face-to-face interactions and recorded audio to internal flash memory. The researchers found that amount of influence over turn taking in a conversation, as modeled by a coupled Markov Model called the Influence Model, correlates highly with centrality in a social network (Choudhury and Pentland 2003).

Eagle instrumented approximately one hundred students and faculty members with smart phones that recorded the behavior of their users. The phones recorded position by cell tower ID, proximate devices with Bluetooth scans, logs from phone calls and text messages, as well as application usage (e.g. alarm clock). By considering patterns of proximity, location, and usage over the several weeks of the study, Eagle could reliably infer the nature of the relationship between study members as "friends" or "not friends" (Eagle and Pentland).

Chapter 3: Social Motion

Problem Statement

How can we make mobile devices aware of the face-to-face interactions that continuously happen all around us? Furthermore, how might we apply these models to one of the most important problems facing social mobile computing: privacy and information sharing? Without answers to these questions, mobile devices will either remain unaware of their social surroundings or become wide open to abuse by detractors such as spammers and nefarious peers.

This type of social intelligence depends on maps of the many and varied human relationships that span our lives. Over the last decade, a significant effort has gone into charting human relationships with online social networking sites. While these sites capture large human networks with some success, they require manual entry, remain largely static in content, and do not have the capacity to capture the subtle differences in human relationships across the many contexts of everyday life.

Outfitting mobile and wearable computers with sensors and models of social interactions opens an avenue into tackling the dynamic side of social awareness that goes beyond the domain of online social networking sites. However, we are immediately faced with several challenges:

Sensor Selection. While one could outfit a handful of wearable computers with a vast number of sensor combinations, not all sensors have the ability to scale to large numbers of devices. This is true for several reasons. First, the need for social acceptance prohibits the use of many sensors that would be considered an invasion of privacy (e.g., high quality microphones and video cameras). Second, personal devices must be designed with both style and usability in mind; sensors that require intrusive placement on the body or present a fashion faux pas will likely fail (Starner 2001).

Model Generality. Mobile devices go nearly everywhere that people go. Useful models of social interaction that enable social intelligence should be applicable across these varied circumstances. Additionally, models should generalize across people as much as possible. Without this property, people would have to train individual models for each social setting to derive any usefulness.

Architecture for Multiple Devices. The combination of social and mobile essentially demands that useful models incorporate features from multiple devices. In order to use these models in a real-time scenario, this means that software infrastructure must be put in place to ensure flexibility of device configurations and to handle limitations in connectivity, bandwidth, and computational capacity.

Operation with Minimal User Input. Compared to the world of online social networking sites, mobile users have more divided attention that is usually dominated by the demands of the physical world. Socially aware mobile devices must be able to operate without continuous input from the user.

Modeling Privacy. The final challenge is a large one: how does one create a functional model of privacy and information sharing based on face-to-face interactions?

Social Motion Framework

Our framework, called Social Motion, adds social awareness to mobile devices. In this thesis, we will outline the components of Social Motion and analyze work we have done towards implementing these pieces. Social Motion addresses the aforementioned challenges in the following ways:

Lightweight, orientation-independent sensing. We select sensors that either already exist in mobile phones or have a straightforward path towards adoption by mobile phone manufacturers. Imposing this constraint ensures that our framework remains viable for widespread adoption. The two categories of sensors that we consider are proximity and motion sensors. By proximity sensors, we are referring to sensors capable of detecting other devices within a fixed proximity of the sensing device. We put both RF scanning and face-to-face IR sensors into this category, but only the former fully satisfies the imposed constraint. For motion sensing, we are using accelerometers, which are both small and useful without a known orientation.

Measuring Social Interaction. With the end goal of creating general models of social interaction, we analyze the structure of data collected from several large data sets of group interactions. We identify correlations in the users' states that can be used to create classifiers of friendship and company affiliation. Working from these correlations, we describe a hierarchical model that captures social context at the individual, dyadic, and network levels. We discuss the potential for using one particular network model, called the Influence Model, which has been shown to be useful for modeling social interactions with wearable devices (Choudhury 2003).

Distributed Modeling. We have developed a software infrastructure called Enchantment that is capable of supporting distributed modeling in a network transparent manner. Enchantment implements two inter-process communication mechanisms that enable this functionality: a whiteboard and a signaling system. The whiteboard allows processes to be run simultaneously both locally and remotely while decoupling consumer from producer. The signaling system allows for the high-bandwidth transfer of stream-oriented data from point-to-point. Enchantment can be used for offline classification by simply playing back previously recorded information through the system (DeVaul, Sung et al. 2003). We have developed a large variety of applications with the Enchantment system and will show that it is a versatile tool. It was developed with context classification in mind, and we have developed a system that showcases this use.

Automatic Classification and Context-aware User Interfaces. We do not rely upon having input from the user available for our modeling. Both proximity and motion sensing operate completely autonomously. This allows for the continuous generation of social context metadata for chronicling and other offline uses. We can use the context likelihoods to augment mobile user interfaces with social context. We have made steps towards realizing such an interface with the Electronic Lens project. For this project, we created a set of social widgets that allow users to create and edit social groups on the mobile device. We will describe ways that these elements can be extended with social context information. These elements can also contribute to the refinement of the underlying models by providing ground truth for the learning process.

Privacy as Inverse Function of Predictability. We propose a simple privacy model that we test with empirical data that we have collected. This model holds that the better entity A's state predicts entity B's state (as determined by the influence value, in our case), the lower the privacy barrier B should assign to A. Therefore, two people who are closely coupled will grant each other low privacy levels while two random, non-interacting people will have high privacy levels. Clearly, there are cases where this model breaks down, and these will be interesting to evaluate.

Chapter 4: Sensor Selection

Our goal in sensor selection is simple: we want to sense proximate and face-to-face interactions using sensors that can be easily deployed into current and near future mobile devices. Mobile devices come equipped with the necessary interfaces and resources to connect a multitude of sensors that could be used to accomplish this goal. However, these sensors each have advantages and disadvantages with no single sensor offering the ideal feature set at all times. Instead, we consider an array of suitable sensors that can be combined into multi-modal models. The sensor characteristics are summarized in Table 1 at the end of the chapter.

Infrared Transceivers

Infrared transceivers (IR) have long been used to recognized face-to-face interactions. With this technology, the transmitter emits an infrared signal that is digitally encoded with an identifier that is associated with the transmitting party. A receiver within range of the transmitter can receive this signal, and the receiving device can know with high reliability that the transmitting device is nearby within a tight cone of space. False positives are nearly impossible (with a well designed encoding), but the highly specific range of the transmission means that some face-to-face encounters will be missed. For example, two people in conversation but facing in the same direction will not register with each other.

Since the IR signal requires line of sight to transmit, typically each person wears both a transmitter and a receiver that face outwards from the chest towards the location of a potential interacting party. This configuration is difficult to integrate into mobile phones as they are worn in the United States. In other parts of the world, where phones are increasingly worn around the neck, this is possible but not necessarily reliable. Work in the Ambient Intelligence at the MIT Media Lab has demonstrated the potential for working infrared technology into Bluetooth wireless headsets (Merrill and Maes). This location could enable even finer grain detection of conversation participants. However, the number of people who wear a Bluetooth headset throughout the day is still low. Another project under development in the Human Dynamics group is creating a badge the size of a Star Trek communicator that would allow for the required placement under everyday use.

For our experiments with IR face-to-face encounter detection, we used the UbER-Badge. The UbER-Badge is an electronic conference badge developed in the Responsive Environments group at the MIT Media Laboratory (Laibowitz and Paradiso 2004). The badge is worn around the neck and directs the IR transceiver outwards at chest height. Approximately every second, the badge transmits an IR packet containing the badge's unique ID. These packets are received at up to six meters away within an unobstructed conical field of view of approximately ninety degrees.

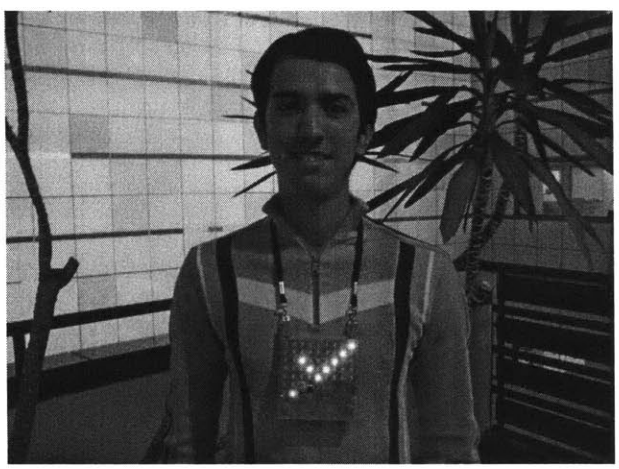


Figure 1. The UbER-Badge records accelerometer and audio signals as well as IR encounters with other badges and fixed beacons called Squirts.

Within the badge hardware, the individual IR packets received are amalgamated to form IR encounters. An encounter begins when an IR packet is received. Each time an additional packet is received within thirty seconds of the previous one, the encounter's end time is extended. After the time limit between packets is exceeded, the encounter is recorded to flash memory.

Proximity Scanners

Research and industry alike have widely employed proximity scanning for sensing social context. Typically, short-range radios will broadcast their IDs to surrounding devices in much the same fashion of an IR based system. However, since radio frequency with omni-directional antennae are used instead of directional infrared, the radio device does not need to be located in any particular orientation or location. Thus, proximity scanners have higher sensitivity but lower specificity than IR based systems at detecting face-to-face interactions.

Many RF scanning systems also have the ability to capture a Received Signal Strength Indicator (RSSI) along with each packet received. This measure can be transformed into a rough approximation of distance. It will not always be accurate due to the way that RF signals propagate in different types of environments (e.g. open air vs. cluttered environments).

Bluetooth is a short-range RF system that works in the unlicensed 2.4GHz band. An increasing number of mobile devices include Bluetooth, which allows people to use wireless headsets, transfer files to desktop machines, and access the Internet through their mobile phones. Several studies have used discovery capability of Bluetooth to track the co-presence of devices over time.

Bluetooth offers several advantages that make it a logical choice for proximity scanning. As Bluetooth comes included in many phones, users do not need to carry around additional hardware. Mobile devices with Bluetooth generally include user interfaces that allow people to turn off the feature when they desire privacy or power-savings. Most importantly, the wide-scale availability and interoperability of devices with Bluetooth makes it an attractive choice.

As it was not designed specifically as a social sensor, Bluetooth has several disadvantages. The Bluetooth software stack is quite large and imposes substantial overhead on proximity scanning what could be a simple piece of functionality in a different implementation. The result is that scans take at least ten seconds to complete and extra energy is expended that can limit the runtime of the entire system. Eagle found that he needed to set Bluetooth scanning to an interval of five minutes in order to achieve an acceptable runtime from a Nokia 6600 (Eagle and Pentland 2006).

Perhaps the largest downside to using Bluetooth as a social sensor is the globally unique BD_ADDR value assigned to every Bluetooth device. This address serves functions at the lowest levels of the Bluetooth protocol to identify a device to other devices. It is set by the device's manufacturer and cannot be changed by the user. While the uniqueness of the BD_ADDR is convenient, it severely limits the ability of an RF scanning system to put privacy mechanisms in place. Without the ability to issue and revoke IDs, it becomes trivial to track a Bluetooth device across time and space.

In our experiments, we have used a Sharp Zaurus SL-5500 PDA equipped with compact flash Bluetooth cards and scanning software. Since we collected data in short-term experiments, we set the scanning rate as high as possible and achieved an average scan period of approximately one minute.

Accelerometry

Accelerometers are not typically thought of as social sensors. Most research using these devices as on-body sensors focuses on individual activity classification. We have taken the novel approach of calculating a feature that captures the mutual information in the motion energy ($MIME$) of two accelerometer signals as an indicator of social interaction. This dyadic feature is symmetric and intended to capture the natural synchronizations that arise from face-to-face interactions. As Erving Goffman put it, a face-to-face interaction is "the reciprocal influence of individuals upon one another's actions when in one another's immediate physical presence" (Goffman 1959). This accelerometer feature is designed to represent these synchronizations.

We calculate the $MIME$ across two people for a given time period (from t_0 to t_1) in the following way:

1. Signals from accelerometers attached to the core (e.g. waist or chest) of two individuals are time synchronized.

Figure 2. Subjects are wearing a hybrid UbER-Badge / Sharp Zaurus PDA system that includes accelerometers on waist and chest, a close-talking microphone, an IR transceiver, and a Bluetooth scanner.

2. The signals are binned into segments that are T_{Bin} seconds long, where T_{Bin} is approximately 2 seconds. This value has been chosen experimentally and reflects the timescale of the interaction components that we are measuring.

3. The energy in each of the bins' signal data is calculated using one of two techniques:

    a. DFFT. We add the square of the discrete fast Fourier transform frequency components without the DC component and divide by the number of samples in the bin. This is the best method but incurs the highest processing cost.

    b. Standard Deviation. We calculate the standard deviation of the samples in the bin. This is an approximation of the energy and suitable for real-time microcontroller implementations.

4. $MIME$ is calculated as the mutual information over the bivariate bin samples between t_0 and t_1.

    a. Mutual Information is determined by the following formula:

    $$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

    where the samples of X and Y have been made discrete by binning their continuous values into a finite number of bins. For our purposes, we divided the continuous energy values into five bins that were evenly distributed over the range of the signal.

Our general interpretation is that the $MIME$ feature is proportional to the amount of synchronization and therefore the amount of interaction in the behavior of the two subjects over the specified time period. For this interpretation to hold, we must assume the nonexistence of foreign influences on the two subjects' motion. For example, two people playing the game "Simon Says", where a third party instructs participants to follow his lead, will have a high *MIME* but are not interacting with each other. More mundanely, two people who are walking up a flight of stairs one after the other will have a high *MIME* but may not be interacting at all. The same would be true for passengers riding on the subway, where the motion of the subway car would dominate and coordinate the motion of the riders.

We can foresee methods to deal with the falsely high readings created by foreign influence. Many of these sources of error, such as the subway car, induce motion that is outside the range of typical human motion. With precise enough synchronization of data, it should be possible to detect this type of motion and remove it from accelerometer signals.

Our use of accelerometers has several advantageous qualities. We only require that a single accelerometer is located near the core of the body so as to measure the motion of the wearer's center of mass. Locating an accelerometer within a mobile phone or device would satisfy this requirement. We do not require a set orientation for the accelerometer as we are calculating only the energy in the signal, which in the case of a three-axis accelerometer is orientation-independent. We do not need to worry about calibration of the accelerometers to a true G scale because we are looking at the correlation across accelerometers, and it is only important that the accelerometers stay consistent over the period of the feature we are calculating.

Accelerometers as social sensors impose some hard constraints. The primary constraint is that the possible pairings of devices needs to be known *a priori*. Calculating the *MIME* feature for every pair of devices known to the system quickly becomes intractable. Instead, this feature relies heavily upon a substantially reduced set of potential interaction pairs as is easily produced by a proximity-scanning sensor. Real-time use of this feature presents an additional constraint where at least one device's energy features must be transmitted to a peer (most likely over RF) in order to calculate the feature.

Alternative Sensors

Mobile devices afford many additional social sensing possibilities through their diversity of input signals. On-device email, phone logs, text messages, and calendars are all rich with indicators of social interaction. While these sources may be used to increase likelihoods of particular social contexts, they do not give any direct evidence of face-to-face interactions occurring.

Audio recorded from a wearable device can be used as a social sensor in many different ways. We have previously described work that extracted structural components studied in Conversational Analysis using wearable microphones to show correlations in these turn-taking patterns to position in a social network. Work like this shows the promise of using wearable sensors to identify the larger structural underpinnings of a social context.

More simplistic audio measurements that do not presuppose the availability of a reliable face-to-face interaction sensor can benefit social interaction analysis. The amplitude of the audio signal from the microphone on the UbER-Badge, for example, can function as a rough indicator of the presence of human conversation. This type of audio feature can potentially be extracted from the microphones built into mobile devices while being stowed on the person.

Several technologies on phones can be used to sense a user's location. Global Position Service (GPS) can yield location accuracies down to a few meters depending on environmental conditions. Cell tower IDs available both on the mobile device and the carrier's servers can be used to sense coarse granularity location that is available both indoors and outside. Another technique is to use Bluetooth scans to recognize devices that have fixed locations (e.g., printers, desktop computers).

The location of a mobile device user can be a powerful social context indicator. Clear mappings exist between physical locations and the types of social contexts that they support. For example, dog parks tend to host pet-related social activities. A soccer field would likely be associated with sports activities. An important social context is the home. Knowing that a device is at home has large implications for the types of social contexts likely to be encountered. Interactions with the same people at work versus at home can relate to two totally different social contexts. There is a substantial learning phase to using location as a social context sensor in this way.

We are interested in inferring face-to-face interaction, and location does have something to offer to this cause. Co-location can function much like proximity sensing in that it is also a necessary condition of face-to-face interaction. Even though location sensing is less robust for this purpose (due to GPS unreliability and cell tower coarseness), it offers something that proximity scanners alone do not offer. This is the asymmetric quality of the data that tells us qualities about the interaction beyond the fact that it occurred. For example, it could tell who confronted whom and in what order the participants departed. This information can be acquired through combining accelerometry with proximity scanning but not with scanning alone.

Time of day, much like location, is useful for determining the types of social interactions that people are likely to occur. During work hours, work social contexts are likely whereas at night and on the weekends personal contexts are more likely. These patterns generalize better than spatial patterns but still require learning on an individual basis.

Sensor	Hardware Cost	Power Consumption	Face-to-face Resolution	Design Constraints	Learning Required
Infrared	Low	Low	High	High	Low
Bluetooth	Medium	Medium	Medium	Low	Low
Accelerometer	Medium	Low	Medium	Low	Low
Microphone	Low	Medium	High	Medium	Low
Location (GPS)	High	High	Low	Medium	High
Location (Cell Tower ID)	Low	Low	Low	Low	High
Time of Day	Low	Low	Low	Low	Medium

Table 1. Analysis of social sensors along multiple dimensions.

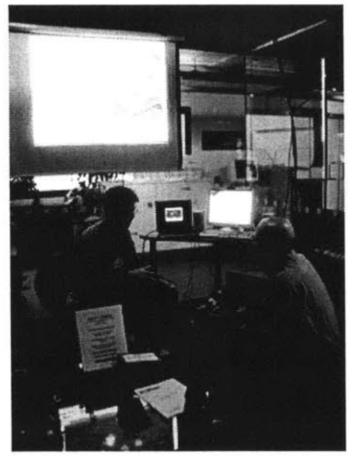
Chapter 5: Measuring Social Interactions

We have instrumented several large gatherings of people with wearable sensors described in the previous chapter. Our primary goal in analyzing this data is to identify the features from the social sensors that correlate the strongest to the known social structure of the subjects within the context of the collected data. Once we have identified these features, we can combine them across all the members of the sensed population and use the resulting adjacency matrix to cluster the individuals.

We collected data from three different social activities. The first data came from the open house at the MIT Media Laboratory where sponsors wore the UbER-Badge for an entire day's worth of presentations, meals, and demonstrations. The second data set comes from a three-day career fair where Scottish middle school students roamed an indoor sports complex looking at booths with their friends. The final data set was collected during a treasure hunt where two teams competed against each other to find clues distributed around a floor of the Media Laboratory. The ground truth social structure came from profiles, surveys, and team assignment, respectively.

Co-workers

In the spring of 2005, eighty-four corporate sponsors of the MIT Media Laboratory wore the UbER-Badge as they attended a semi-annual sponsor day. After attending presentations in an auditorium, attendees had a three-hour period of time to roam around the building, socialize, and attend demonstrations put on by students.

We analyzed this part of the day to see if our social sensor measurements could be used to infer the underlying company affiliations. We found two useful features, which can be used independently or in combination.

Cumulative time spent face-to-face with someone as measured by IR encounters has a medium correlation with whether two people are affiliated or not ($r=0.4674, p<0.001$). We also found that the *MIME* feature had predictive power in determining affiliation between badge wearers. *MIME* showed a medium-low correlation with two people being from the same company ($r=0.35, p<0.001$). Figure 1 shows a multi-dimensional scaling of the *MIME* and IR features of the attendees during the open house.

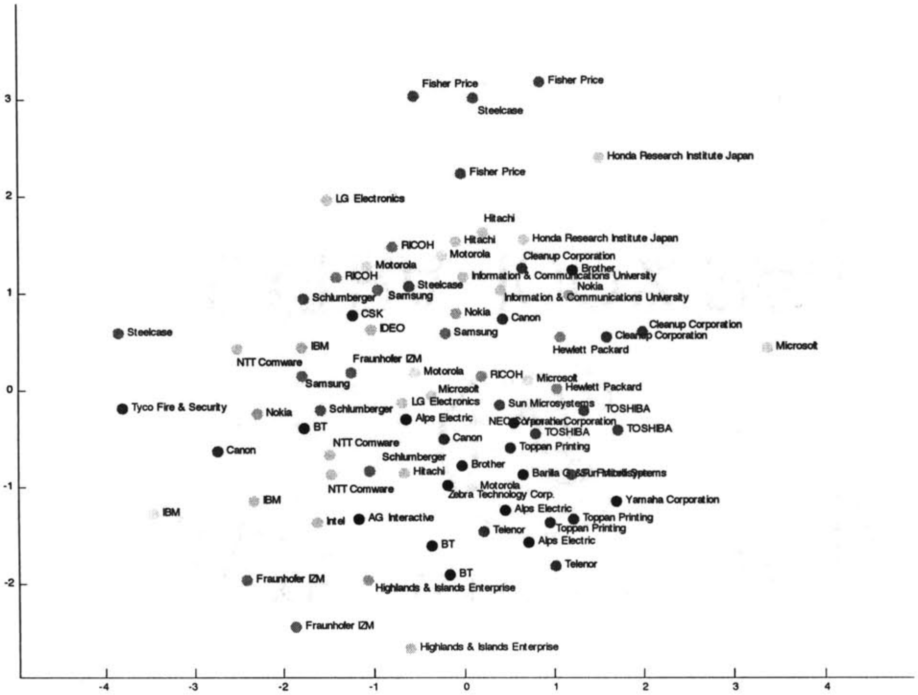


Figure 1 Multi-dimensional scaling of *MIME* and IR encounter time data collected from badge participants during a sponsor meeting at the Media Laboratory.

Friends

We collected a data set from students at a career fair under different conditions. In this data set, 261 students from 27 schools wore badges over the course of three days. The students were divided into groups consisting of between six and ten students drawn from a single school. Each group wore badges for an average duration of approximately one hour. Beacons were placed at exhibits, but we did not place emphasis on giving bookmarks to exhibits and therefore do not rely on these marks in our analysis. Breaking from the sponsor meeting design, students were told to bookmark their friends instead of using the badge as a way to exchange contact information, which would not have made sense in this context.

We analyzed the within-group dyadic interactions for all three days of the career fair. We looked at two features: 1) the natural logarithm of the time two badges spent face-to-face 2) the correlation in the energy of their accelerometers over the time the wearers used the badges. Each of the interacting dyads is plotted in Figure 2. We have removed those dyads that did not spend any time face-to-face from the data set.

From looking at the plot, one can see that the two features have a medium correlation ($r=0.56$, $p<0.001$). As face-to-face time increases so does the motion energy correlation. While unsurprising, this supports the idea that accelerometers can tell us something about face-to-face interactions when IR sensors are absent.

The green marks specify dyads where a bookmark was exchanged. We can see that most dyads of students (84%) who spent a non-zero amount of time face-to-face exchanged a bookmark. We provided an incentive for students to bookmark their friends by raffling off an iPod Nano™ to one person who wore a badge and one of her bookmarked friends. Apparently, the popular strategy was to over-bookmark.

As students returned the badges, we administered a brief survey that asked each person to answer questions about their friendship with another participant. In Figure 2, the circles indicate answers to the question "How well do you know the other person?" The answers were given on a scale of one to seven, with one corresponding to "not at all" and seven to "very well", in this case. The size of the circles is proportional to the answer given. A limitation with the survey data is that it was only administered to dyads – each participant did not fill out a survey about everyone else in her group. Interestingly, however, we noted that small groups of students would return at the same time after exploring the event and fill out surveys about each other. This had the effect of making the presence of a dyadic survey response function as a sort of "buddy" indicator. This observation is supported by the data, which shows that these survey dyads have both higher time spent face-to-face and higher motion energy correlation than the non-survey dyads (see Table 2).

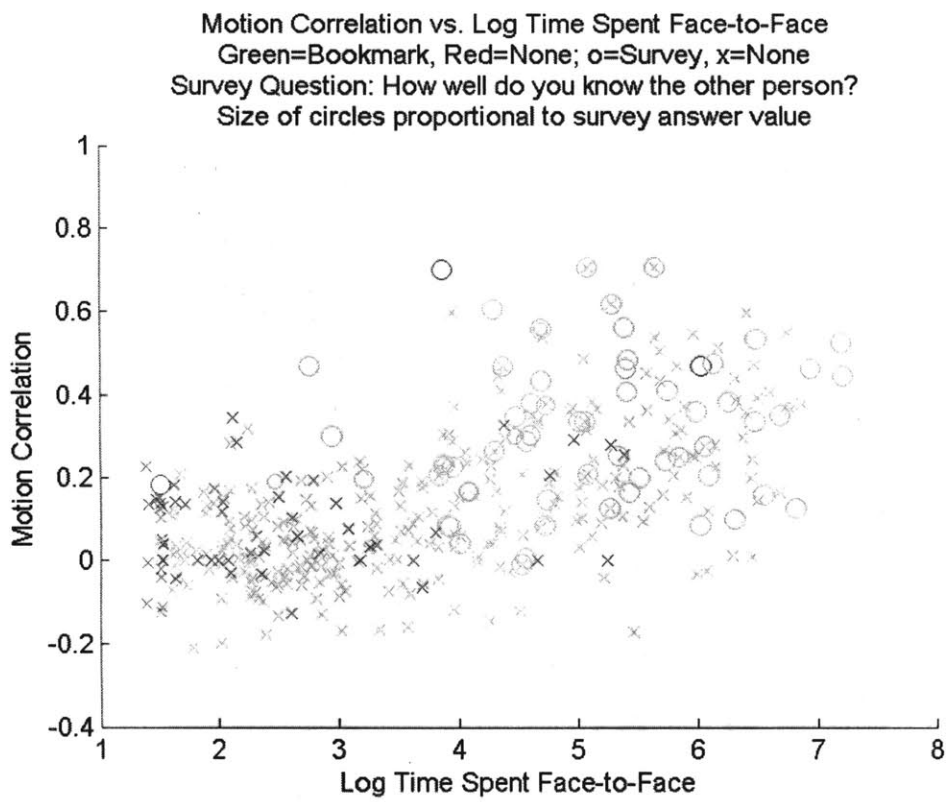


Figure 2. Scatter plot of features calculated for dyads of school classmates.

Dyadic Relationship	Log IR Face-to-Face Time	Motion Energy Correlation
Surveyed	$(\mu = 5.12, \sigma = 1.13)$	$(\mu = .308, \sigma = .166)$
Bookmarked, but not Surveyed	$(\mu = 3.71, \sigma = 1.48)$	$(\mu = .120, \sigma = .170)$
Neither Bookmarked nor Surveyed	$(\mu = 2.56, \sigma = 1.25)$	$(\mu = .096, \sigma = .166)$

Table 2. The means and standard deviations for three different label categories of dyadic relationships in the career fair data set.

Teammates

We ran a treasure hunt experiment where two teams of people each competed to gather clues distributed around the third floor of the Media Laboratory. Our goal was to identify the team membership of the subjects by only using the sensors that they wore as they participated in the experiment. The short nature of the experiment (twenty minutes) combined with the restricted space that subjects could explore (approximately half of the third floor of the Media Laboratory) presented significant challenges to the classification problem.

For each of a total of three trials, we formed two teams of three people each to compete in a treasure hunt. We told the subjects that approximately twenty clues were distributed around the third floor of the Media Lab in areas that were highlighted on a map that we gave them. Their task was to find as many clues as possible over the course of twenty minutes and also get their teammates to see the clues as well. Figure 3 shows an example clue from the treasure hunt. Each clue consisted of a photo of members of the Human Dynamics group holding various gadgets, a number in a red circle, and a Squirt IR beacon.

Figure 3. An example clue from the treasure hunt (left) and two subjects wearing the wearable gear used in the hunt (right).

We motivated the subjects to search out the clues by telling them that they would use the information they found in the second half of the experiment, which would determine the winning team. The second half of the experiment was a market experiment that had the subjects combine their information to get the best approximation of the total distribution of information across all the clues. This analysis was performed as part of a separate research project.

While they participated in the treasure hunt, subjects wore a rig that consisted of two UbER-Badges and a Zaurus. One UbER-Badge was located on the subject's chest, where it is typically worn, and a second UbER-Badge was placed inside of a fanny pack that rested at the base of the back. These provided two measurements of acceleration as well as face-to-face interaction detection. We placed a Zaurus PDA inside of this pack that contained a compact flash Bluetooth card that performed repeated Bluetooth scans at its maximum (but variable) rate of approximately one scan per minute. The Zaurus also recorded full quality11KHz audio from a headset microphone. Figure 3 shows a subject wearing the data collection rig.

We ran three runs of the Treasure Hunt experiment involving a total of eighteen subjects. We found that both the *MIME* feature ($r = 0.533, p < 0.001$) and the total IR encounter time ($r = 0.438, p < .01$) had medium correlations with two people being on the same team. The Bluetooth scanning failed on four of the devices from the third trial. Using the remaining twelve subjects, we did not find any significant correlation between the number of times devices scanned each other and the subjects being on the same team. We need to experiment further to compare Bluetooth with our other measures.

The correlations of the *MIME* and IR features to team membership are encouraging. Figure 4 contains the features for one trial of the experiment in matrix form. Players 1, 2, and 3 are on one team with 4, 5, and 6 on the other. Figure 5 shows a dendrogram formed by clustering the *MIME* and IR features for the same data. The clustering correctly separates out the two teams.

In this example, the IR features most clearly delineate the teams. The *MIME* features provide the second best differentiator. Interesting, the *MIME* and Bluetooth features are structurally similar, and yet they are noticeably different than the IR features. One explanation for this is that most of the time the subjects were following each other around but not necessarily engaging in face-to-face interaction that would register with IR. While only one example, this suggests that the accelerometer feature may be a suitable proxy for proximity scanning in certain instances.

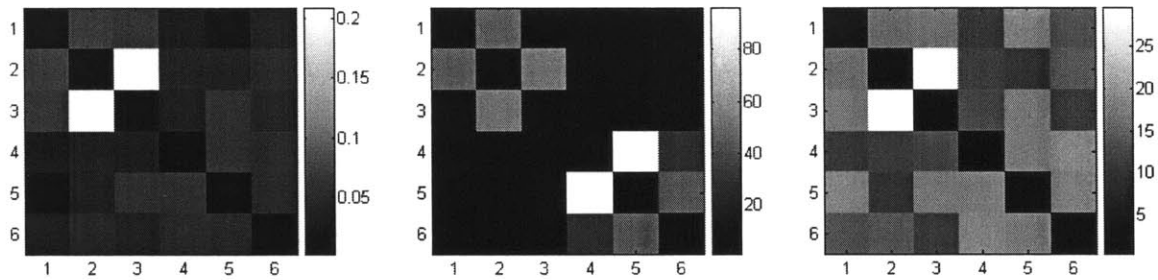


Figure 4. *MIME* (left), IR Encounter Time (center), and Bluetooth Scan Count (right) from a trial of the Treasure Hunt experiment.

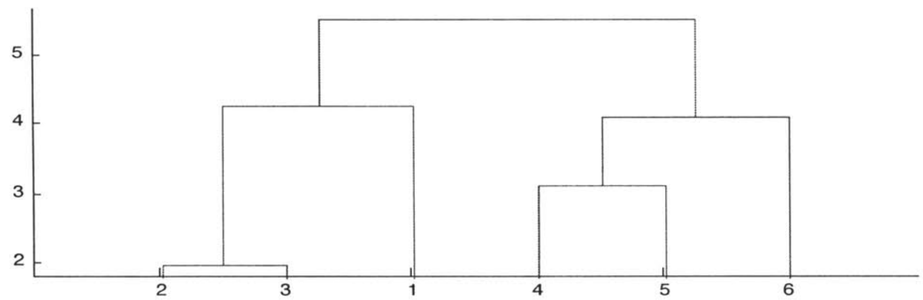


Figure 5. Clustering of the players from a run of the treasure hunt experiment. Team members are correctly clustered into their teams (1,2,3) and (4,5,6) by using the *MIME* and IR features.

Chapter 6: Design of a Socially-Aware System

We have designed a system that uses the ability to sense interaction couplings between people to inform the design of real-time, mobile applications. This system incorporates sensing and feature extraction; networking multiple sensor nodes; group membership management; and social context-aware widgets embedded in a mobile application platform. While we have not yet integrated everything into one fully functional system, the components have functioned individually. The complete system deployment is future work.

Sensing and Feature Extraction

In our architecture, each person modeled corresponds to one sensor node. This node may contain multiple sensors of different types that take measurements such as motion activity, vocalization, and GPS location from the individual. The sensor node contains modules that process these raw measurements into a feature vector. For example, accelerometer data may pass through a Fast Fourier Transform that extracts the strength of the signal across frequency bins. Audio data may pass through both an FFT and a Hidden Markov Model to create features corresponding to voiced or non-voiced audio.

We have implemented a sensor node-based system that handles sensing and feature extraction in real-time on Linux Personal Digital Assistants or other POSIX compliant systems. Software modules that process input signals into transformed output signals form the core of these nodes. We have written modules that read from a variety of sensor hardware including accelerometers, IR transceivers, and biosensors, as used in the LiveNet health monitoring system (Sung and Pentland 2004). After sensor sampling, the outputted signals feed into modules that perform transformations such as down sampling and FFTs. These intermediate signals then flow into modeling modules that classify the state of the sensor node using Gaussian Mixtures Models, Hidden Markov Models, Influence Models, Decision Trees and the like.

In order to network these modules together in as flexible a way as possible, we have designed and implemented an inter-process communication system called Enchantment. The Enchantment library functions as the glue that combines the sensor node modules. It accomplishes this through two mechanisms. The first is a socket-based signaling system that ensures reliable transport of point-to-point time encoded data between processes. The second is a whiteboard server that allows processes to publish and subscribe to signal handles while decoupling the producer from the consumer. This provides the important ability for modules to come up in any order as well as for modules to be "hot swapped" without bringing down the entire system. The whiteboard is also used to publish data that does not continuously vary and has some limited permanence.

In summary, we define a sensor node in our system to be the combination of hardware sensors, software feature extraction modules, the model module used for state inference, and a locally run copy of the Enchantment whiteboard server. Figure 6 contains a diagram of a sensor node.

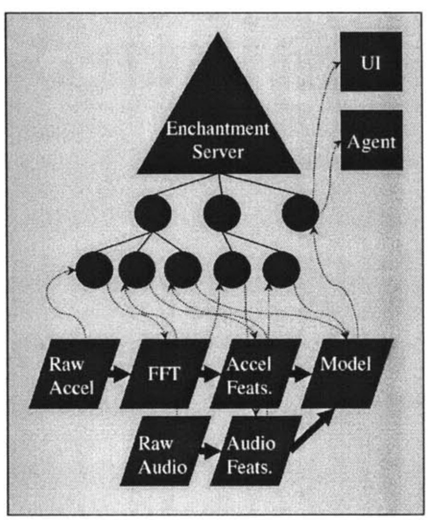


Figure 6. Sensor node with feature extraction and model processes connected using the Enchantment inter-process communication system.

Networking Multiple Sensor Nodes

In a system that models only the activity of a single sensor node, the model component can simply be run on the sensor node itself. However, modeling multiple individuals, as we are doing with our Social Motion architecture, requires data pathways across multiple sensor nodes. This can be accomplished through configuring the inter-process links in a number of ways.

Centralized. In typical client-server fashion, the sensor nodes connect to one special node that runs the model. This has the advantage of minimizing both the overall amount of processing done as well as the theoretical number of data transmissions. A disadvantage to this approach is that there is a single point of failure for the entire classification system.

Distributed. Each node runs the modeling code locally and sends updates to each of the other nodes. While this increases processing and communication, the system is robust to node failure and state inference can be made at each node.

Figure 7 shows examples of centralized and distributed configurations.

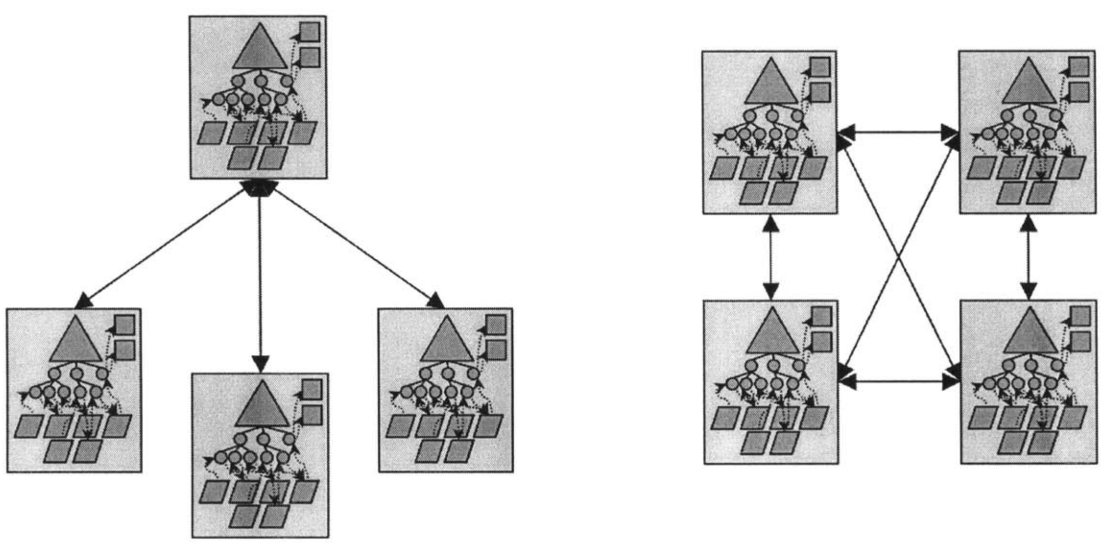


Figure 7. . Sensor nodes connected in centralized (left) and distributed (right) topologies.

Our system allows for the possibility of three types of messages to be passed between nodes to maintain a continuously updated model: raw sensor signals, feature vectors, and latent state likelihoods. The first of these types, raw sensor signals, offers the most flexibility but also has the highest bandwidth requirements. Each of the two remaining types has its advantages and disadvantages.

Feature Vectors. These vectors need only be transmitted at a rate determined by the model, which can be dramatically lower than sampling frequency of the raw signal. In the case of a centralized model, feature vectors are the natural candidates due to the lack of any modeling at the sensor nodes themselves. Distributed models may also communicate feature vectors that will be used as observations in the full model run at each of the nodes.

Latent State Likelihoods. Running the model at each of the sensor nodes, as in the distributed case, enables the transmission of the latent state likelihoods at each time step. Minimally, only the likelihoods for the states pertaining to the sensor node making the calculations need to be transmitted to each of the other nodes.

Figure 8 shows the feature vector and latent state configurations for inter-node messages.

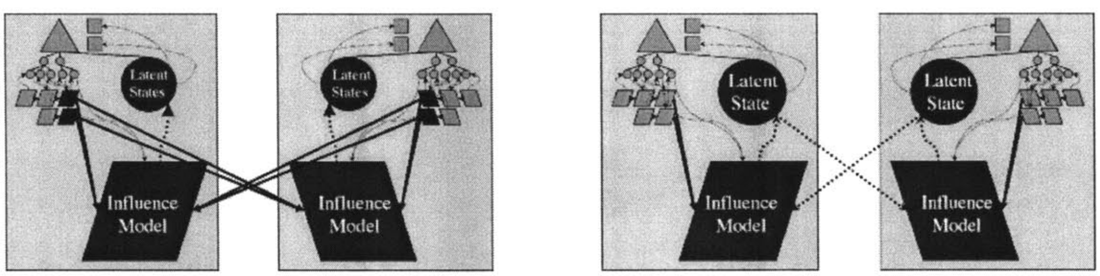


Figure 8. Sensor nodes can transmit feature vectors (left) or latent state inferences (right).

Group Management Software

Enchantment gives us the ability to create real-time models of social interactions, but it does not provide a way to perform the group management operations that a social-context aware application requires. We have created a package of software called Constellation whose role it is to provide this piece of infrastructure.

Constellation comprises a database-backed server that handles connections from mobile clients through a context server called xLink (Sadi 2005). The server handles requests to join and leave "networks", create links between members of networks, and set users' current active status on a per-network basis. The client library includes a set of widgets written in J2ME that access the functionality of the server.

We have integrated Constellation into an urban mobile computing project called the Electronic Lens. The Electronic Lens allows users to participate in asynchronous discussions that are indexed by and accessed through spatial and social keys. Figure 9 shows the main screen of the Electronic Lens running on a Motorola A1000 smart phone. The button in the upper left hand corner of the screen (1) leads to the Constellation control screen where people can edit their network membership. Button (2) enters spatial mode, where users can scan and create visual tag markers that are intended to be affixed to points of interest (Costanza 2006). By pressing the lower button (3), users can select a social space that corresponds to one of the social networks to which they belong. After either scanning a tag or selecting a social space, the user enters an asynchronous discussion space that presents user-generated media clips in a browsable, graph-based interface called RadioActive (Zinman and Donath).

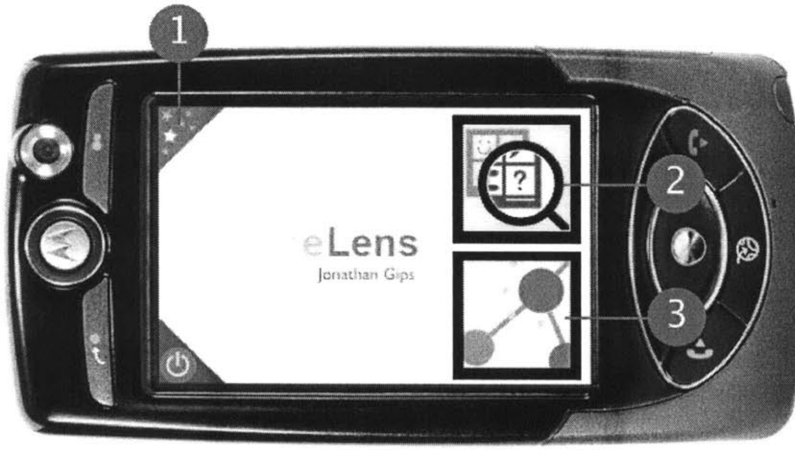


Figure 9. The main screen of the Electronic Lens. Users can edit their social connections through accessing the Constellation control screen (1), or they access media through scanning visual tags (2) or selecting social spaces (3)

The Constellation control screen, which is shown in Figure 10, allows the user to join existing networks or create new ones. The list of networks (1) includes both networks to which the user belongs as well as publicly visible networks that the user might join. When a user is a current member of a network, its entry is bolded, and the right side of the list entry (2) displays the total number of members. Public networks that the user does not belong to are shown in light grey. The text on the right is "Locked" if the network requires a password to join. Otherwise, it displays "Non-member" and can be joined by pressing the button (4) that displays a "+" for join. When a joined network is highlighted this button displays a "-" to un-join the network. Highlighting a locked network will change the icon to a key, which can be pressed to unlock the network with a password. Not all of the networks are initially available in the list. When creating a network with the new button (3), the user can specify the network as private. This makes the network visible once the user presses the private button (4) and enters the network's name. The left arrow (6) brings the user back to the main screen.

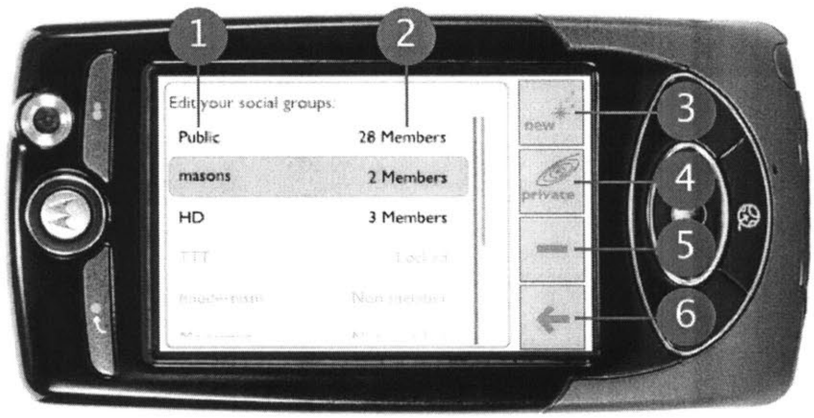


Figure 10. The Constellation control screen. The list shows the available networks (1) and the network's status (2). Buttons on the right allow users to create new networks (3); access private networks (4); join, un-join, and un-lock networks (5); and (6) return to the main screen.

After pressing the social mode button on the main screen, the user enters the network selection screen in Figure 11. This screen allows the user to check the message count of each of her member networks (1). By selecting a network and pressing the right arrow (2), the user continues on to the asynchronous message screen.

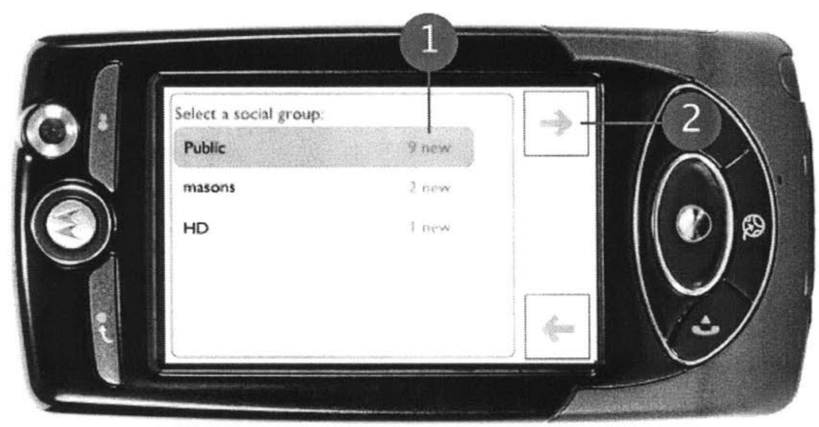


Figure 11. The network selection screen. The user can see the new message count for each of her member networks (1) and enter the selected network's discussion space (2).

Chapter 7: Conclusion

In this thesis, we have presented the foundational components of a mobile system that enables social context awareness and networking through the sensing of human behavior. We showed that several sensing options exist that can be integrated into current mobile devices without disrupting the usage patterns and form factors to which people have grown accustomed. Through our experimental analysis, we have demonstrated that these sensor signals contain the information necessary to infer the underlying social structure of groups of interacting people about which no information is known *a priori*. We have developed infrastructure to make these inferences of social structure available to the mobile user on a real-time basis. With the Electronic Lens project, we have created and tested a platform that integrates social computing into a multi-faceted mobile experience.

Future Work

Going forward, we aim to integrate these components into a new type of sharing application that leverages real-time social awareness to guide user-generated media dissemination. Working from the Electronic Lens platform, we would augment the selection screen to prioritize networks based on the inferred social setting of the user. At the very least, this would cut down the complexity of navigating through all the groups that the user belongs to in order to share a photo. Ideally, the inference would be reliable enough that people would be able to trust the system to distribute the content through the system automatically. We would also implement an online training mechanism where each selection of a social network would provide the system with a trained example of what social context the user considered her self to be in at that time.

Interruption management and synchronous communication support are two additional areas where we can put social awareness to good use. Understanding the relationship of a caller to the receiver's social context could help make intrusive incoming calls more acceptable by only letting pertinent calls through. Conversely, the user's device could proactively contact users related to the user's social context but not currently involved in the interaction. This type of opportunistic communication could enable interactions that would not occur without this capability.

Lessons Learned

Testing and deploying social applications on mobile devices present a substantial challenge. In the United States, the mobile landscape is filled with many players who have failed to converge on a standardized platform that allows full-featured applications to run on a majority of devices. Unifying technologies like J2ME and Symbian have advanced towards this goal, but they have largely fallen short. Upcoming platforms such as Flash and optimized web browsers, such as those from Opera, are striking out on a similarly difficult path. Instead of relying on open standards, we see tightly integrated, closed applications coming from major online players like Google and Yahoo who have used their deep pockets to create their own platforms that span the numerous devices that are on the market.

Even these efforts cannot tap into the true potential of the mobile device as a multi-modal sensor node due to the restrictions that wireless carriers impose on device manufacturers. The carriers pressure these manufactures to lock down their devices with the idea that they alone will reap the benefits of delivering expensive content to their customers. Without a standardized hardware profile, mobile applications are undermined to the point where they become severely limited desktop alternatives.

Consequently, we have seen several new players enter the mobile market that are taking on the roles of device manufacturer, application developers, and wireless carriers. These companies, including the likes of Amp'd Mobile and Helio, aim to unlock the intrinsic value of mobile devices through offering fully integrated hardware, software, and wireless services to their customers. While this approach can offer an improved mobile experiences that takes advantage of what a single mobile device has to offer – such as high quality video feedback, seamless interface design, and location-based services, it creates new barriers along branded device lines for social applications that need the participation of large numbers of people to realize their potential.

Two pathways show particular promise for bringing mobile networking to the masses. The first is Microsoft's increasing role in the mobile device market. Windows Mobile is currently at version 5.0, and the number of devices that support the operating system is growing. Microsoft may well be posed to bring the same type of standardization to the mobile market that they brought to desktop personal computers. With so many devices on the market, it may take a force like Microsoft to achieve such a goal.

The second avenue is more of a parallel pathway than a market trend. Wearable devices that are designed specifically with social applications in mind have the potential to fill the void in today's mobile device market. A plethora of low-power, low-cost radio transceivers, microprocessors, and storage solutions are hitting the market and spurring on the sensor network community. For real-time applications, these devices could be interfaced to mobile devices through Bluetooth or simply work in a stand-alone mode. Ubiquitous Wi-Fi networks and USB accessible desktop computers offer zero-cost ways to upload information to Internet based services for asynchronous applications.

Final Thoughts

Social computing on mobile devices presents a way for people to put their social capital work across all segments of life. By using sensors to automatically infer social context, we can remove the hurdle of manual annotation for disseminating new content and minimize the attention necessary to search for content from other people. Together with advances in location sensing, social context-awareness promises to bring computing out into the every day world where the majority of the world's population will be using computing devices in the years ahead.

References

3G.co.uk. (2005). "How Operators Can Get From 1.5 To 3 Billion Mobile Phone Users." Retrieved March 15, 2006, from http://www.3g.co.uk/PR/March2005/1102.htm.

Bao, L. and S. S. Intille "Activity recognition from user-annotated acceleration data."

Blum, M. (2005). Real-time Context Recognition. MIT Media Lab, ETH Zurich. Cambridge, MA USA. **Master of Science:** 98.

Borovoy, R., F. Martin, et al. (1998). "Meme tags and community mirrors: moving from conferences to collaboration." Proceedings of the 1998 ACM conference on Computer supported cooperative work: 159-168.

Choudhury, T. and A. Pentland (2003). "Sensing and modeling human networks using the sociometer." Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on: 216-222.

Choudhury, T. K. (2003). Sensing and Modeling Human Networks, Massachusetts Institute of Technology.

Costanza, E., Leinss, Mirja. (2006). Telling a Story on a Tag: the Importance of Markers' Visual Design for Real World Applications. Mobile Interaction with the Real World (MIRW 2006). Espoo, Finland.

Davis, M., S. King, et al. (2004). "From context to content: leveraging context to infer media metadata." Proceedings of the 12th annual ACM international conference on Multimedia: 188-195.

DeVaul, R., M. Sung, et al. (2003). "MIThril 2003: applications and architecture." Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on: 4-11.

Dictionary, M.-W. O. (2006). "proxemics".

Eagle, N. and A. Pentland "Reality Mining: Sensing Complex Social Systems."

Eagle, N. and A. Pentland (2005). "Social serendipity: mobilizing social software." Pervasive Computing, IEEE **4**(2): 28-34.

Eagle, N. and A. Pentland (2006). "Reality mining: sensing complex social systems." Personal and Ubiquitous Computing **10**(4): 255-268.

Eagle, N. N. (2005). Machine Perception and Learning of Complex Social Systems, Massachusetts Institute of Technology.

Goffman, E. (1959). "The Presentation of Self in Everyday L fe." Garden City, NY: DoubledayAnchor.

Hall, E. T. (1963). "A System for the Notation of Proxemic Behavior." American Anthropologist **65**(5): 1003-1026.

Hall, E. T. (1968). "Proxemics." Current Anthropology **9**(2/3): 83-108.

Holmquist, L. E., J. Falk, et al. (1999). "Supporting Group Collaboration with Interpersonal Awareness Devices." Personal Technologies **3**: 13-21.

Hooke, R. (1961). Micrographia; or, Some physiological descriptions of minute bodies made by magnifying glasses, with observations and inquiries thereupon, Dover Publications.

Infield, H. F. (1943). "Sociometry and the Concept of the Moment." Sociometry **6**(3): 243-244.

Iwatani, Y. (1998). "Love: Japanese Style." Wired News **11**.

Jiang, X., J. I. Hong, et al. (2002). "Approximate Information Flows: Socially-based Modeling of Privacy in Ubiquitous Computing." Proceedings of Ubicomp: 176-193.

Laibowitz, M. and J. A. Paradiso (2004). "The UbER-Badge, A Versatile Platform at the Juncture Between Wearable and Social Computing." Advances in Pervasive Computing, Oesterreichische Computer Gesellschaft: 363-368.

McCarthy, J. F., D. H. Nguyen, et al. (2002). "Proactive displays & the experience UbiComp project." ACM SIGGROUP Bulletin **23**(3): 38-41.

Merrill, D. and P. Maes "Invisible Media: Attention-sensitive informational augmentation for physical objects."

Moreno, J. L. (1937). "Sociometry in Relation to Other Social Sciences." Sociometry **1**(1/2): 206-219.

Musolesi, M., S. Hailes, et al. (2004). "An ad hoc mobility model founded on social network theory." Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems: 20-24.

Paulos, E. and E. Goodman (2004). "The familiar stranger: anxiety, comfort, and play in public places." Proceedings of the 2004 conference on Human factors in computing systems: 223-230.

Rhodes, B. J. (1997). "The Wearabte Remembrance Agent: A System for Augmented Hemory." Personal Technologies **1**: 218-224.

Sadi, S., Maes, P. (2005). xLink: Context Management Solution for Commodity Ubiquitous Computing Environments. UBICOMP Workshop ubiPCMM: Personalized Context Modeling and Management for Ubicomp App. Tokyo, Japan.

Salber, D., A. K. Dey, et al. "The Context Toolkit: Aiding the Development of Context-Enabled Applications."

Starner, T. (2001). "The challenges of wearable computing: Part 2." Micro, IEEE **21**(4): 54-67.

Starner, T., S. Mann, et al. (1997). "Augmented reality through wearable computing." Presence: Teleoperators and Virtual Environments **6**(4): 386-398.

Sung, M. and A. Pentland (2004). "LiveNet: Health and Lifestyle Networking through Distributed Mobile Devices." Proceedings WAMES.

Terry, M., E. D. Mynatt, et al. (2002). "Social net: using patterns of physical proximity over time to infer shared interests." Conference on Human Factors in Computing Systems: 816-817.

Thorp, E. O. (1998). "The Invention of the First Wearable Computer." Proceedings 2nd. Int'l Symp. on Wearable Computers: 4–8.

Want, R., A. Hopper, et al. (1992). "The active badge location system." ACM Transactions on Information Systems (TOIS) **10**(1): 91-102.

Want, R., B. N. Schilit, et al. (1995). "An overview of the PARCTAB ubiquitous computing experiment." IEEE Personal Communications **2**(6): 28-43.

Weiser, M. (1999). "The computer for the 21 stcentury." ACM SIGMOBILE Mobile Computing and Communications Review **3**(3): 3-11.

Zinman, A. and J. Donath "RadioActive: enabling mobile-based audio forums."