Providing Lightweight Telepresence in Mobile Communication to Enhance Collaborative Living

by

Natalia Marmasse

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

DOCTOR OF PHILOSOPHY IN MEDIA ARTS AND SCIENCES

at the

MASSACHUSETTS INSTITUTE of TECHNOLOGY

September 2004

Author: ______________________________

Program in Media Arts and Science
August 23, 2004

Certified by: ______________________________

Christopher M. Schmandt
Principal Research Scientist, MIT Media Laboratory
Thesis Advisor

Accepted by: ______________________________

Andrew B. Lippman
Chair, Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Providing Lightweight Telepresence in Mobile Communication to Enhance Collaborative Living

by

Natalia Marmasse

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning in August 2004
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MEDIA ARTS AND SCIENCES

Abstract

Two decades of computer-supported cooperative work (CSCW) research has addressed how people work in groups and the role technology plays in the workplace. This body of work has resulted in a myriad of deployed technologies with underlying theories and evaluations. It is our hypothesis that similar technologies, and lessons learned from this domain, can also be employed *outside* the workplace to help people get on with life. The group in this environment is a special set of people with whom we have day-to-day relationships, people who are willing to share intimate personal information. Therefore we call this *computer-supported collaborative living*.

This thesis describes a personal communicator in the form of a watch, intended to provide a link between family members or intimate friends, providing social awareness and helping them infer what is happening in another space and the remote person's availability for communication. The watch enables the wearers to be always connected via awareness cues, text and voice instant message, or synchronous voice connectivity. Sensors worn with the watch track location (via GPS), acceleration, and speech activity; these are classified and conveyed to the other party, where they appear in iconic form on the watch face, providing a lightweight form of telepresence. When a remote person with whom this information is shared examines it, their face appears on the watch of the person being checked on. A number of design criteria defined for collaborative living systems are illustrated through this device.

Thesis Advisor: Christopher M. Schmandt
Title: Principal Research Scientist, MIT Media Laboratory

Providing Lightweight Telepresence in Mobile Communication to Enhance Collaborative Living

by

Natalia Marmasse

Thesis Committee

Thesis Advisor:

Christopher M. Schmandt
Principal Research Scientist
MIT Media Laboratory

Thesis Reader:

Deb K. Roy, Ph.D.
Assistant Professor
MIT Media Laboratory

Thesis Reader:

Joseph L. Dvorak, Ph.D.
Distinguished Member of Technical Staff
Motorola

Dedication

To Isabelle and Rapi for your unconditional love and support, for the moon in the right dose, proteins, for not choosing the jungle option, and for much more than any words could express.

You have been my fortitude in the last few years.

Acknowledgements

First I thank my advisor Chris Schmandt who provided me with the opportunity and the environment in which to carry out my research. I also wish to thank my committee members: Deb Roy, with whom I have taken several classes and especially benefited from his Pattern Recognition class, and Joe Dvorak, without whom this project would not have been possible. Joe was enthusiastic from the moment I scratched out some drawings on my whiteboard and has provided hardware and support for the project ever since. I am also grateful to Motorola for their help and support, and for awarding me a fellowship.

MIT and the Media Lab are a unique environment in which to learn. In my years here I have met many people. Some have inspired me, others have challenged me, some have helped me along the way, and others have made the journey more pleasant and enjoyable. I have been inspired by professors such as Marvin Minsky, Patrick Winston and Whitman Richards. I feel indebted to Michael Hawley for enabling me to participate in the Everest 1998 project; the weeks spent in Nepal and on Mt. Everest were an unforgettable experience, and since then GPS and other sensors have found their way into almost all my projects. Many thanks also to Linda Peterson for all her support, and to my peers and colleagues for making the lab the uncommon place that it is.

I would like to acknowledge all members, and honorary members, of the Speech Group, in particular: Stefan, who started this long journey at the same time; Sean, my officemate and friend of many years; both Kwan and Jang, whom I would have liked to know better; Vidya and Nigel, for great conversations that hopefully will continue in the years to come; Assaf and Joseph, newer officemates but together we made a good team, and last but not least, Sunil, an honorary member of the group and friend.

Special thanks to: Emmanuel Munguia and Gerardo Vallejo for long nights of hardware hacking and deep conversations; David Spectre without whom the watch would have looked very different and for the numerous 3 am breaks to watch the Simpsons; Assaf Feldman for help with the speech detector and his excellent cooking; Joseph Corral who worked as an undergrad on the project, and Anna Pandolfo for help with the questionnaires.

I am very fortunate to have such good and solid friends, on different continents. I am grateful to Tanzeem and Raj for more than years of delicious food and a warm home, and to numerous friends in Israel, both on and off the kibbutz, who after so many years still have the feeling that they do not see me much just because I am having another long day at the office but could walk in at any moment. And finally, I am especially thankful to Isabelle and Rapi, Lucienne and Sead.

Table of Contents

Table of Figures

1. Introduction

1.1 Collaborative living

Over the last twenty years, much research has been done to understand how people work in groups and explore the role of technology in the workplace. This is the domain of Computer Supported Cooperative Work (CSCW), a term coined in 1984 [Greif88]. “Cooperative work” was defined by Marx (1867) as “multiple individuals working together in a planned way in the same production process or in different but connected production processes”. CSCW is a broad domain that focuses on work, that is, the tasks people perform and the technologies that support their collaboration; it is often also referred to as collaborative, collective, or group work.

CSCW systems have been divided into four categories based on the function they play in improving the group’s work [McGrath94]. These categories are: systems for intra-group communication, systems for communication with information, systems for external communication (often the same systems used for internal group communication), and systems that structure group task performance. The first category, systems for intra-group communication, include Group Communication Support Systems (GCSS) such as email or teleconferencing, or any system that reduces communication barriers in the group. Systems for communication with information are those that connect the individuals to shared databases, and multi-editor tools. Lastly, the group task performance category includes Group Decision Support Systems (GDSS), and agenda setting systems, as well as workflow and notification systems such as used in air traffic control. On a larger scale, this category also includes organizational memory, the processes through which the knowledge of an organization is stored and retrieved. On the other hand, a different classification based on the task the group is performing [Rodden99] clusters: message systems, conferencing systems, coordination systems, and co-authoring tools.

Although these two classification schemes focus heavily on the task-solving aspects of teamwork, the importance of informal workplace communication for social bonding and social learning has also been recognized [Nardi00]. The emerging themes associated with communication management tools are awareness and availability, interruption filtering, and multi-modal communication. Interruptions certainly have an impact on performing multiple tasks, but as Hudson puts it "attitudes toward interruption are marked by a complex tension between wanting to avoid interruption and appreciating its usefulness" [HudsonJ02]. An active area of research, within the CSCW domain, has been systems that provide background awareness. This is grounded on evidence that awareness plays an important role in keeping co-located groups coordinated and motivated [Isaacs02]. Moreover, it has been found that teams that cannot have opportunistic and spontaneous interactions perform less well [Kraut90].

CSCW has a rich history of deployed technologies with underlying theories and evaluations. Our thesis is that similar technologies, and the lessons learned from this domain, can also be employed *outside* the workplace to help people get on with life. We call this *computer-supported collaborative living*.

In collaborative, or collective, living, our focus is on people with whom we have day-to-day living social relationships, individuals who are willing to share intimate personal information, people with whom we share our lives, especially those with whom we share a household. We refer to these people from our "inner circle" as *insiders*.

A fundamental aspect of maintaining a functional and socially healthy household is communication; that is, communication to coordinate daily activities as well as to build and sustain the emotional relationships between the household members. The former is more task-oriented and typically involves the direct transfer of information, whereas the purpose of the latter is relationship building and is less focused on the actual content. This communication is often one and the same, however, for simplicity of explanation, they will be referred to as separate forms.

Modern western society is fast-paced, often complex, and very mobile. Members of families are frequently geographically distributed, making it challenging to share time and space. Even people who share the same home often have different schedules and time limitations and thus have little time for face-to-face interaction. Not only do adults have complex schedules, often juggling work and home duties, but so do children with their numerous after-school activities. Therefore, much of the communication between a couple is about the coordination of everyday life, and it is amplified in families with children. Modern parents often spend time synchronizing their children’s activities, either by remote-parenting and/or chauffeuring. This type of communication serves a very practical logistic function. Additionally, security of the children is a big concern and parents wish, or need, to know their children’s location and that they are safe and sound. Besides the logistic importance, there is a feeling of comfort attained in knowing the whereabouts of the people we are emotionally close to.

While task-oriented domestic communication is for functional (household) maintenance, a large part of the family communication is connectedness-oriented, i.e. to maintain and strengthen the existing relationships. The content of the conversation itself is often much less important to the family members, rather the point is spending time together. This type of communication has a social and emotional function.

Needless to say, telecommunication is already used in families to fulfill their need of staying in touch for both practical and emotional reasons, however current telecommunication is limited. Moreover, in technology-mediated systems to enhance collaborative living we are not proposing to replace face-to-face communication, rather we see the need for tools that better support the communication requirements of mobile household-members. As CSCW research has shown, knowledge of presence, availability, and location (PAL) can improve the collaboration of work teams. This thesis is an attempt to use these same types of technology to enhance the coordination, communication and collaboration of this special team: the family unit.

1.2 System overview

Today’s telecommunication has several limitations. The telephone enables communication at a distance, in shared time, but lacks the richness we have when co-located. Up to now the focus has been on verbal communication, restricting the non-verbal expression. It facilitates explicit communication, such as a phone call or a text message, but not the implicit message, for example, of a glance or a smile. It lacks back-channels which help us maintain awareness of those with whom we periodically share communication space, and moreover affords no way of inferring a person’s situation before the communication has been initiated. Although family members often call each other during the day, they often interrupt at inconvenient times, or have to negotiate a phone call by way of a text message.

In technology-mediated communications there is an asymmetry between the initiator of the interaction and the recipient: it is the initiator who dictates the time and medium for the interaction. At best the recipient knows who is trying to contact him, for example from caller-id or the person’s buddy-name, however, the subject/motive is previously unknown. A recipient can screen calls, filter email, or simply ignore the voice mail, email or instant message (IM). Nevertheless, often the communication is desired, though perhaps not at that precise moment or via that particular medium. Our availability, or interruptibility, may vary according to our immediate context: the content of the potential interaction, identity of the initiator, what is on our mind, etc. Many phone calls start off with “Do you have a moment?” or “Is now a good time?”. Absurdly, the recipient is possibly interrupted in order to know whether or not he is interruptible/available; moreover, this interruption can also be disruptive to others physically present.

The telecommunication asymmetry problem can be addressed at different points of the communication flow. A common approach is to use a call-management agent that mediates the communication while it is *in progress* –this could be anything from simple call screening on an answering machine, to a more complex agent that establishes a dialogue with the potential recipient, negotiates the relevance or urgency of the call, and learns the user's preferences for future transactions. An alternative approach, the one we have taken, is to provide potential callers with enough information *prior* to the interaction, thus enabling them to make an informed decision regarding the availability of the targeted party, and the option to use a perhaps less intrusive, albeit less rich, medium. This approach is analogous to what happens between co-located people in the physical world, where one can appraise the situation prior to making a move to communicate and, based on that assessment, decide how best to act. Both these approaches, and their variants, clearly complement each other and certainly address different audiences and situations.

Furthermore, we have all been asked when arriving home, "So, how was your day?" or received a phone call from one of our intimates during the day asking "How are things going?". Evidently we want to be aware of what goes on with the people we are emotionally close to. Having a persistent awareness of the people we care about would be a form of communication in itself, a non-verbal form, which could strengthen the existing sense of connectedness and foster additional communication.

WatchMe is a context-aware personal communicator in the form of a wristwatch. It provides several channels of communication: synchronous and asynchronous verbal communication (text messaging, voice IM, phone call), as well as channels of implicit and non-verbal communication. The initiator of the communication can choose the appropriate modality based on the inferred situation, or use a channel to negotiate availability and then migrate to a mutually appropriate alternative channel.

In essence, the system uses multiple sensors to automatically collect aspects of external context. The data is abstracted to relevant telepresence snippets and conveyed to the people who have a strong influence on our emotional state, in other words, on our internal context. The information provided is iconically represented and includes the person's location (taken from GPS data), his physical activity or mode of locomotion (classified from wireless accelerometers), and whether he is in a conversation (obtained from audio analysis).

We believe that augmenting telecommunication with relevant telepresence will help enhance collaborative living between mobile household-members, or insiders. Nonetheless, this telepresence is not intended to give the remote person a sense of "being there", as defined by Minsky (1980). Rather, it is lightweight relevant context snippets that insiders can interpret and make sense of. Insiders are people with whom we have an established relationship, some common grounding, and a basis of trust, so they are people familiar with our routines, habits and preferences. Providing them with small pieces of our surrounding context can help them assess our availability for communication and choose the most appropriate medium. It can assist in the task-oriented communication aspects of collaborative living, i.e. household logistics and coordination of daily activities.

Undoubtedly, there are numerous privacy concerns vis-à-vis disclosing to others our surrounding context information, which would enable them to infer our availability. Therefore, this system would only be appropriate for a subset of people in our lives, people emotionally and socially close for whom we want to be available, that is, insiders only.

WatchMe also supports the connectedness-oriented facet of collaborative living communication, when household members are not co-located. The goal is not to replace face-to-face interaction between insiders, but rather to provide communication channels that can help maintain and strengthen their existing sense of connectedness and emotional need of staying in contact, when they cannot share the same space. The lightweight telepresence can provide a general sense of remote awareness, which in itself is a form of non-verbal communication, as well as comfort in knowing where family members are. When an insider checks another's telepresence status, either to infer how best to contact him or just for general awareness, this "thinking of you" information is conveyed to the remote person. This can strengthen the affective bonds and foster more communication between the individuals.

Computers can efficiently collect data and detect patterns, however they are less effective at understanding the subtleties and complexities of our life. Humans are experienced in understanding situations, gauging the importance of messages, and gracefully negotiating availability. Because insiders are familiar with our habits and preferences, they can effectively interpret our communication availability based on small pieces of relevant context information. Telecommunication devices are already widely used between remote insiders to keep in touch. By enhancing mobile communication devices with lightweight telepresence and various communication channels, we can support both task-oriented and connectedness-oriented communication needed for functional and socially healthy collaborative living.

1.3 Social presence, awareness and connectedness

WatchMe is a platform for communication and awareness between members of a closely-knit group. It seeks to maintain and strengthen the existing sense of connectedness between insiders. Although the terms awareness, connectedness and social presence are strongly related, there is no consensus on their exact meaning in the literature on mediated-communication [Rettie03]. Some use the terms to describe the medium, while others apply it to refer to the participants' perception.

Social presence has been defined by Short *et al*. [Short76] as the "degree of salience of the other person in a mediated communication and the consequent salience of their interpersonal interactions". They associate social presence with concepts related to face-to-face situations: intimacy and immediacy. Intimacy [Argyle65] is a function of verbal (e.g. conversation topic) and non-verbal behaviour (e.g. physical proximity, eye contact). Immediacy [Weiner68] is the psychological distance the participants place between themselves. Immediacy behaviours can be verbal or non-verbal, including facial expressions such as smiles or nods, gestures such as forward leaning, and choice of words (e.g. “we”, as opposed to “I” or “you”, conveys a feeling of closeness). These behaviours help maintain intimacy and enhance social presence [Gunawardena95]. Others [Danchak01] believe that the level of intimacy is a function of bandwidth and immediacy behaviours.

Awareness is “an understanding of the activities of others, which provides a context for your own activity”, according to Dourish *et al*. [Dourish92]. The numerous definitions –often interchangeably used– of awareness have been categorized [Christiansen03] into four groups: workplace awareness, availability awareness, group awareness and context awareness. Moreover, as Schmidt [Schmidt02] points out, the term awareness is ambiguous and unsatisfactory and therefore it is being used in combination with another adjective (e.g. peripheral awareness, mutual awareness, general awareness, etc.); it is “being used in increasingly contradictory ways… it is hardly a concept any longer”. As occurs with social presence, some definitions understand awareness to be an attribute of the system that enables the feeling of awareness, whereas others refer to the user’s perception.

Maslow’s Hierarchy of Needs (1943) includes five categories: physiological, safety, belonging/love, esteem, and self-actualization. According to this theory, humans are motivated by unsatisfied needs –the lower needs (physiological) must be satisfied before the higher ones (self-actualization) can be achieved. The term connectedness, generally defined as the feeling of being in touch, relates to Maslow’s categories of belonging, love and self-esteem. Connectedness is an emotional experience that is caused by the presence of others, however it is independent from it [Rettie03]. Even when there is no exchange of messages, the knowledge that others are online in IM conveys a sense of connectedness [Nardi00].

The “affective capacity” of a channel has been defined as how much affective or emotional information a channel lets through as compared to the total amount of information that is passed [Picard97]. This is not necessarily the same as bandwidth –pointing a camera at a wall uses bandwidth but it does not transmit affective information [Danchak01]. The affective capacity of a channel is related to what is referred to as media richness.

We understand connectedness to be an affective state. One of the goals of systems that support collaborative living is to maintain and strengthen this feeling of being in touch, since this feeling can decay over time [Patrick01] when people are not co-located. The sense of connectedness can be stimulated or boosted via communication. The strength of the stimulus is not necessarily a function of the channel bandwidth or the content of the message, but rather what it triggers in the recipient and/or sender. An asynchronous text message can be very intimate, a full-duplex synchronous voice conversation does not always carry emotional richness, just knowing that someone is thinking of you at that moment can be very powerful even if it is only conveyed through a vibration or blinking light.

Connectedness is a goal and communication is a means to stimulate it. Awareness is a form of communication in itself and therefore the sense of connectedness can be enhanced via awareness systems.

1.4 Context-awareness

This work shares certain features with what are commonly called context-aware systems. These systems usually refer to the current setting an individual is in [Schilit94, Pascoe98, Dey01], e.g. their location, whether they are logged into a system, what devices they are using, what exactly they are doing, or physical characteristics of the environment such as the temperature, noise levels, or lighting conditions. A person's social context usually refers once again to the immediate situation: whether others are present, is it a formal or informal gathering, who is speaking, etc. These two types of context are a person's *immediate context*. These are things that can, at least to a certain extent, be measured or sensed.

Nevertheless, more often than not a person's context is strongly affected by events and/or people that are neither physically co-located, nor are co-temporal. This cannot be measured with sensors, classified with machine learning techniques, or modeled. Yet it is arguably the most significant part of context, at least from that individual's personal/social perspective.

Although this research addresses context and (tele)communication, its emphasis is not on how we engage with the settings we are in but rather how we engage with others who are not in the same setting, and how the fact that they are not co-located has relevance on our interaction.

1.5 Document overview

This chapter has explained the motivation for technology-mediated collaborative living, given a brief overview of the *WatchMe* system, and defined some terminology. Chapter 2 discusses design issues to be considered when building technology to support collaborative living. The larger part of this thesis, Chapters 3 and 4, describes the design and implementation of the prototype in detail. Results of the different evaluations carried out are reviewed in Chapter 5. Chapter 6 situates this piece of research in the related work and, finally, Chapter 7 concludes with the contributions of this research and discusses future work.

2. Design criteria

Technologies developed for, and lessons learned from, the domain of computer-supported collaborative work may also be applied to the collaborative living environment. Despite certain parallels, this environment differs from traditional CSCW and has distinct requirements. For instance the eight challenges of CSCW [Grudin94], such as the disparity between who does the work and who benefits from it, or the challenges in obtaining critical mass, do not apply. Likewise, metrics of productivity and efficiency are not relevant. In organizations, members of work groups do not always have shared goals, meanings, and histories [Heath96]. In contrast insiders, by definition, are people with whom we have an established relationship and some common grounding. In this chapter we describe design criteria to be considered when developing technology-mediated systems intended to support collaborative living. Some of these address the functional or task-oriented aspects of the communication, whereas others refer more to the social or staying in touch (connectedness) facet.

WatchMe is a personal communicator, in the form of a wristwatch, that aims to enhance communication and awareness between members of a closely-knit group, specifically people with whom we have day-to-day living social relationships, i.e. insiders. We find that a fundamental aspect of maintaining a functional and socially healthy household is communication; that is, communication to coordinate daily activities as well as to build and sustain the emotional relationships between the household members.

Awareness between remote insiders can serve several purposes. It can help one infer the other's availability and suitable modality for communication, hence reducing the "Is now a good time? Do you have a minute?" genre of phone calls. Beyond enabling more appropriately timed communication, often regarding the coordination of household activities, it can also directly facilitate the coordination –for instance by knowing where the kids are and who has not been picked up yet, or that someone is in a certain vicinity and can run an errand. Awareness, however, serves more than these very task-oriented facets of domestic communication. It is a form of communication in itself, and there is a certain comfort associated with knowing where our loved-ones are, and what they are up to. A goal of the system is to help users reduce interruptions at inconvenient moments, yet still be available. As one of our subjects noted, it “builds relationship to be available if possible”, i.e. it is also a way to strengthen connectedness.

WatchMe provides several channels of communication, enabling the initiator to choose the appropriate modality based on the inferred situation, or use a channel to negotiate availability and then migrate to a mutually appropriate alternative communication channel. The system includes multiple sensors to automatically collect aspects of external context. This data is abstracted to relevant telepresence snippets that can be displayed iconically on the watch; it includes a person’s location (from GPS data), his physical activity (from wireless accelerometers), and whether he is in a conversation (from audio analysis). The GPS data is analysed for frequented locations, which the user can label, meaning that when next there this information will be shared with his insiders. The physical activities displayed are the person’s mode of locomotion, such as walking, biking, or in a vehicle. The audio data is simply analysed to determine whether it is speech, no speech recognition is performed, nor is any of the audio saved.

A number of CSCW systems have addressed awareness, interruptions, and inferring the best time for communication. For instance, Horvitz’s notification system [Horvitz99] used probabilistic Bayesian models to infer a user’s focus of attention based on his computer activity (use of software, mouse and keyboard actions) and location (predicted from scheduled appointments in calendar, and ambient acoustics in user’s office). The criticality of an email was assigned based on text classification; for example influenced by whether the mail was from a single person or an email alias, whether it was only to the user or to a list of people, certain time related phrases like “happening soon”, etc. Messages were delivered if the calculated cost of deferring it was greater than the calculated cost of interrupting. This scheme is certainly effective provided the user's activity is confined to the computer, or detailed in a calendar. However in life outside of the workplace (whether that be in an office, or take-home work) most of our activities are not measured by keyboard use. Furthermore, although people use calendars in the home setting, they are typically less detailed and structured than shared work calendars, and are usually not online.

Milewski's Live Addressbook [Milewski00] helped people make more informed telephone calls, from a PDA or desktop browser, by providing dynamic information of where the person was (phone number) and his availability (Available / Urgent only / Leave message / Do not disturb). This information was revealed to the person's Buddy List, as well as an optional text presence-message such as "working". Although this system presented the potential caller with information to help him make a decision regarding its appropriateness, this came at the expense of burdening the callee, who had to remember to manually update it. The system, however, would prompt the user when it found mismatches, for instance if based on keyboard activity the person was understood to be on a PC in one location but his last update indicated a different one. While this system offers the user a lot of control, only revealing private data when he wants, the awareness information is only relevant if it reflects the user's current state, i.e. if he continually updates it. Another example, and more can be found in Chapter 6, is the Hubbub system [Isaacs02] which addressed awareness, opportunistic conversations, and mobility. This system extended a text Instant Messaging (IM) client, on a mobile device and desktop, with awareness through musical sounds unique to each user, enabling others to know (without looking) who had just turned from idle or offline to active. A significant fraction of the communication occurred immediately after the person turned active, suggesting the usefulness of awareness information for opportunistic interaction. This system also had location information manually updated by users.

The idea of inferring the best mode of communication, or predicting the best time for it, is not new. Looking at these systems, and others, we identified several powerful concepts, such as multi-modal communication, automatic detection, awareness, and the prediction of availability. These ideas are also applicable outside of the workplace, however, in dealing with everyday life, things are much more complicated and there are a number of design considerations that need to be addressed. We enlist them below.

2.1 Human-centric

The main goal of collaborative living systems is to help insiders better bridge the times when they cannot share the same location. Hence, any system designed to support this closely-knit group should take a very human-centric approach, as the key purpose is connecting people. The technology should facilitate the exchange of communication and help maintain and strengthen the existing relationships in the social network. The human-centric aspect cannot be over-emphasized, it is the core principle that the other criteria derive from. Where applicable, not only should a human-centric approach be considered in the design of the system functionality, but also its physical form. The human-centric theme is also pertinent to many CSCW communication systems.

2.2 Anywhere, anytime

Life is not confined to an office or a home, it occurs everywhere. Hence, collaborative living systems, to be useful both to the user and to the rest of his network, should be accessible all the time. This has design implications both regarding ubiquitous functionality as well as form factor. One can envision a system that connects a parent’s office to his child at home and enables some form of awareness/communication that helps strengthen the bond between the two. Although this system would fall under the category of collaborative living, it is, in fact, only facilitating the connection provided the individuals are in very specific locations –it is connecting the people, and not just the locations, however it is conditional on the location. In conventional CSCW, although there are systems that support mobility, it is not a main system requirement.

2.3 Awareness

Collaborative living systems have two main goals: facilitate and foster communication between insiders, and help maintain and strengthen their existing sense of connectedness. Awareness features can contribute to both these aspects. Systems that provide availability-awareness help share the decision of initiating communication between the two parties, aiding them to communicate at an opportune moment. If the system supports multiple communication modalities, this information can also help the initiator choose the most suitable medium. Moreover, systems that provide general-awareness can help strengthen the social connections between the existing closely-knit network.

Not all collaborative living systems need to support both of these aspects. However, in either case, serious consideration should be given to exactly which information is provided, how much of it, at what level of abstraction, and how it is presented. All these have implications not only on privacy, but also on the usefulness and effectiveness of the system. To the extent possible, the data should be gathered automatically, requiring minimal explicit user-input.

As mentioned previously, awareness systems are an active research area within CSCW. Awareness has been found to be conducive to collaborative work, people prefer being aware of others within a shared space, and use this information to guide their work [Erickson99].

2.4 Multiple modalities

The key motive for providing multiple communication channels is to enable flexibility, permitting the users to converge on a mutually convenient mode. Different modalities, whether verbal or non-verbal, have different affordances and facilitate different degrees of expressiveness. Task-oriented communication typically involves the direct transfer of information, hence systems supporting this aspect should enable lightweight fast interaction;

immediacy is important since timing is often critical. For communication that is more socially-oriented, systems should support ways to maintain a persistent “conversation”, perhaps across different modalities, over long periods of time.

Different tradeoffs must be considered: synchronous vs. asynchronous, voice vs. text, production vs. consumption, verbal vs. non-verbal. Which mode is the most appropriate depends on the situation and the particular preferences of the users. The more modes provided, the better the users will be able to adapt to the different situations. This is especially relevant for mobile systems. Additionally, providing several communication channels permits users to negotiate availability via a lightweight channel and then upgrade to another, for example from asynchronous text to full-duplex synchronous voice. Other aspects to take into account are bandwidth requirements and potential intrusiveness of the different channels. These issues must be deliberated in any computer-mediated communication system.

2.5 Attentional demands

Collaborative living systems are intended to support our communication needs with our insiders, typically while we are going about our life and performing other tasks. Hence they should be designed to demand minimal attention and have low intrusiveness –this is especially important when the user is mobile. They must have low intrusiveness yet have the ability to catch our attention and alert us when relevant. The exact balance between these two seemingly contradictory aspects varies from application to application. The system should be able to alert in a subtle way and then let the user decide when to attend to the information. Besides intrusiveness to the user, it is also important to consider the intrusiveness to others in the environment. Attentional demands must be considered in any multi-tasking environment.

2.6 Alerting mechanisms

The systems should not be constantly intrusive, nor continually demand the user's attention, however they must be able alert the user. Alerts can generally range from private to public, and from subtle to intrusive [Hansson01]. Different combinations are more appropriate for different situations, and the user's preferences can vary in this regard. Although we believe that subtle and private alerts are generally preferable for collaborative living systems, the choice should be left to the user, and systems should provide different options.

A factor to consider in these systems is the persistence of the cues. Alerts can be visual, auditory or tactile. Persistence is mostly related to visual cues, however it could also be relevant to haptic ones, e.g. in the case of a device that emitted heat to indicate some remote state. Whether an alert should be persistent or not depends on the nature of the information it conveys. The systems should not create new obligations, rather they should either create opportunity for communication and/or help the user recognize a communication opportunity – these are two different cases.

2.7 Emotionally engaging

This principle relates to the connectedness aspect of collaborative living. The more emotionally engaging the system is, the more likely it is to maintain or strengthen the affective bond between the individuals and foster additional usage. The objective is the affective state it triggers in the recipient, regardless of the actual content of the message transferred. Photographs can be emotionally rich, for example if they remind us of cherished moments or people. A system that enables the sharing of photographs has the potential to be emotionally engaging –in this case the actual content transferred, the photo, can have emotional value. A system that lights an LED (light emitting diode), in a situation meaningful to the parties of the communication, can also trigger an affective state although the content of the transfer was simply a command to turn on the light.

2.8 Privacy

There are many different aspects of privacy that should be considered: what and how much is being conveyed; what patterns emerge from this information; is the data persistent; is their plausible deniability; who has access to the information. The last point refers both to strangers and family/friends. Interesting social situations may arise if A considers B his insider, but not vice versa. Or in the case of asymmetric relationships such as parenting, where a parent may want to know the location of a child, whereas the teenager (who no longer considers himself a “child”!) objects. Other factors to consider are the transfer mechanisms and security of the data: is it encrypted, is it peer-to-peer or being routed through a server.

The tradeoffs of awareness and privacy are a recurrent theme [Ackerman00]. People want control on an ongoing basis regarding what information is given to which individual [Goffman61]. Insiders, people with whom we have an established relationship and are emotionally close, typically already are privy to a lot of information about us. For this reason, in some cases privacy might be an even greater concern, though in others, the constraints can be relaxed.

The following two chapters provide an in-depth description of the *WatchMe* implementation and show how the aforementioned design criteria were addressed.

3. WatchMe

The previous chapter discussed design criteria to be considered when building technology-mediated systems to support collaborative living. This chapter describes the *WatchMe* prototype and how those issues where addressed. The description starts with motivation, followed by the functionality provided through the graphical user interface, before going into details of the architecture and hardware. Specifics of the classification algorithms, the outputs of which are displayed iconically in the interface, are explained in the following chapter.

WatchMe is a system that provides insiders with awareness of each others' activities, and multiple channels for communication between them. We believe that having a level of understanding of a remote person's context, especially people with whom we are emotionally close, will help evaluate whether and how to communicate. Furthermore, we believe that this awareness can foster communication at opportune moments.

Cues from the physical world often help us infer whether a person is interruptible or not. An office with a closed door, for example, may indicate that the person is not around, or does not want to be disturbed. From prior knowledge we may know that this individual is easily distracted by outside noise and therefore keeps the door shut, and that it is perfectly acceptable to simply knock. If a door is ajar and voices can be heard, then perhaps the person is unavailable –that could depend on the nature of the relationship and the urgency of the topic. When someone is in a conversation, and consequently unavailable, there are different acceptable ways to try and catch their attention indicating that we would like to talk to them, such as trying to catch their eye, or lingering in the periphery of the conversation without intruding on it. The targeted person may provide verbal or non-verbal feedback regarding his near-future availability, or he may simply ignore the gesture. *WatchMe* provides lightweight telepresence cues (location, physical activity, presence of conversation) to our insiders, helping them infer our interruptibility, as well as mechanisms for them to catch our attention, such as causing the watch to vibrate or its display to light up. Insiders, being people know us well and share our life, have insight into the potential meaning of these cues, and their implications on our availability.

During mobile phone calls, people have a tendency to say where they are [Laurier01]. This not so much to provide exact geographical information, rather to coordinate future moves, such as meeting up, or to provide context as to whether they can talk or not. Location information is provided on the watch to insiders to help them infer a suitable time and mode for communication. Moreover, it assists in the coordination and synchronizing of the household members and activities. For parents, being aware of the whereabouts of their children is also a security issue. Location information is gathered by GPS and analysed to determine previously unnamed frequented locations. The user has the option to name the identified locations, or indicate that they should be ignored by the system. When the user is next identified at a named location, this information is automatically shared with his insiders. In this manner, the user decides which locations he wants to share, simply not naming the others. The system also analyses routes, enabling the watch to display a user's current location, or next predicted one and expected time of arrival, or time elapsed since his last known one.

In addition to the presence of speech, and a remote user's location, situational context is also provided by displaying the person's physical activity –classified from wireless accelerometers. As has been demonstrated [Bao03], with accelerometers it is possible to classify a fine level of activity, including ironing clothes or brushing your teeth. Whereas this may be useful to monitor elderly people, to determine whether they are capable of performing the activities of daily life, we believe that most people would not feel comfortable revealing that amount of detail, on a continual basis, even to their insiders; it is also unclear whether the potential recipients of the information would want that granularity. Which activities are classified should be up to the user to decide, based on the amount of information he is willing to share, as well as the number of on-body accelerometers he is willing to wear. We have helped develop wireless accelerometers (Section 3.3.3) that are small and light enough to be attached to clothing without constraining the wearer's movement, in order to provide this flexibility of choice. Our implementation focuses on modes of locomotion (static, walking, running, biking, vehicle), believing that this is the least common denominator of what most people would be willing to share, however, *WatchMe* is not limited to these activities.

What the combination of location, presence of conversation, and a person's mode of locomotion indicate as far as communication availability will vary from person to person. A person identified to be at "work", "static" and "in conversation" might be considered unavailable if the situation is inferred to be a meeting. These same parameters, for example, for a construction worker might indicate that he is on lunch break talking to his buddies, perhaps an appropriate time for communication. Knowing that a person is "on the way home" and "walking" may indicate that they are running an errand in a store, since their commute involves driving, and this is perhaps an opportune moment to remind them to get milk. The situation is to be evaluated by our insiders, people who can interpret their meaning, know our routines and preferences. They are the ones to judge the importance of the message versus the "cost" of the interruption.

WatchMe provides various communication channels enabling users to converge on a mutually appropriate mode, while giving them some control over the awareness data revealed. The users decide which locations they want to share, only naming those they are willing to reveal. Although modes of locomotion can be inferred to a certain extent from GPS data, GPS does not work everywhere. The accelerometers are a much more reliable source of information and enable the classification of many more activities, however the user always has the option not to wear them. A user can also manually set that he is out of cellular range, or simply exit the application. We want to enhance communication and awareness, and create opportunity for communication between insiders, not create obligation or intrude on privacy.

The user's privacy is further protected by symmetry, or reciprocity, of awareness. A user cannot receive context data without sharing his own. As stated by Licklider in his seminal paper [Licklider68], "A communication system should make a positive contribution to the discovery and arousal of interests". In *WatchMe*, when an insider views another's context data, this action is revealed to the person in question, by displaying their photograph on the watch. Hence there is not only reciprocity of awareness, but also reciprocity of interaction.

We believe that enhancing telecommunication devices with lightweight telepresence information can help insiders infer another's availability for communication. Augmenting the communication device with multi-modal channels will permit flexibility, enabling them to converge on a mutually appropriate modality. Furthermore, the telepresence snippets can provide general awareness, helping maintain or strengthen the insiders' existing sense of connectedness.

3.1 Watch – form and function

3.1.1 Why a watch?

A watch is an artifact very assimilated into our lives. It is an object many people wear, and glance at numerous times a day. It is always accessible, always on, and in the periphery of the wearer's attention. Watches are very noticeable, but in a non-intrusive manner. Besides serving a defined function, they are also often related to our emotional lives, having been given to us as a present from a dear one, or passed down from a family member of a previous generation.

The *WatchMe* device had to be easily accessible and frequently visible. It had to include mobile phone capabilities since one can hardly imagine a system for intimate telecommunication that does not include duplex synchronous voice. Building such a system into a watch is a challenge, due to its physical size. A key requirement of the user interface is that it must convey a lot of information in a relatively small amount of space, and in an aesthetically pleasing manner. An additional requirement was a device that could comfortably support switching between the modalities. A watch is a convenient location for conveying information as well as being in a place that is easily manipulated –albeit with one hand.

3.1.2 Watch User Interface

A wristwatch is a personal device, but it is also very public. We frequently look at other people's watches to know the time when it would be socially awkward to look at our own. Watches are also often a fashion statement, intended to be observed by others. Since it is at the seam of the personal and the public, our graphical interface has tiers of different levels of information, with different levels of privacy.

The face of the watch is visible to all and conveys information accessible to all, i.e. time. People glance at their watch more often than they perceive. By embedding this high-level information in the watch's default mode, we can keep track of our loved-ones subconsciously and continually throughout our day. The top level (Figure 3.1), the default screen, also embodies other information meaningful only to the owner. The owner of the watch chooses a unique icon and position around the watch face for each insider; although this is visible to others, they do not know the mapping from icons to names.

Fig. 3.1 The *WatchMe* prototype displaying the main screen (right) with icons representing insiders. The left image shows the size of the current version.

Research has shown [Isaacs02, Grinter03] that with text messaging clients, users interact recurrently with 5-7 people on a general basis. To play it safe, we initially chose to display icons for up to eight insiders, however, as explained below, we later reduced the total to six. At this top level the colour of the icon indicates availability, fading to the background colour in thee degradations: the most faded colour indicates that this insider does not have cellular coverage (either because he is out-of-range, or has chosen to be "disconnected"), the midway colour indicates that the person is in a conversation and hence probably less available.

From a full-colour icon it is not possible to infer availability without going down a level in the interface and seeing more detail (Figure 3.2); we refer to this as zooming in. This is achieved by selecting the corresponding icon, via the Left/Right navigational buttons, and then pressing the Down button. On this screen a pre-selected image of the insider appears lightly underplayed in the background, as do the continuous lines of the design.

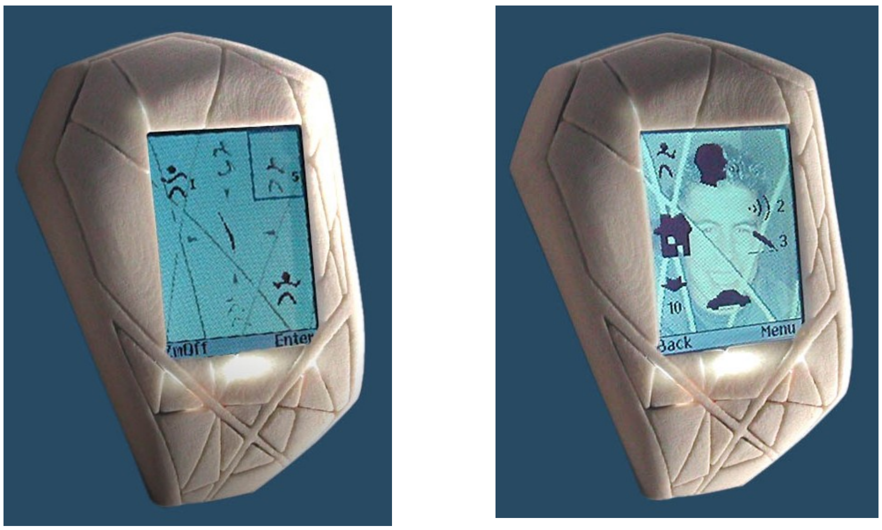


Fig. 3.2 “Zooming” into a user’s context. The left image displays the top level of the interface with the cursor over the icon of the insider we want more detail on. The detailed context screen is on the right.

The more detailed information that can be viewed here (described clockwise from the top left) is the specific person’s assigned icon, whether that person is engaged in a conversation, how many voice and text messages this person has left, and the person’s activity or mode of locomotion (walking, vehicle, biking, etc). Also displayed is his current location or next predicted one and expected time of arrival (ETA), or his last known location and time elapsed since departure. For example, in Figure 3.2, we see that Joe left home 10 minutes ago, that he is driving and in a conversation, and that he has sent us two voice messages and three text messages.

Since Joe is driving and also talking, this would probably not be a good time to phone him. For an insider, these snippets of information can go a long way. With a combination of prior knowledge and a form of lightweight telepresence provided by the watch, we believe that it is possible to quickly form a meaningful interpretation. For example, knowing Joe and judging by the time and that he is driving and talking, it is possible to presume that he has already picked up his buddy and is heading to the gym. If “gym” is a location Joe has revealed, once the system has enough information to predict he is heading there, the icons will automatically change to reflect that (gym icon, direction arrow, and ETA). We are not trying to present an inferred abstraction of the user’s availability, based on a combination of sensor data, but rather abstracting the data from each individual sensor and letting the remote person decide on their aggregate meaning and implication on the best time and mode for communication.

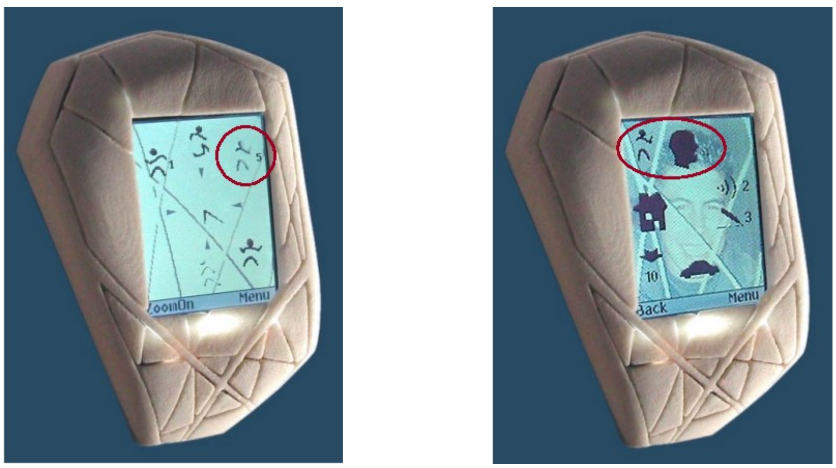


Fig. 3.3 The detection of speech is a strong indicator of unavailability. Hence, it is conveyed in the main screen (left) colour-coded into the icon, and not only in the detailed context screen (right).

The user interface must convey a lot of data on a small display. Much effort was put into delivering this information in an aesthetic form and in a format that can be interpreted at a glance. Too many features would require close examination and the user’s direct attention, defeating the purpose of having a device at the periphery of the user’s vision. Not conveying the most salient features at the top level would require added manipulation, as the user would have to navigate to other screens. Hence, the top level displays the detection of speech (Figure 3.3), the most relevant context snippet, as well as the number of messages received from each insider (Figure 3.4). Speech is indicative of social engagement, and it has been found to be the most significant factor in predicting availability [HudsonS03], therefore it was coded into the icon colour. Since *WatchMe* is ultimately a communication device, the number of messages received from a person should be readily visible. At this top level only the total number of messages is displayed. In order to know whether they are text or voice messages, it is necessary to zoom into that insider. Access to the messages themselves, and the sending of messages to this person, is achieved by zooming in an additional level –this is a total of two key-presses once the cursor has been placed on the relevant insider.

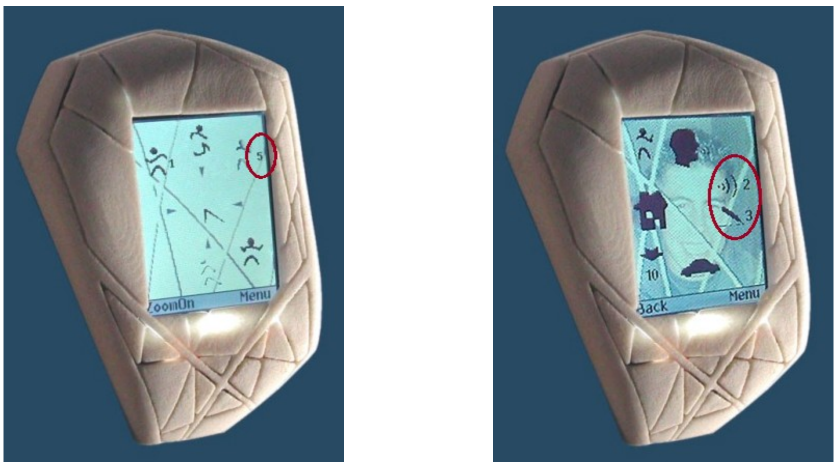


Fig. 3.4 The main screen (left) shows the total number of unread and unheard messages received from each insider. The detailed context screen (right) indicates whether they are voice or text messages.

3.1.3 Thinking of you

A fundamental part of communication is its reciprocal characteristic, or its "mutually reinforcing aspect" [Licklider68]. When an insider views another's detailed information (in this case, that she is biking to work and expected to arrive in 18 minutes), her image appears on the reciprocal wristwatch (Figure 3.5). In this way one can have a notion of when a specific insider is thinking of the other, and this information may subsequently stimulate an urge to contact that person. This conviction is supported by [Isaacs02] where a significant fraction of the communication happened immediately after a party appeared online.

Fig. 3.5 When an insider thinks about another and views her detailed context data (left), the “viewer’s” photograph will appear on the “viewed” insider’s watch (right).

Knowing that someone is thinking of you can foster communication. When the picture appears on the “viewed” insider’s watch, one of the following could occur:

- The picture popping up may go unnoticed, especially since it disappears after a couple of minutes, so the “viewing” insider is not interfering with the “viewed” one in any way.
- The “viewed” person notices the picture but decides not to reply or divert attention from his current action.
- The “viewed” insider notices the picture and responds by querying the availability of the other user, which causes his or her picture to appear on the other’s watch, similar to an exchange of glances or smiles without words.
- The “viewed” insider decides to phone the “viewer” or engage in another form of verbal communication, i.e. text or voice messaging.

This “thinking of you” mode, which we also refer to as a “smile”, creates opportunity for communication between the two insiders, without creating obligation. It is analogous to co-located situations, as exemplified in the following scenario:

> *Two close friends in a living room each reading their own book. One glances up and checks on how the other is doing. That person may not notice the gesture at all, or may somehow sense it and yet decide to ignore it, or might smile back providing a subtle emotional acknowledgement before they each continue reading their respective book, or a (verbal) conversation may be triggered.*

A photograph is an affective way of conveying the smile. Pictures can be very expressive and emotionally engaging. We believe that having a spontaneous non-verbal communication exchange, triggered by a remote insider thinking about you, could be a powerful experience.

3.1.4 Communication modes

The watch supports text messaging (also called Instant Messages, or IM), voice messaging (also referred to as voice Instant Messages or voiceIM), and phone calls. The content of the text and voice messages, as well as the ability to compose messages or place a phone call to the insider, is accessed through the third layer of the interface (Figure 3.6).

Instant Messages are viewed and composed on the same screen (Figure 3.6b). The text is entered by choosing letters from a soft-keyboard, using the four navigational buttons –this can be done with just one thumb. An additional keyboard, including numbers and symbols, can be toggled. A scroll bar on the side enables scrolling through the dialog. Texting on any small device can be very tedious, and this interface is no different. The IMs in *WatchMe* are intended to provide insiders with a non-intrusive channel for fast interchanges, for coordination of activities or to negotiate communication on a different channel. IMs can be semi-synchronous provided there is good network coverage, however this near-synchronous exchange can only take place with one insider at a time –incoming text messages from others will be saved, but not displayed simultaneously.

Voice Instant Messages are another form of asynchronous communication. They are, however, considered to be more synchronous than voice-mail is. If both parties are available, the communication exchanges can be fairly fast, less so than using Push-To-Talk, yet still semi-synchronous. VoiceIMs are recorded and sent via the interface shown in Figure 3.6c; they are received through the menu shown in Figure 3.6a. For the sender, voiceIM requires less effort to produce, however for the recipient, text is faster to read. For both sender and recipient, the fact that the message could be overheard may be a concern.

(a) (b) (c) (d)

Fig. 3.6 (a) Interface through which all messages from a specific insider are accessed or sent. (b) an IM dialog, showing messages sent and received, an a soft-keyboard. (c) Interface used to record and send voice IMs. (d) Interface for sending “smiles” or “gestures” .

Both text and voice IM are convenient for small transfers of information, enabling a persistent connection with short spurts of dialog throughout the day. But the content of the message need not necessarily be limited to the direct transfer of data, it could just as well be communication for the sake of feeling connected. These asynchronous channels are also a means to converge on a mutually suitable communication mode, for example to negotiate a phone call.

The watch also supports non-verbal forms of communication. These can be implicit or explicit. When an insider views another's context data, that insider receives a smile in the form of a photograph; as explained previously. Although the smile was triggered by the user's action, it was not proactively sent. The picture appearing on the remote person's watch makes him aware of whom is thinking of him, and in addition indicates that he may soon be contacted by this person; or gives him the option to contact them. A user can also explicitly send another a smile (Figure 3.6d). These include either triggering a picture to appear on the remote watch, or causing the backlight to turn on, or the watch to vibrate. The more intrusive forms of the smiles, i.e. the light and vibration messages, we sometimes refer to as "gestures". The implicit form is analogous to thinking about someone, or remembering a certain event, and a smile coming to your face; except that the remote person sees it too. The explicit forms are similar to trying to catch someone's attention, for example across a room, by tentatively catching their eye or gesturing. When co-located, people have different strategies for negotiating contact with someone who seems to be busy; these often involve non-verbal cues letting them know that you are waiting for a moment with them at an opportune time. The non-verbal channels provided in the watch give the user options to do this remotely. In certain cases, the cue is the message, such as a smile or a wink. Insiders may also develop their own coded messages using the non-verbal channels.

The watch has several alerting modes, giving the wearer flexibility of its intrusiveness. The number of unread and unheard IMs are indicated next to the icon of each insider, viewable by the user when he glances at his watch. Since a user may want to be alerted, the system provides any combination of turning on the backlight of the display, causing the watch to vibrate, and playing an audio file. These can be defined independently for text and voice IMs, per insider. So a text message from person A may cause the backlight to turn on for two seconds, whereas a voiceIM from that person may cause the watch to vibrate half a second. A text message from person B may trigger the backlight of the display and the sound of a bird chirping, whereas notification of a voice message from them may be a pre-recorded snippet of their voice. When a user views another's context data, his photo appears on the other person's watch. Although this occurs automatically, the recipient can control how long the picture will be displayed before subtly disappearing. Additionally, the recipient can set his watch to light up or vibrate when smiled at, for example if wearing long sleeves and not wanting to miss the information. An insider may explicitly send another a gesture, in the form of light or vibration. Though these are intended to be more intrusive, the recipient can control their duration –the defaults are 500ms for the vibration and 2000ms for the backlight. The settings for the smiles and gestures are global, and not defined per insider. The system provides flexibility, enabling the wearer to decide how private and subtle, or how public and intrusive [Hansson01] the alerts should be.

3.1.5 Design

The user interface design was a continual process. The initial version displayed up to eight insiders at a time, represented by icons of cartoons. Although people found it playful and colourful, it was too overloaded to be understood at a glance. For instance, Figure 3.7 (left) shows Bart greyed-out, indicating that he is in a conversation, however it was often not noticeable to users until specifically pointed out. People also indicated the stereotypes of the icons, and what it might mean associating the Homer Simpson icon to an insider. In consequence, we decided to reduce the number of displayed insiders to six, and design simplified, more aesthetic, and more abstract icons. The new set of icons, those in the current version, went through several colour iterations until we found a combination which had enough contrast between icons and background, and could be seen even without the backlight turned on.

Some people have noted that the current icons are perhaps too similar, making it hard to differentiate between some of them, but that they would remember whom they represented by where they had placed them around the watch. Although we would like the icons to be aesthetic and consistent throughout the interface, we have implemented an option enabling users to download and associate new icons, both to insiders and to locations. If an icon does not exist for a specific location, its text name is displayed instead.

An initial version enabled a user to define the urgency of a text or voice IM. The screen with the context data displayed the urgency of the last message sent (Figure 3.7, top right corner), if one had been defined. The icon represented a clock and an arrow indicating whether the urgency of the communication was rising or decreasing as a function of time. People found this unintuitive and an unnecessary complication, hence it was eventually removed.

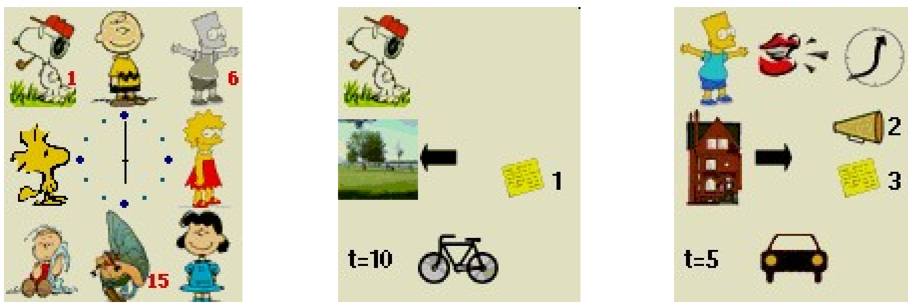


Fig. 3.7 Examples of the initial graphical user interface.

The form of the watch also went through different phases. Figure 3.8 shows exploratory variations of shape and screen rotation. In the version implemented, the orientation of the screen is at an angle, maximizing its legibility when placed on the left wrist; the bottom surface is curved in order to better conform to the body. In addition to the counter-clockwise skew, the top right corner of the screen is pitched slightly upward to reduce glare. The design of the outer shell was intended to make it more like a piece of jewelry, with lines of the shell continuing across the screen, emphasizing that it is one entity. The lines are reminiscent of those on the palm of the hand.

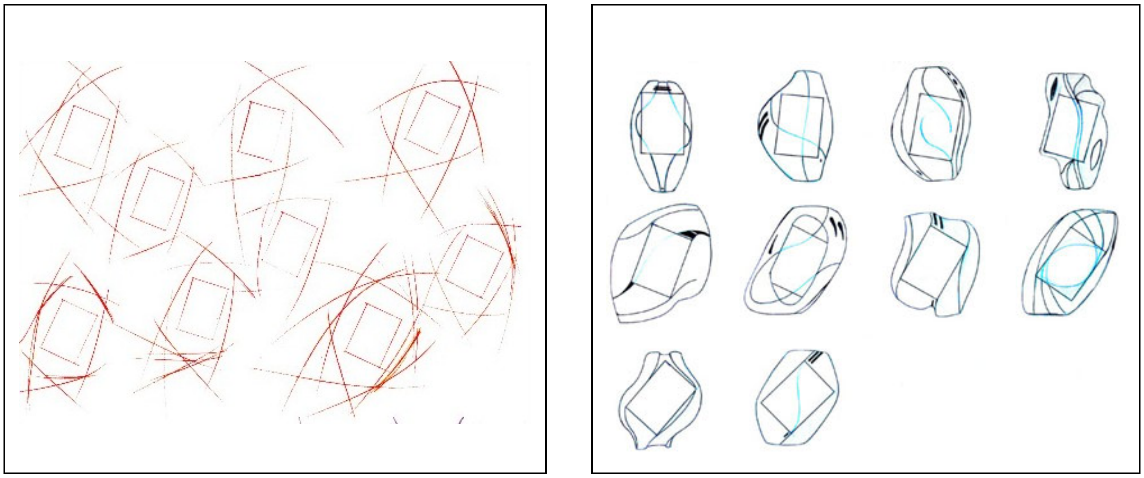


Fig. 3.8 Hand sketches exploring preliminary variations of shape and screen rotation. (Drawings by David Spectre)

The functional buttons were intentionally concealed, following the theme of private vs. public. There are two main axes subdividing the bottom half of the design. The point at which these axes meet is the main reference point in the location of the functional keys (Figure 3.9); this is also a point easily accessed blindly. At direct horizontal and vertical angles relative to this point are the navigational keys (Up, Down, Left and Right). Following these axes upward, the keys manipulating the lower edges of the screen (both left and right), between them and just under the screen is the Menu key and finally along the far ends of the axes the Connect (upper right) and Disconnect (upper left) keys with almost reversed appearance.

Some people liked that it was not apparent that it was a “piece of technology”, noting that it was “more like jewelry, less nerdy”, and said they would like the ability to have different outer shells, of different shapes and colours, for different attires and occasions. Others found it hard to distinguish which were the functional buttons, and remember what each one did. Figure 3.9 shows the rendering of a metallic shell we had envisioned making out of aluminum and titanium. In this version, although the buttons are revealed, they are not labeled.

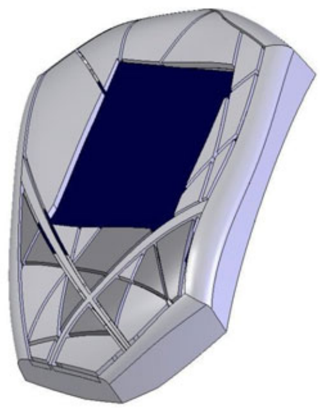


Fig. 3.9 Rendered prototype of metallic shell, emphasizing placement of buttons around two main axes.

3.1.6 Privacy

In any awareness system some of the information that is revealed is sensitive to some of the participants at least part of the time. In the course of developing *WatchMe* we encountered a number of privacy issues.

sensitive information: *WatchMe* reveals a lot of information about a user, but only the locations that he has chosen to name; raw geographic coordinates are never revealed. A user might see that another is at the bookstore, but where the particular bookstore is physically located is not displayed. Additionally, *WatchMe* has been designed from the beginning to be a system used by people who are intimate friends, who already share much personal information. People whom we are really close to know much more sensitive information about us than, for example, how long ago we left our house.

photographs: Photographs are very personal and a watch face is semi-public. People may be more sensitive in other cultures, but in ours we often display pictures of family, especially children, in offices and homes. We often carry them in wallets or purses, both to look at ourselves and to show to others. We now have them on phones as well, so displaying pictures of our loved ones on a watch is not that different. The detailed context information would not be readily understood by someone looking at our watch from a distance. It is also invoked only by specific user action.

reciprocity: *WatchMe* enforces reciprocity of data. A user cannot receive context data from another unless he is also sending his. There is also reciprocity of interaction: when user A views B's context data, A's photograph appears on B's watch. So a person cannot "spy" on another without them knowing they are doing so, regardless of whether it carries a positive or negative connotation.

peer-to-peer vs. server: The current implementation depends on a server to relay the messages between the users. Now that there is better support of server sockets on the phones, the architecture could be modified to be peer-to-peer, over a secure socket, adding another layer of security. Even in this version, no data is stored on the server.

plausible deniability: The user has control over the locations he decides to share with his insiders, and at any given time he can manually make it seem that his watch is "out of service" (out of cellular range), or that he is in a conversation. We have thought about randomly invoking the "out of service" mode to provide the users with plausible deniability and prevent them from having to explain why suddenly they were disconnected. In this way it can be attributed to a supposed bug in the system, when in fact it is a privacy feature. The user's location is only transmitted to others when he is somewhere he has previously chosen to name, however the hardware that he is wearing is keeping a history of where he has been, to detect these patterns and perform calculations of ETA. In addition to giving the user the option of not sharing the location, he should also have the option of not logging it at all or the ability to delete certain sections from it. No acceleration data or audio is saved.

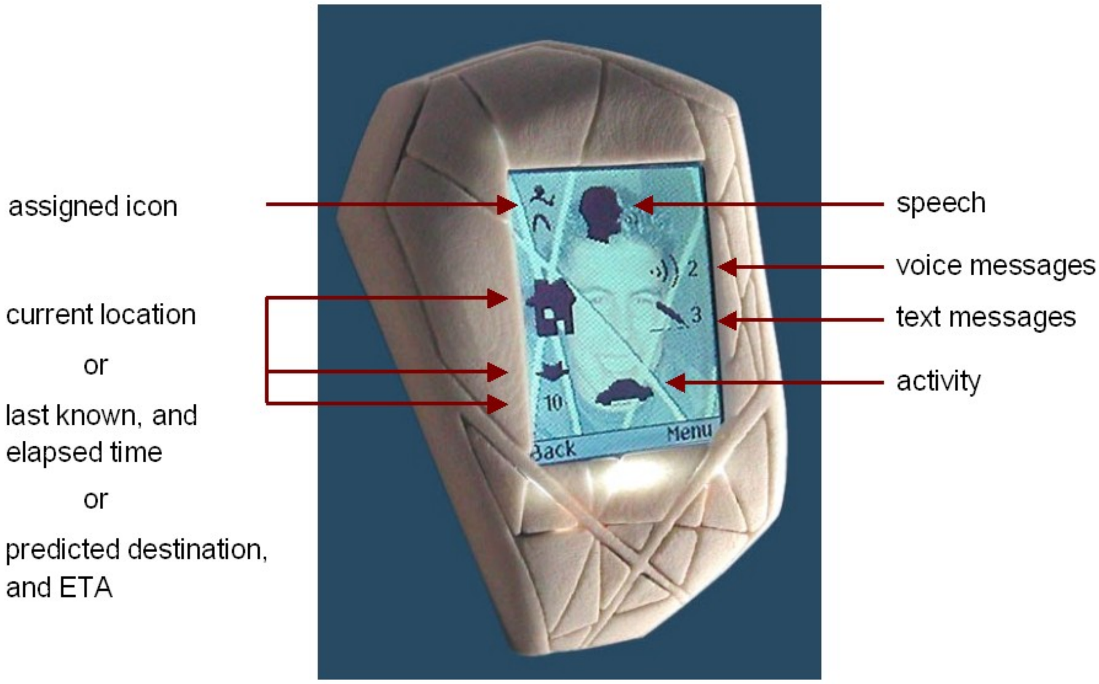
3.1.7 Summary

WatchMe illustrates a number of design criteria defined for collaborative living systems. It does not intend to replace face-to-face communication between insiders, rather support their communication needs when they cannot be together. The same platform supports both the more task-oriented type of communication, used to coordinate and synchronize household activities and members, as well the connectedness-oriented communication, helping to sustain the existing relationships.

WatchMe is not only a wearable computing device, but it is also a “conformable” [Dvorak03] –a device that conforms a body’s contours, attaches to the body in a non-intrusive way, does not impede natural movement, is aware of and utilizes the wearer’s context, imposes minimal cognitive load, and conforms to the users preferences, activities and patterns.

The following is an example of a scenario it could support:

> *Dad knows that Mom has had a hectic day at the office. Her work group has a major upcoming deadline. Through his watch he has been aware that she has been in and out of meetings all day, and therefore concludes that she must be quite stressed. He sends her a text message telling her that he will pick up their son from guitar practice; unknown to her he also plans to get some wine for a relaxing dinner. Mom checks her watch to see the time and notices the text message. This reminds her that she wanted to talk to her daughter. By way of the watch she checks the child’s whereabouts and sees that she is at the mall, presumably with her friends, so decides she will talk to her later. Mom’s action however has automatically caused her face to appear on the daughter’s watch. The daughter notices it and sends her back a picture.*



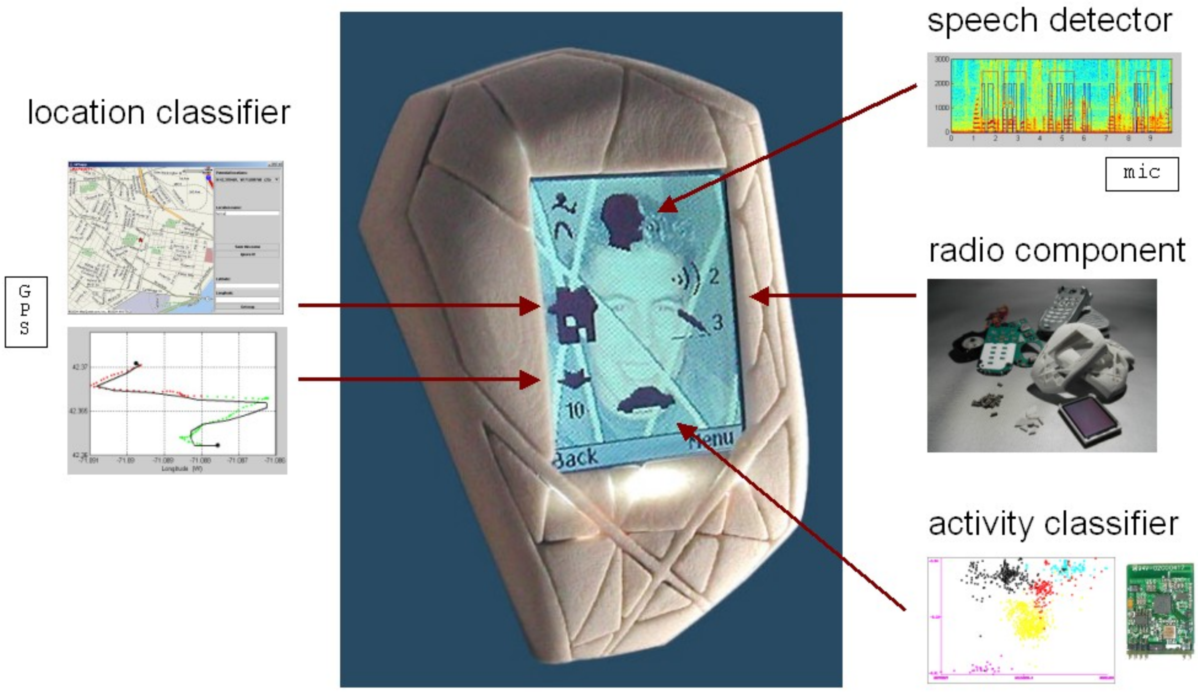


Fig. 3.10 Overview of (a) all the context features displayed on the watch and (b) the system hardware and software components.

3.2 Architecture

Our initial design rationale required that the user interface be easily accessible and frequently visible, which lead to a watch-based design. But to date appropriately sized hardware is not available, nor could we build such tiny phones. Although we see a rapid evolution of phones (display, processing power, size) such that a watch is a reasonable hardware target, we were forced to build the prototype in separate parts.

The *WatchMe* system is made up of several hardware and software components (Figure 3.10): the watch display and user input interface, the radio communication unit, software for the GUI and all the communication functionality, the speech classifier and microphone, the activity classifier and accelerometers, and finally, the GPS unit and route predicting module. These numerous components can be connected in different ways.

One option would be to have everything encompassed in the watch, except for a couple of wireless accelerometers on the body, transmitting to the watch. This would require a radio communication unit with much more computing power than we currently have. In an alternative approach, the watch would be a thin client (basically just the display and buttons for the user input) that communicates via a short range and low power Personal Area Network, such as Bluetooth, to a more powerful radio unit. This unit would have an embedded GPS, microphone, and all the software modules. In these two different configurations, the communication could either be peer-to-peer, or through a server. If transmitting context updates to multiple insiders, a server configuration is preferable, since the data need only be sent once to the server, and from there relayed to the relevant people, as opposed to transmitting multiple copies.

Due to existing hardware constraints, we initially took yet a third approach, a hybrid between the two described above. The plan was to couple the display and communication unit, together with hardware we developed to overcome limitations of the phone, and have this unit communicate via Bluetooth to an iPaq PDA running the various classifiers. Although the Motorola iDEN platform does not yet support Bluetooth, it will soon. Other platforms, such as the Nokia 6600 already include the Bluetooth communication protocol, however, at this point we are too invested in the iDEN hardware to switch platforms. Yet another way to connect the numerous components would be to have them individually communicating through the server.

In its current form, the system is not fully integrated. The watch display, user input interface, radio communication component and all the software for the GUI and communication, reside in a fully functional unit ("the watch"). The speech and activity classifiers are located in an iPaq, receiving input from the built in microphone and wireless accelerometers respectively. The iPaq is currently a standalone functional unit, however it does not have the ability to transmit the classification results to the watch. We have used an iDEN mobile phone to collect GPS data and train the desktop location classifier. Location coordinates sent wirelessly from the phone to the classifier are analysed in real-time, and the classification result is sent to the watch through the server. The iDEN radio unit used in the implemented watch does not include an embedded GPS, however the newer units do, so in the future this sensor would reside in the watch component. In summary, *WatchMe* at present comprises two disconnected parts: the watch with full communication functionality receiving location information from a desktop location classifier, and an iPaq with the speech and activity classifiers and wireless accelerometer sensors.

3.3 Hardware

3.3.1 Watch / phone

In order to build the watch, the display was removed from a Motorola iDEN mobile phone and encased in a shell built using a rapid prototyping 3D printer. This same shell includes the buttons for the user input, and is generally (together with the UI) what we refer to as "the watch". At this point the internals of the phone are not in the watch. The display and buttons are tethered to the base of the phone, i.e. the communication component, via a flat flex cable and thin wires (Figure 3.11). The watch shell also contains a speaker and microphone, for the mini sound-card (see Section 3.3.2).

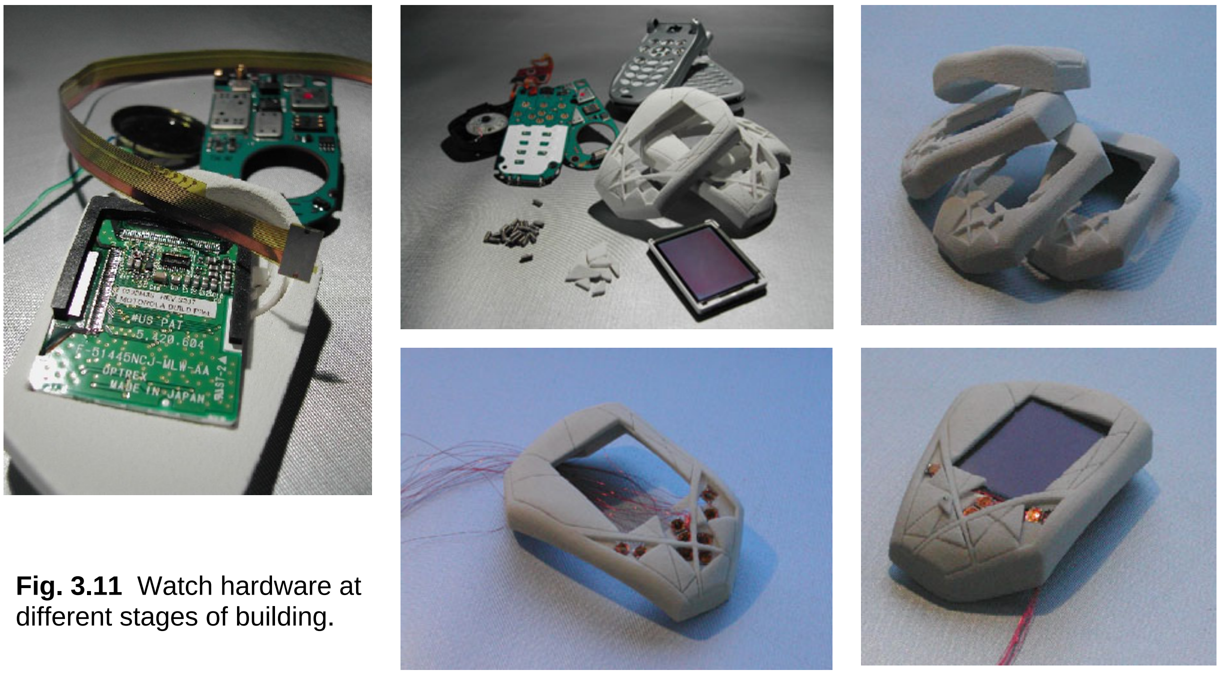
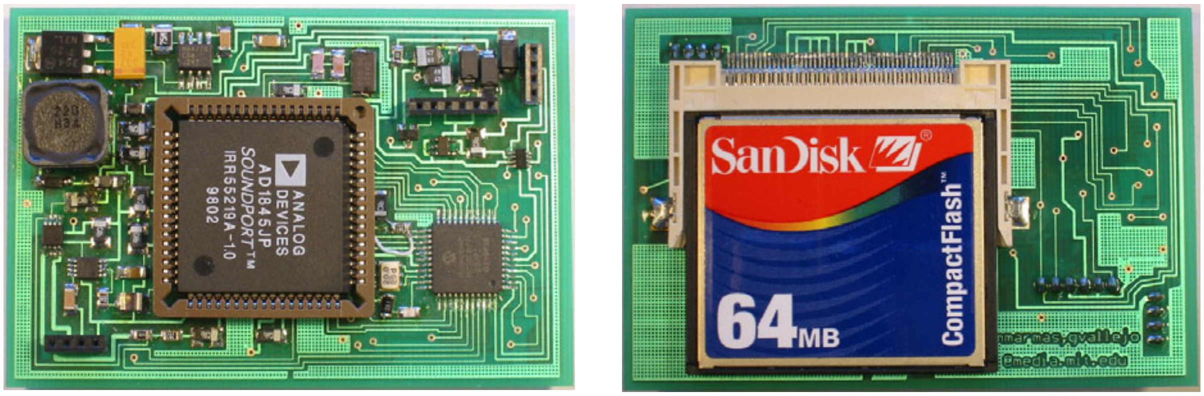


Fig. 3.11 Watch hardware at different stages of building.

The radio component is the base portion of an iDEN phone, i.e. with the display part of the clamshell removed. It is connected to the watch component via a flat flex cable and wires. iDEN is a specialized mobile radio network technology that combines two-way radio, telephone, text messaging and data transmission in one network. It supports an end-to-end TCP/IP connection, the only platform that did so when we initiated this work. Other networks, such as GSM/GPRS, could also support our watch, with a different radio unit. The *WatchMe* system provides text messaging as well as voice messaging, using TCP/IP sockets. It also supports synchronous voice communication, using the ordinary mobile phone telephony functions. In this prototype the phone can be up to 35cms from the watch, limited by the length of the flex cable, so it could be strapped to the user's forearm.

3.3.2 Mini sound-card

When we initiated this project, the Java environment on the cell phones was MIDP1.0 (Mobile Information Device Profile), which did not support audio and multi-media as the enhanced MIDP2.0 does. The only audio formats supported on the iDEN phones were the proprietary VSELP and AMBE formats, and no utilities to convert to or from these formats were readily available. Furthermore, the storage capacity of the phones was very limited. We wanted the ability to store voice Instant Messages, as well as play and store auditory cues. Hence we designed and developed a standalone sound-card (Figure 3.12) that interfaces through a serial port to a mobile phone or PDA. At the time, this sound-card was required to extend the phone capabilities needed for *WatchMe*. It was co-developed with Gerardo Vallejo since we envisioned connecting *WatchMe* and his ListenIn project [Vallejo03]. The idea was to have one of the icons on the watch correspond to the user's home; the context data for the home would be the output from ListenIn, providing auditory awareness of the home environment.



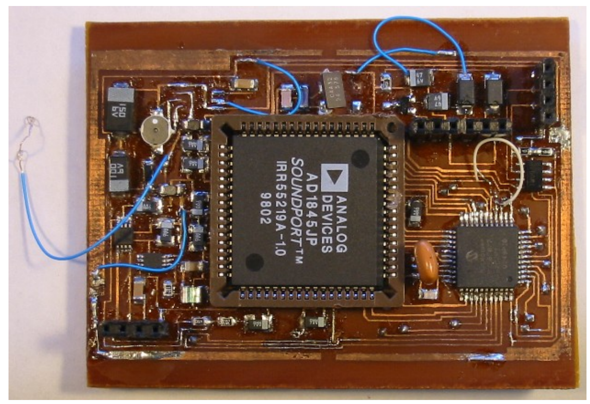


Fig. 3.12 Mini sound-card. The bottom right corner shows the home made version (etched on copper using lithium chloride) used for debugging. The front and back of the final version is shown above.

The mini sound-card was designed around the AD1845 parallel-port 16-bit stereo codec. This codec includes stereo audio converters, on-chip filtering, sample rates of 4-50 kHz, and analog and digital mixing. The audio formats supported are 8- or 16-bit PCM, and 8-bit μ-law or A-law. It is a standalone sound-card with its own file system on a compact flash, receiving commands for audio capture and playback through the serial port. It is powered from an external +3.6V lithium ion battery, or from a phone that can provide 600mA. The size of the board could be reduced by using a smaller package of the same codec, the limiting factor however is the size of the compact flash card.

With the improved audio capabilities of the newer phones, this sound-card is no longer essential for *WatchMe*. We find it very encouraging that mobile phone hardware is improving at the rate that it is.

3.3.3 Accelerometers

In order to collect data and classify human activity in real-life situations, a portable sensing infrastructure is necessary. Motivated researchers, or subjects in experiments, will often bear the burden of cumbersome hardware. However in applications such as *WatchMe*, envisioned to eventually be used in everyday life, the hardware must be as inconspicuous and unobtrusive as possible. The wearer should not have to think of extra hardware he needs to carry around. It should be as simple as putting on your watch in the morning, tying your shoes, and making sure your wallet is in your pocket –the watch, shoes and wallet, or other pieces of clothing, might have embedded accelerometers. We wanted sensors that were small enough to be embedded or attached to clothing, and that would not constrain the wearer's movement in any way. Additionally, we wanted a configuration that would enable real-time synchronized data from multiple sensors, on different parts of the body. Furthermore, we required sensors that could run continuously for a day without battery replacement. None of the existing available sensing modules (such as Hoarders, UCB motes, Smart-Its) are adequate for this purpose. Therefore, we contributed to the development of a new generation of portable wireless sensors [Munguia03].

To the best of our knowledge, these are currently the smallest (1.2 x 1.0 x 0.25 in), lightest (8.1 grams with battery) and least expensive wireless 3-axis accelerometer sensors available (Figure 3.13). A single prototype unit costs $41 and its average battery life, sampled continuously at 200Hz, is 20.5 hours.

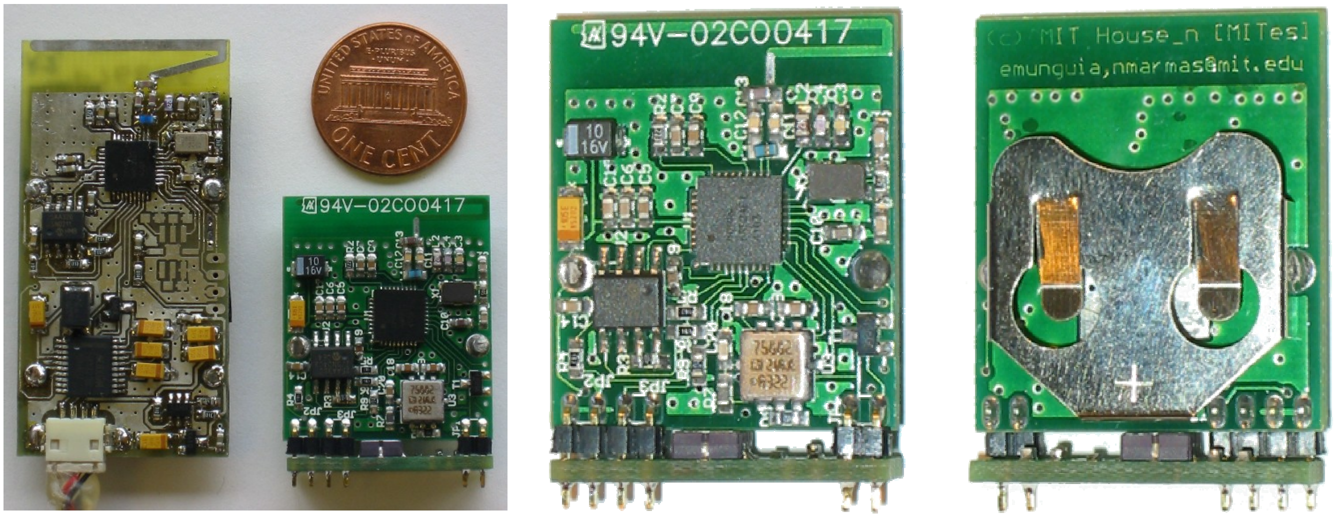


Fig. 3.13 Accelerometers. The receiver (left) is connected to the serial port of the iPaq. Wireless transmitters (middle and right images show front and back) can be placed on different locations of the body. Dimension: 1.2 x 1.0 x 0.25 in. Weight: 8.1 grams with battery.

The wireless sensors are designed around the 2.4GHz Nordic nRF24E1 transceiver, which includes an 8051 compatible micro-controller running at 16MHz, an analog-to-digital converter (ADC), universal asynchronous receiver-transmitter (UART) and serial peripheral interface (SPI), pulse width modulation (PWM), timers, and I/O pins. The transceiver operates in the 2.4GHz ISM band, which is available without a license, and uses a proprietary protocol, which does not interfere with 802.11 or Bluetooth. The Nordic transceiver provides data rates up to 1Mbps and 125 Tx/Rx channels for multi-channel communication. The sensor boards include a 4K EEPROM program memory, ±2g or ±10g ADXL202/210 accelerometers, and a 50Ω antenna matching circuit between the transceiver and onboard micro-strip 3cm antenna. Attached to the main board is an additional accelerometer on a side daughter board, providing the third axis of acceleration. The sensors are powered by a single 3V, CR2032, coin battery.

The receiver board can interface with the RS232 serial port of any PC or PDA. In our configuration it is connected to an iPaq. It includes the same circuitry as the wearable wireless sensors, as well as an RS232 level converter (MAX3218) and a voltage regulator (MAX8880) for external power supplies of +3.5V to +12V, and consumes an average of 28mA. It has room for an accelerometer, as well as a side daughter board, so can also provide 2 or 3 axes of acceleration. The receiver board currently can receive acceleration data from up to six wireless sensor boards simultaneously, however, it could be modified to receive from up to 125 boards.

As a result of their dimension, weight, cost, and power consumption, they could be used in real-life. We foresee their use in numerous applications, in different domains.

The following chapter describes the classification algorithms. The outcome of each classifier (a text label such as “biking”, “speech”, or “left home, 10”) would be sent to the watch and the corresponding icons updated for each insider.

4. Sensors to sense

A person's external context is obtained by gathering and classifying sensor data. We find that this data must be abstracted in order for it to be meaningful at a glance within one's peripheral vision; more effort is required to interpret raw sensor data, and furthermore location coordinates, or G values of acceleration, are not very significant to most people. As shown in the previous chapter, the context in WatchMe is represented by icons displayed on the watch. This chapter describes the classifiers that determine which icons are shown.

4.1 Location learning

The watch displays a remote insider's current location, or his next predicted one and expected time of arrival (ETA), or his last known location and the time elapsed since departure (Figure 4.1). This involves three different components: the learning of recurrently visited locations, or endpoint detection; the identification of routes between these endpoints (sources and destinations); and inferring the currently travelled route and predicting ETA.

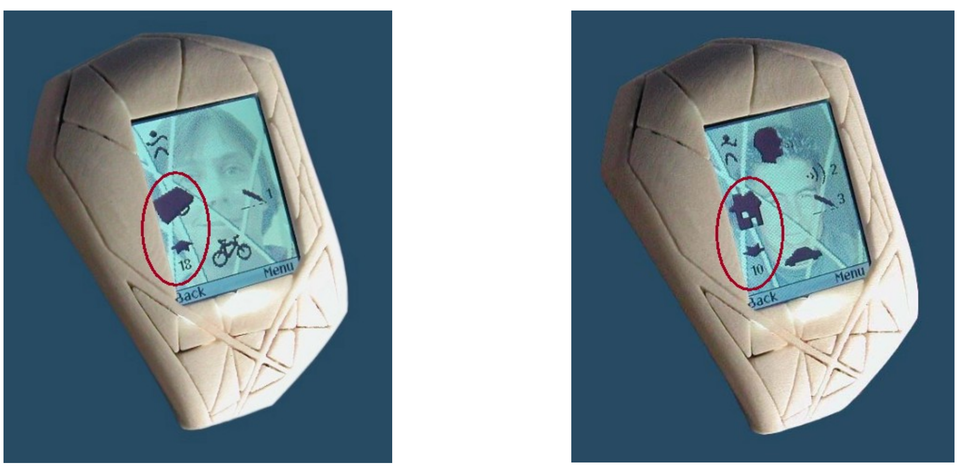


Fig. 4.1 Examples of the context screen of the user interface, highlighting how current location, expected time to arrival, or time elapsed since departure, are graphically represented. (a) heading to work and should be there in 18 minutes, (b) left home 10 minutes ago.

Endpoint detection

Location coordinates, from an embedded GPS, are collected throughout the day on the watch/phone and uploaded to a desktop computer. The GPS data is analysed for personal landmarks, i.e. locations recurrently frequented, which are then presented to the user to be labeled (Figure 4.2). The graphical user interface shows the point on a map, and permits the user to name and save the location, or indicate that it is irrelevant and should be ignored. This *WatchMe* application dynamically downloads the maps from MapQuest. Named locations are considered endpoints of routes. Once a location has been marked as irrelevant, it –and any point within a given radius (default 40 m)– will not be considered again. An irrelevant location is for example a bus stop, or train platform. The user can define how many times he must be seen within a radius of a site in order for it to be considered as potentially interesting.

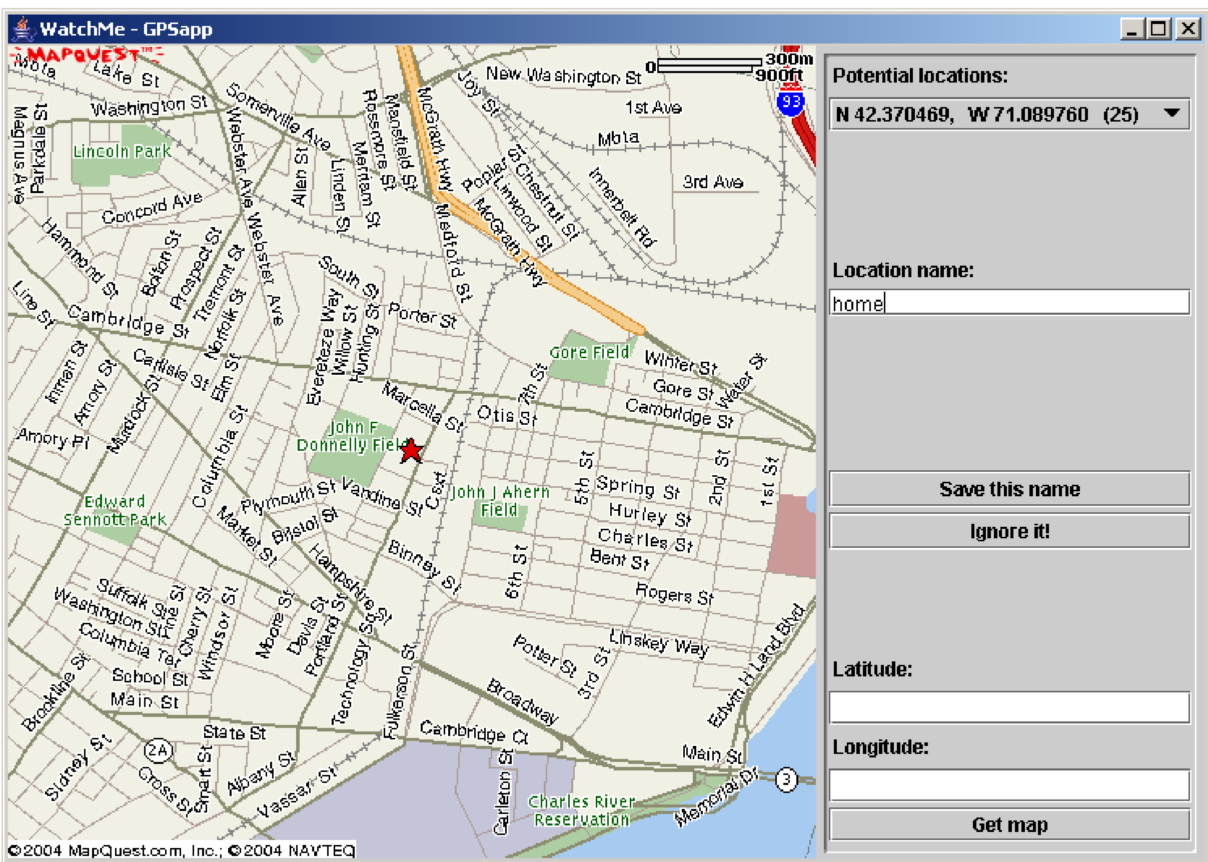


Fig. 4.2 Interface through which a user sees locations marked as potentially relevant. The locations can be named and saved, or marked as irrelevant.

The system can mark a location as potentially relevant for a number of reasons:

- **static fix:** if the GPS fix is within a given radius for over 10 minutes (user-defined value), the location should be considered. In previous work [Marmasse02] we had counted on the fact that GPS signal was lost in most buildings. With Assisted GPS (AGPS) this is often no longer true. A standalone GPS receiver acquires a position by searching for satellite signals and decoding their messages. This requires strong signals, line-of-sight to at least 3-4 satellites in an adequate constellation, and time to search for the different satellites and calculate a position. The GPS units in mobile phones are assisted by the cellular network. Since the network already roughly knows the phone's location (from cell triangulation), it can assist the unit by pre-computing the doppler shift of the satellite signal, making it easier to find. Additionally, the base station can send the phone some of the coefficients needed for the positioning calculation.

- **breaks in time or distance:** the system searches the GPS track and analyses it for breaks in time, or breaks in distance. The following example shows a break in time (indicated in the GPS data just for clarity). The user arrived home and did not get a GPS fix until leaving the following day.

```
4/21/2004,  2:10:45, Lat: N 42.370416, Lon: W 71.089723,
4/21/2004,  2:10:50, Lat: N 42.370416, Lon: W 71.089723,
4/21/2004,  2:13:54, Lat: N 42.370363, Lon: W 71.089701,
---
4/22/2004,  12:18:28, Lat: N 42.370523, Lon: W 71.089701,
4/22/2004,  12:18:32, Lat: N 42.370568, Lon: W 71.089701,
4/22/2004,  12:18:37, Lat: N 42.370555, Lon: W 71.089701,
```

  The location saved is always the one before the break, providing there was a "clean" GPS track up until that point. A track is considered no longer clean if the reported accuracy of the GPS fix is above a threshold, or if the distance between the consecutive points is incompatible with the speed the user is progressing at. In the case of a clean track, followed by a sequence of low accuracy readings, and then a break, the saved location is the fix before the sequence of bad readings. Although GPS data is much more reliable since Selective Availability (intentionally transmitted random errors) was turned off in March 2000, fluctuations still occur, especially in certain areas. Figure 4.4 shows the data recorded from a static AGPS receiver, over the course of four hours.

The following example shows a break in time and distance. The user entered a train station (marked as 1 on Figure 4.3) and exited in a different location. The location saved is the train station entered. The location after the break, i.e. the station where exited, is not considered since, even with AGPS, the receiver often takes time to acquire a fix and, depending on the speed of travel, the user may be far from a potential personal landmark when the unit manages to lock on a position.

```
4/20/2004,  13:3:58, Lat: N 42.362293, Lon: W 71.085963,
4/20/2004,  13:4:1, Lat: N 42.362328, Lon: W 71.085963,
4/20/2004,  13:4:4, Lat: N 42.362357, Lon: W 71.085947,
4/20/2004,  13:7:8, Lat: N 42.362400, Lon: W 71.085920,
---
4/20/2004,  13:20:30, Lat: N 42.374053, Lon: W 71.118752,
4/20/2004,  13:20:34, Lat: N 42.374008, Lon: W 71.118731,
4/20/2004,  13:20:37, Lat: N 42.374008, Lon: W 71.118731,
```

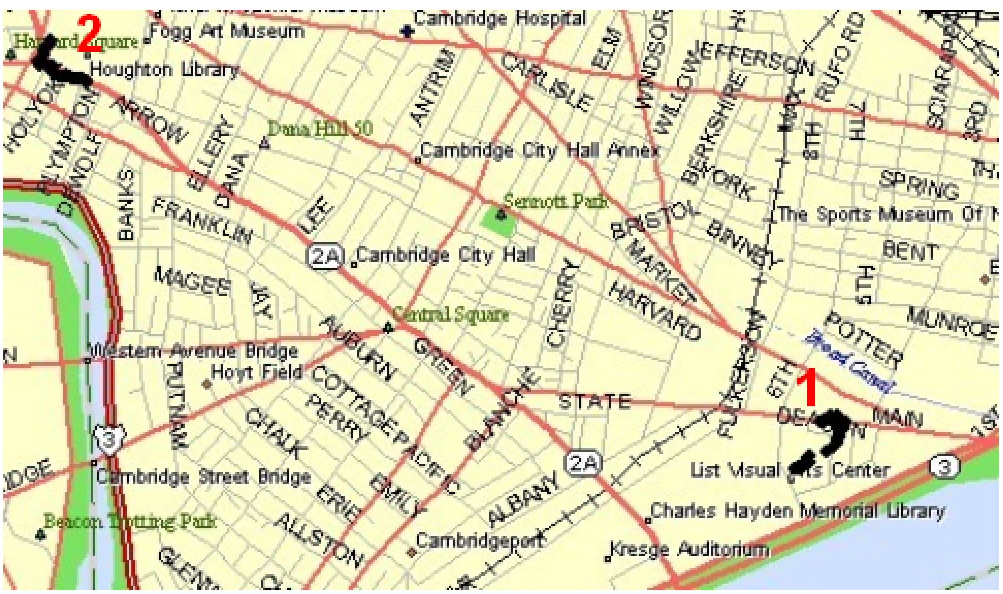


Fig. 4.3 Mapping of GPS data showing the break in the track where the user entered a train station (1) and exited at location (2).

- **low accuracy:** several consecutive fixes (>5), within a radius (50 m), with reported low accuracies is considered a potentially relevant location. Despite AGPS, in some buildings no fix is acquired, whereas in others the receiver gets a position albeit with low reported accuracy.

- **manually saved:** the user has the option of pressing a button and saving the current location. This might be used to save locations not often frequented but considered important enough to mark, or alternatively to save locations which the system does not manage to pick out. For instance, an outdoor location often visited but only for a very short time would not be identified, since it could just as well be a long traffic light. A specific example is a user who frequently buys lunch from a food truck, and would like others to know when she is there, or on her way. The location is outside and the person is typically there for only a couple of minutes each time.

Fig. 4.4 Mapping of the GPS fixes recorded over the course of four hours from a static assisted GPS receiver.

Route identification

Routes are defined as trajectories between named locations. The GPS track between the two endpoints is purged of fixes with low accuracy, and saved as an instance of the route. Positions with low accuracies are ignored under the assumption that they are random; if there is a pattern, and they are not simply noise, they will be identified over time and named or marked as irrelevant. A canonical route representative, or template, is calculated from the instances of a given route (Figure 4.5). This involves dividing the route into chunks of 100 metres of traveled distance and finding an average latitude and longitude for each chunk. The average elapsed time from the source, and average speed, are also included in the template. The individual instances are compared, before composing a route representative, to ensure that they are indeed different instances of the same route. An instance of a route that varies more than one standard deviation from the average of all instances of that specific route is automatically excluded from the template. Since a person could travel from point A to B on different routes, multiple templates corresponding to different trajectories between the same locations can be created. Routes A to B, and B to A, are considered to be two different routes.

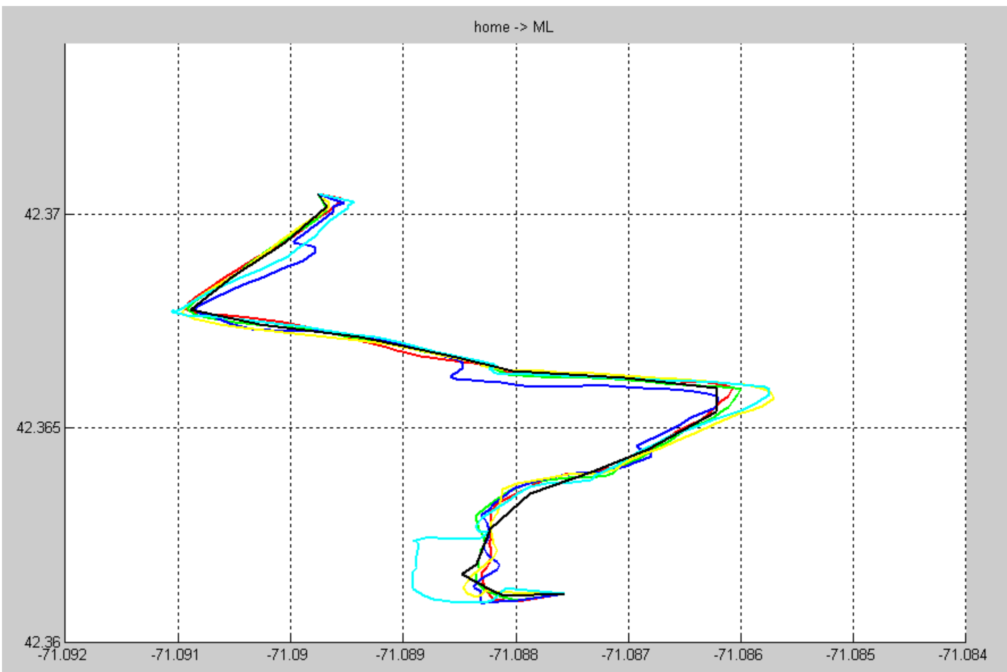


Fig. 4.5 Canonical route representative (black) generated from different instances of the same route.

Route prediction

The canonical route representations are used to predict where the user is going and estimate the time of arrival. Initially the system will indicate how long ago the user left a known place, until it has enough information to predict the destination. The route the user is currently traveling is identified by trying to align it to all the templates. The alignment is a piecewise comparison of the route segments, generating a distance error for each segment, averaged into a total route error. The predicted route is the one with the smallest error, below a threshold (150 m). The route templates provide a means to predict the traveled route, and based on the average time required to reach the current segment, predict an estimated time of arrival. If the user's travel time to the current segment is significantly different than that of the template, the ETA will be proportionally adjusted. For example, if a user normally bikes to a certain destination, however one day decides to walk, assuming the path traveled is similar, the ETA will be adjusted to reflect the slower pace.

If the full traveled route does not align well with any of the canonical routes, the system tries to align the last portion (500 m) of the trajectory. Figure 4.6 shows the trajectory traveled by a user from *Home* to *Westgate*. The red points indicate the part of the trajectory traveled so far, the green points are the remainder of the complete path, and the black path is the canonical route the trajectory best aligns to. When the user leaves *Home* she is predicted to be heading to the *Lab*, as her full traveled path aligns well to the canonical HL route. As she proceeds and does not turn to the *Lab*, the system can no longer predict the destination, all it knows is the time elapsed since she left *Home*. As she progresses on her journey, the system manages to align the last portion of her path to the LW canonical route and predicts she is going to *Westgate* (as she actually was). However, as it was a beautiful day, the user deviated and went for a bike ride along the river before eventually going to *Westgate*.

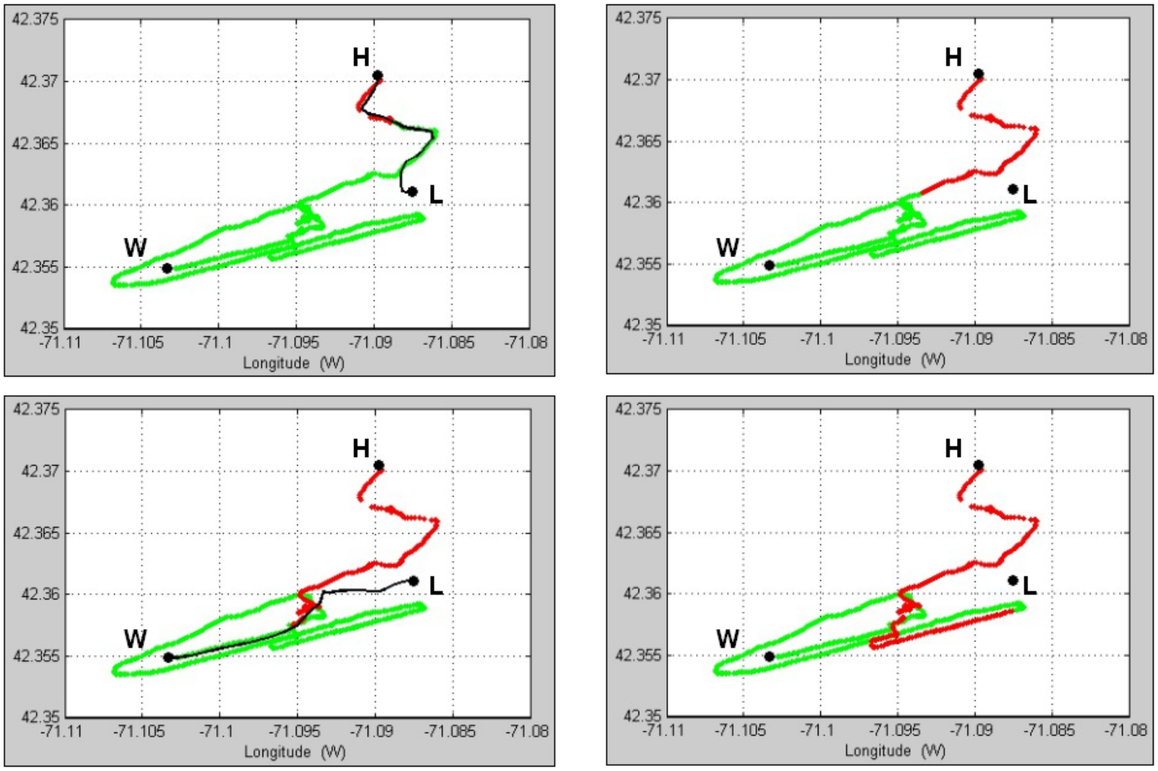


Fig. 4.6 Examples showing how the system infers the destination by aligning the full traveled trajectory where possible, or alternatively the last traveled portion, to the canonical routes. Personal landmarks are indicated by letters (*Home*; *Lab*; *Westgate*). The path traveled so far is red, the remainder is green, and the canonical route it aligns to is black.

Since routes are defined as trajectories between named locations, if a user always passes point B on his way between A and C, when traveling in that direction he will always be predicted as heading to B although in some cases he might actually be going to C. Once he has passed location B, the prediction will be reassessed. Hence, the system is always predicting the current leg of the journey and the ETA to the next known location, even if this is an intermediary place on the way to a final destination.

Currently all the location learning and prediction is done on a desktop computer. However, the route templates could reside on an iPaq. With a Bluetooth connection between the iPaq and watch, the latter would send the location coordinates to the iPaq for analysis, and a real-time classification would be sent back to the watch. The classification outcome would be relayed over the cellular network to the wearer's insiders, and their watch graphical interfaces updated to reflect the new context. See discussion on architecture (Section 3.2).

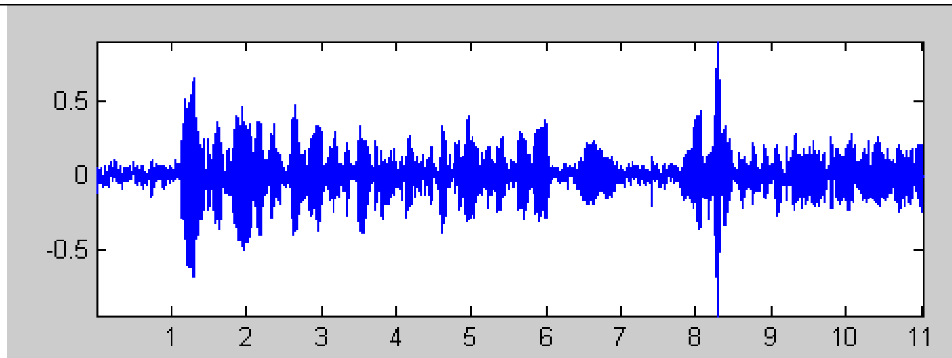
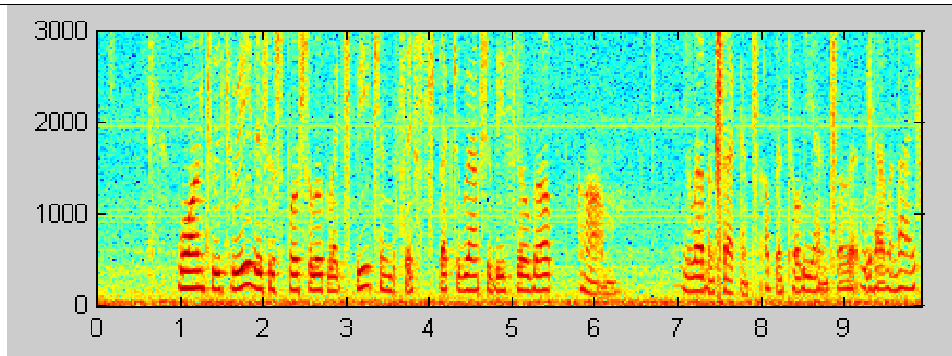
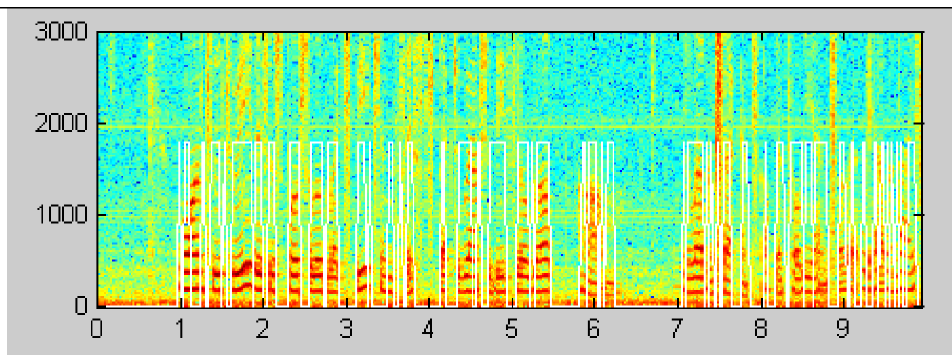
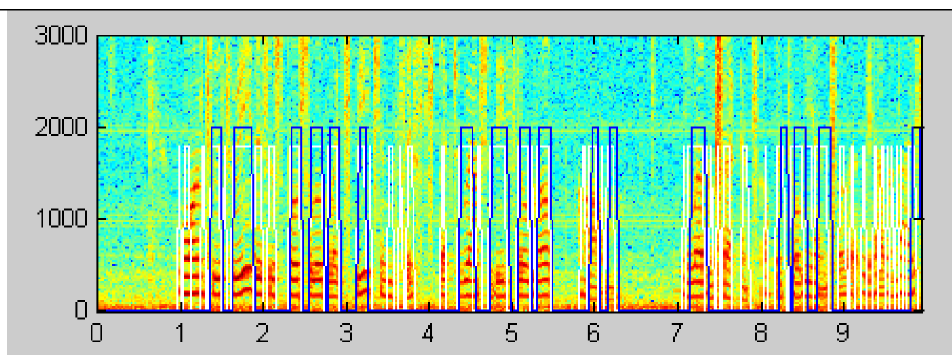
4.2 Speech detection

Speech is indicative of social engagement. Hudson *et al.* [HudsonS03], in a Wizard of Oz setup, simulated sensor data to assess which sensors where the most useful in predicting interruptibility. They found that the ability to detect that someone was speaking was the most promising sensor for this problem; it provided an accuracy of 76% in their inference models.

Fig. 4.7 Example of the context screen (left), highlighting how the detection of speech is represented on the graphical interface. Screen shot of the speech detector running on the iPaq (right).

Because the presence of speech is such a salient indicator of unavailability, we developed a speech detector to run on the iPaq. Figure 4.7 shows how the detection of speech, in the close proximity of the specific remote person, is displayed on the graphical user interface.

The speech detector is a binary discriminator, which analyses the temporal patterns of the voiced segments in the audio signal, and determines whether they correspond to speech. The algorithm is illustrated in Figure 4.8.

record audio 16-bit, mono, 11 kHz	
calculate audio spectrogram band: 0-3000 Hz	
determine voiced chunks based on pitch analysis (white)	
only consider chunks the duration of which is above a minimum threshold (blue)	

cluster voiced and unvoiced chunks into speech blocks (red)

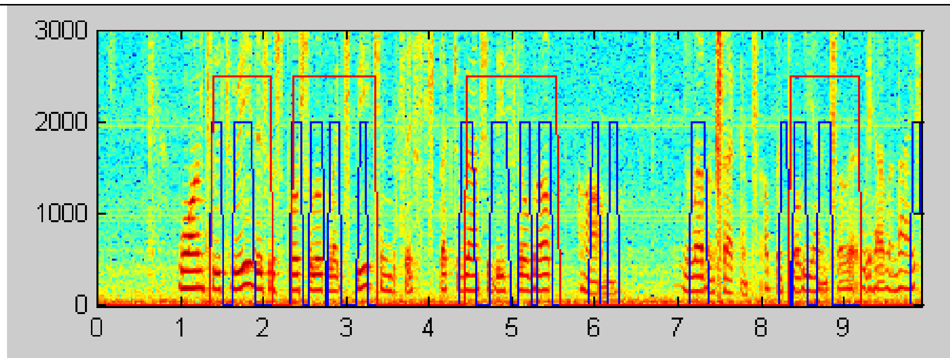


Fig. 4.8 Speech detector algorithm, illustrated step-by-step.

The speech detector records ten seconds of 11 kHz, 16-bit, mono audio. It analyses the digital signal, generates a speech/non-speech classification, and starts again; no audio is recorded during the processing, and none of the audio is stored. A common method for speech detection is the use of formants [Rabiner93], i.e. harmonics found in vowels resulting from resonance in the vocal tract. Our algorithm identifies voiced regions, or chunks, by finding the three largest amplitude peaks, multiples of the pitch (F0), and analysing the pitch structure. An energy threshold –similar to that used by Rabiner and Sambur [Rabiner75] however in the frequency domain– is used to eliminate some of the ambient noise. Only voiced chunks larger than 80 ms are considered; this corresponds to approximately two spoken syllables. The voiced chunks are then grouped into utterances, or speech blocks. In speech there are typically fast transitions between the voiced and the non-voiced chunks, therefore any non-voiced segment above 200 ms is no longer considered part of the speech block. This detector does not aim to exactly determine the speech boundaries, rather merely discriminate whether speech exists in the signal. If two speech blocks are found in the ten second signal, the audio is classified as speech.

4.3 Activity recognition

We believe that the sharing of activity information can help others infer our availability, if they are trying to contact us, or generally provide a level of awareness, if they simply want to know what we are up to. How much information is revealed is a personal choice. Although it is possible to recognize numerous fine-grain physical activities [Foerster99, Bao03], we have focused on ambulatory ones, believing that these are the least common denominator of what most people would be willing to share. Many people may feel that it is sufficient to disclose the fact that they are home and walking around. Others may be happy to share with the family that they are raking the yard (perhaps the children will hurry home and play in the piles of leaves, providing quality family time). Few may feel the need to reveal that they are brushing their teeth.

Our hardware infrastructure can support the different granularities of activity, provided the user is willing to wear multiple small on-body accelerometers –these could be embedded in clothing. The software algorithm used can also support the numerous classes. We have trained our implementation on ambulatory activities, however this training data could be extended. Although we have not done so, it would be possible to build a software interface which enabled the user to choose the subset of activities he was comfortable sharing. The relevant training data would then be selected and a decision tree built from it.

Different modes of locomotion have implications on availability and appropriate communication channel. A person can talk on the phone while walking, or engage in voice IM exchanges, but cannot easily participate in a text messaging interaction. There are also implications on communication privacy. For instance, text conversations may be preferable when on the train, as they cannot be overheard, and furthermore do not impose on others. A user’s current activity is displayed on the watch interface, as shown in Figure 4.9.

Fig. 4.9 Examples of the context screen of the user interface, highlighting how the outcome of the activity classification is represented.

The use of accelerometers to detect and classify physical activity is certainly not new; prior research in this domain includes [Bouten97, Mantyjarvi01, Kern03] and many more. However the first study that examined the performance of different recognition algorithms using multiple wire-free accelerometers, on several subjects and various activities in semi-naturalistic settings, was recently done by [Bao03]. In their setup they were using five Hoarder boards [Gerasimov01] that stored acceleration data on a Compact Flash. The Hoarders are unsynchronized and therefore had to be manually/artificially shaken together in a pattern that could later be recognized and enabled data alignment of the different boards – the alternative would have been to wire them together. Their findings show that, of the various algorithms tested, the Decision Tree performed the best, with an overall accuracy of 84%. This is consistent with the conclusions of Hudson *et al*. [HudsonS03], who found Decision Trees to have the highest level of accuracy in their various inference models.

We have contributed to the development of the next generation of wireless sensors [Munguia03], concentrating our efforts in building more appropriate, lightweight and inexpensive hardware in a small enough package that could fit in a watch, or could be attached or embedded to clothing. This new sensor architecture supports multiple transmitters talking to the same receiver, enabling us to synchronize the data from the different sources.

Furthermore, by connecting the receiver to the serial port of an iPaq, we can perform real-time activity classification. The activity recognizer used in *WatchMe* is a variation on [Bao03]. It also uses a Decision Tree [Quinlan93], however only classifies a subset of activities chosen for this application.

The activity recognition is based on acceleration data collected from two wireless sensors (see hardware Section 3.3.3) each with two orthogonal biaxial accelerometers, providing three axes of acceleration. These wireless transmitters are placed on two body locations (hip and ankle) from where they transmit to a receiver connected to the serial port of an iPaq. The classifier was trained and tested with two accelerometers, however for everyday use, classification results are accurate enough if just one of them is used. The preferable location is the ankle, since ambulatory activities, such as walking, running, and biking, are better detected by leg movements. However, as has been shown by [Lee02, Bao3], accuracies of 90-96% for ambulation can be obtained from acceleration data taken from the thigh or hip.

Classification features

The features extracted from the acceleration data are: mean, energy, and entropy for each axis of acceleration. This is a total of nine features for each board, or body location. The mean is the mean acceleration over a time window, giving the average acceleration of the movement. The energy is calculated by taking the sum of the squared discrete FFT magnitudes and dividing by the window length; it is a measure of the intensity of the movement. The frequency-domain entropy is the information entropy of the discrete FFT magnitudes, divided by the window length. It can help discriminate between signals with similar energy, especially when using just one accelerometer on the hip. The features are calculated over 50% overlapping sliding windows, each window length being 512 signal samples. Figures 4.10 and 4.11 show the different features for some of the classes, from wireless accelerometers placed on the ankle and hip.

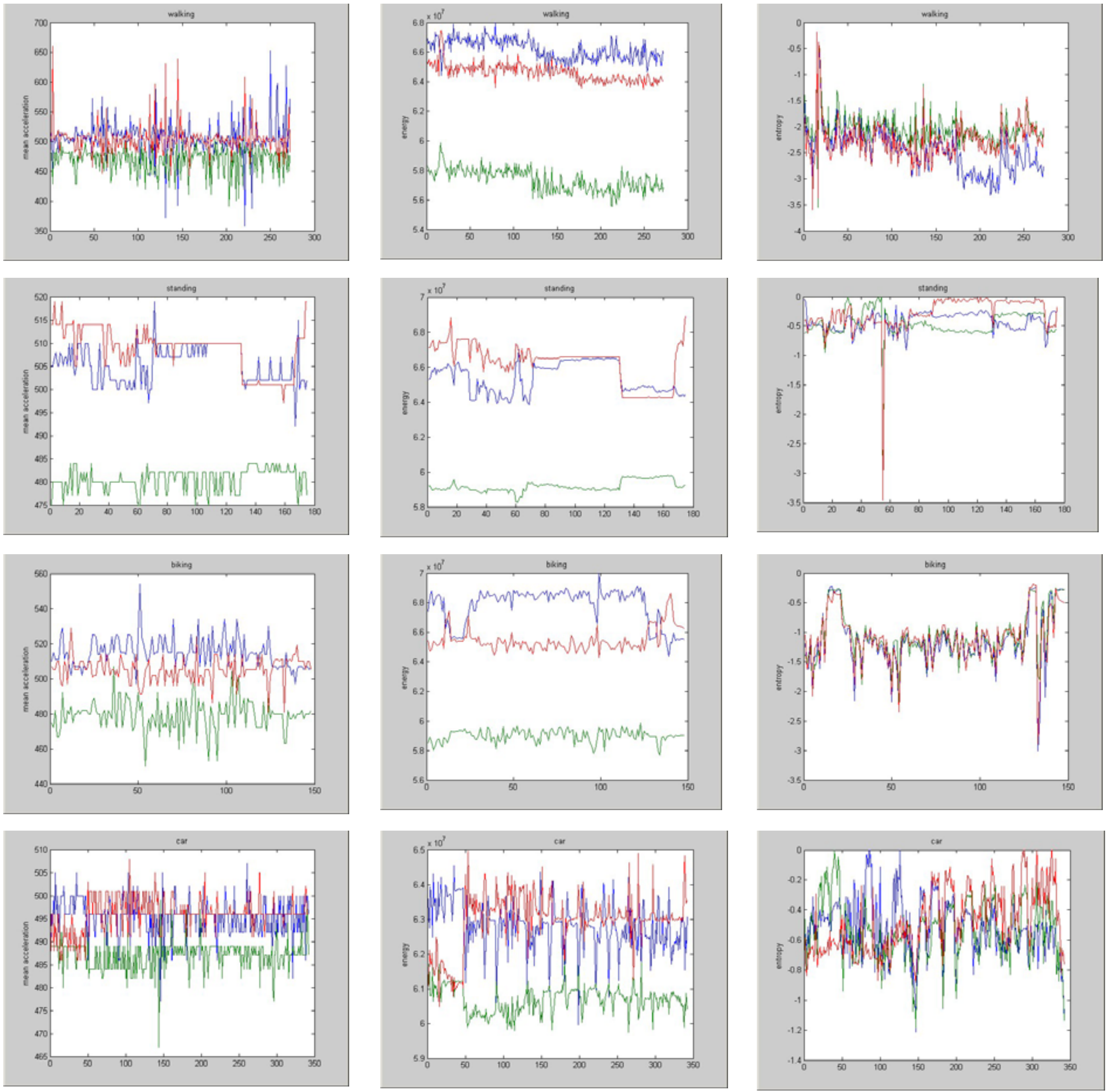


Fig. 4.10 Mean, energy and entropy acceleration features, per class, from a ±10g accelerometer placed on the ankle. The Y-axis of the graph corresponds to the value of the feature, while the X-axis corresponds to the window number. The colours correspond to different axis of acceleration: X-axis = green, Y-axis = blue, Z-axis = red.

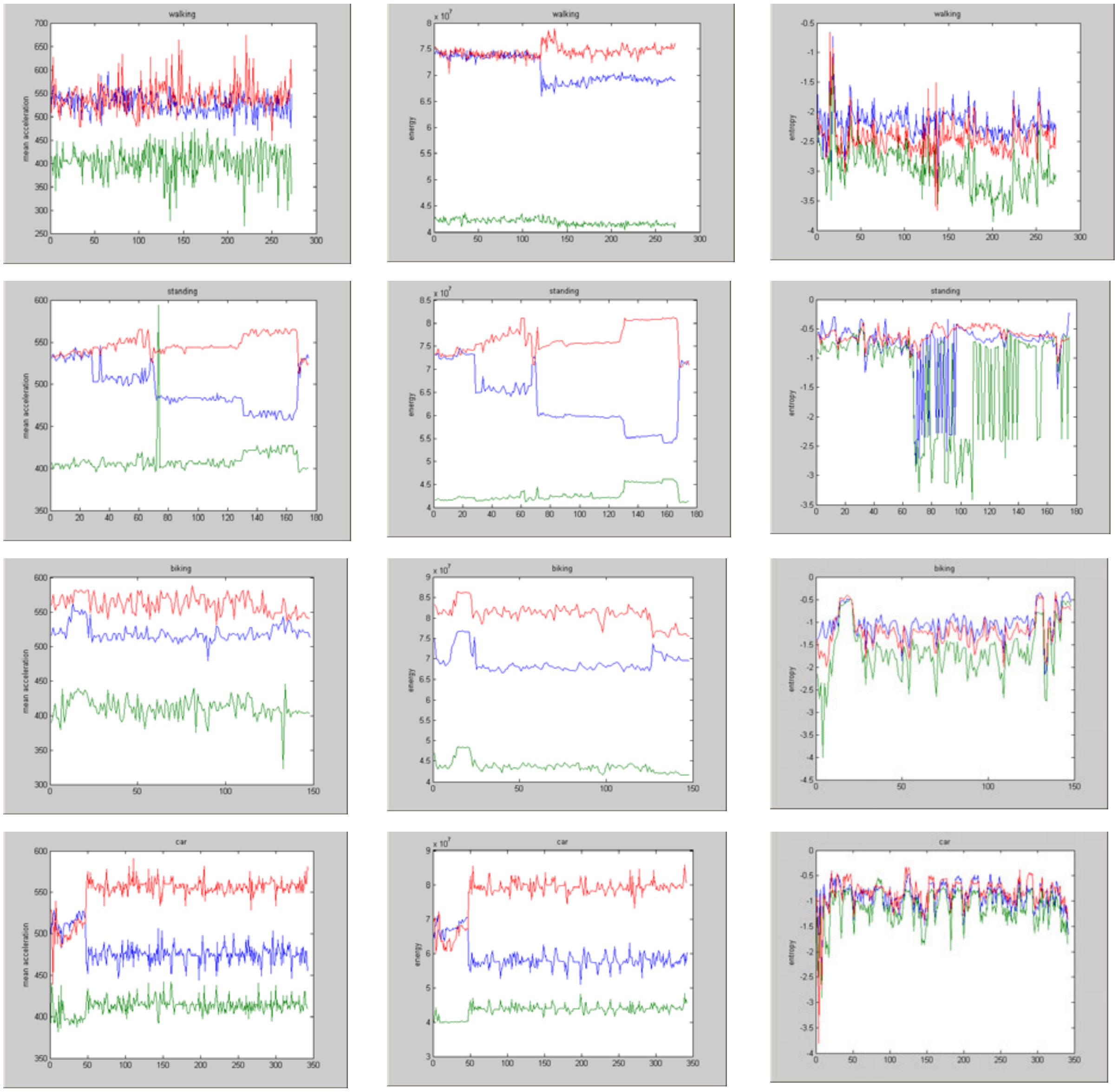


Fig. 4.11 Mean, energy and entropy acceleration features, per class, from a ±2g accelerometer placed on the hip. The Y-axis of the graph corresponds to the value of the feature, the X- axis corresponds to the window number. The colours correspond to different axis of acceleration: X-axis = green, Y-axis = blue, Z-axis = red.

Decision tree

The Decision Tree learning method is one of the most widely used and practical techniques for inductive inference [Winston92, Mitchell97]. The constructed tree, using the C4.5

algorithm [Quinlan93], first performs a binary split on the most salient feature (e.g. the X-axis acceleration energy from the sensor on the ankle), dividing into two branches. It then recursively constructs trees for each branch. The predictive values (e.g. walking, biking, etc.) are assigned to the resulting leaves. To avoid overfitting to the data –which occurs after many tree subdivisions since each leaf then represents only a small number of samples– the tree is pruned. Decision trees are robust to errors in classification of the training samples, as well as to errors in the attribute values of the samples. They are computationally efficient, and their performance is suitable for real-time recognition. Figure 4.12 illustrates clustering of the classes using different classification features; these were generated using the Weka machine learning explorer [Witten99].

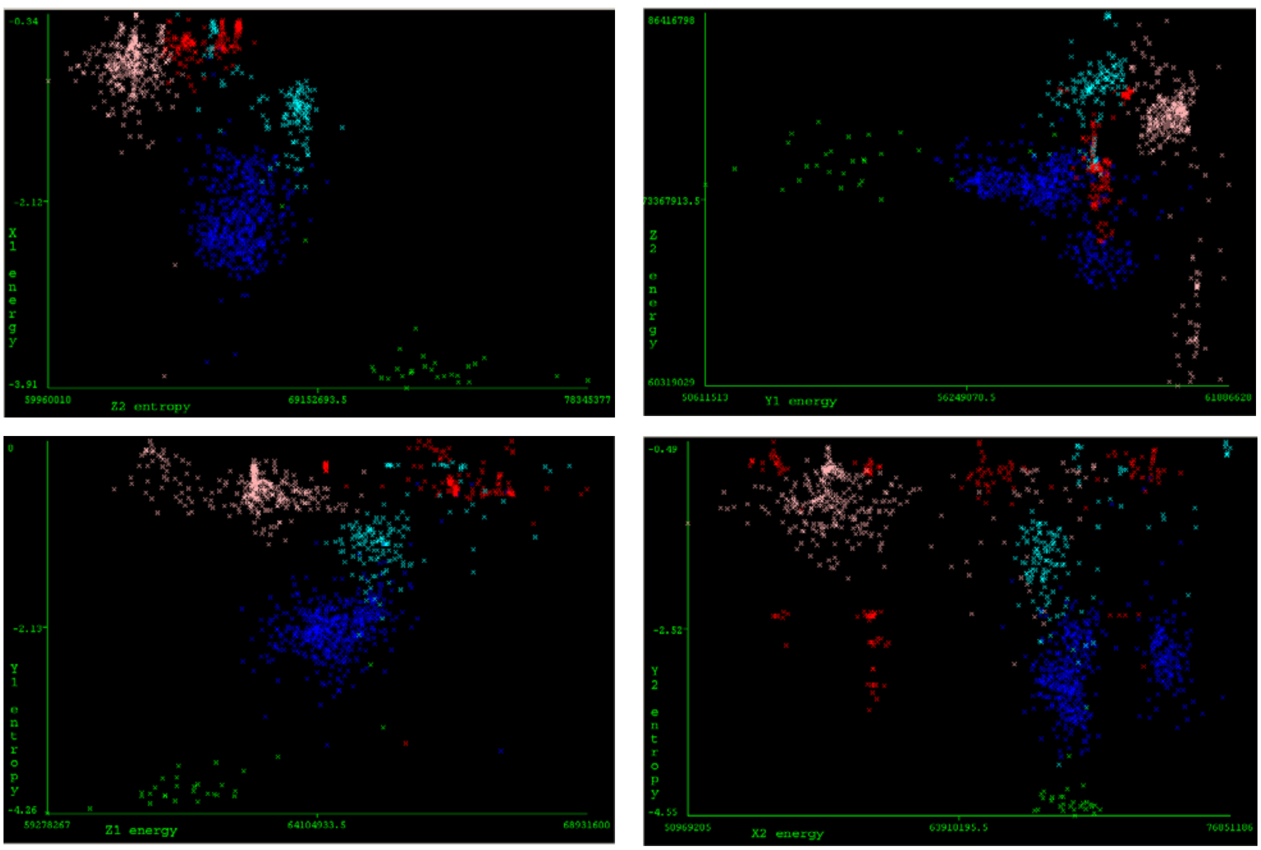


Fig. 4.12 Class clustering using different classification features. X1, Y1 and Z1 are the acceleration axes of the sensor placed on the ankle. X2, Y2 and Z2 correspond to the sensor on the hip. The class colours are: walking = blue, standing = red, running = green, biking = cyan, driving = pink.

The pruned decision tree resulting from the training data is shown in Figure 4.13. The values in brackets at each leaf are the instances correctly classified, or in the case of two values, those correctly classified and the number of incorrectly classified instances.

```
x1-energy <= 64359810
|   z1-energy <= 65320214: car (341.0)
|   z1-energy > 65320214: standing (23.0)
x1-energy > 64359810
|   x1-entropy <= -1.632449
|   |   y1-entropy <= -3.614927: running (27.0)
|   |   y1-entropy > -3.614927
|   |   |   z2-energy <= 76401397: walking (449.0)
|   |   |   z2-energy > 76401397
|   |   |   |   z2-entropy <= -1.956728
|   |   |   |   |   y2-energy <= 44218492: walking (21.0/1.0)
|   |   |   |   |   y2-energy > 44218492: running (2.0)
|   |   |   |   z2-entropy > -1.956728: biking (10.0)
|   x1-entropy > -1.632449
|   |   x1-energy <= 67082559
|   |   |   x1-entropy <= -0.926035
|   |   |   |   x1-entropy <= -1.037409: walking (5.0/1.0)
|   |   |   |   x1-entropy > -1.037409: biking (3.0)
|   |   |   x1-entropy > -0.926035
|   |   |   |   x2-energy <= 68817068: standing (117.0)
|   |   |   |   x2-energy > 68817068
|   |   |   |   |   z2-energy <= 75034414
|   |   |   |   |   |   z1-energy <= 65968799: walking (2.0/1.0)
|   |   |   |   |   |   z1-energy > 65968799
|   |   |   |   |   |   |   y2-entropy <= -0.887443
|   |   |   |   |   |   |   |   z1-entropy <= -0.380332: standing (8.0/1.0)
|   |   |   |   |   |   |   |   z1-entropy > -0.380332: walking (2.0)
|   |   |   |   |   |   |   y2-entropy > -0.887443: standing (28.0)
|   |   |   |   |   z2-energy > 75034414: biking (28.0)
|   |   x1-energy > 67082559: biking (105.0)

Number of Leaves  :     16
Size of the tree  :     31
```

Fig. 4.13 Pruned decision tree. X1, Y1 and Z1 are the acceleration axes of the ankle sensor. X2, Y2 and Z2 correspond to the sensor on the hip.

Having described at length both the design and implementation of the WatchMe prototype, we go on to discuss the evaluations carried out.

5. Evaluation

Twenty years of abundant research have been invested in the field of CSCW, to understand how people work in groups and to explore how technology can support their collaboration. Many theories have been formulated and multiple systems have been built and deployed to evaluate these beliefs. It is our hypothesis that lessons learned from this rich domain can also be applied outside the workplace, to help people get on with life. Technologies similar to those used in CSCW can be applied to *computer-supported collaborative living*. The focus of this research has been to build a working prototype, envisioned to be used in the collaborative living environment, as a proof of concept and to gain insight regarding peoples' attitudes to the underlying ideas.

To truly evaluate the system, all the parts would have to be integrated and encompassed in a robustly engineered platform, to enable its use outside the laboratory over several months. Short of being able to perform such an evaluation, we have addressed the parts individually. We conducted a concept evaluation to understand whether people would be willing to share location information, and if so, with whom. We also wanted to know whether they would share "thinking of you" information and how they would react to receiving it. Likewise, we performed a usability study addressing the functionality of the interface and the comprehension of the icons. Additionally, we wanted to see whether having context information would influence the communication mode chosen. We also conducted a survey to assess how the specific features of context influenced the chosen modality. And, finally, we validated our classification algorithms. The details of each study are described below.

5.1 Concept evaluation

A survey was carried out on a group of 32 people with ages spanning thirteen to sixty five, from four different countries and cultures (the USA, Mexico, Israel and Sweden). The subjects were recruited by email and asked to answer an electronic questionnaire. The respondents were encouraged to forward the questionnaire to their friends. The vast majority of the subjects did not know about the project, but they were family or friends of friends of the researchers. The survey (Appendix A) included two different scenarios and questions about them.

Communication modalities and awareness

The first scenario asked the person to imagine s/he had a device, such as a key-chain or mobile phone, which would enable their friends and family to know their whereabouts. The location information would be automatically available without any effort by either party, it would be reciprocal thus preventing one from "spying" on another, and a person would always have the option of switching the device off. It was pointed out that such a device would, for example, "enable a working mom to know that her husband had already left the office, that her son was still at guitar practice (probably waiting to be picked up by dad), and that her daughter was already at home".

In this population, when face-to-face communication with family and friends is not possible, the most common alternatives are intreaction by phone or email, followed by text messaging (IM, SMS). The majority (28/32) would be willing to share information regarding their whereabouts only with immediate family, that is, spouse and children. A few (8/32) would also share with close friends and siblings. Not surprisingly, some teens seemed much less enthusiastic about giving this information to their family, although an opportunity whereby the parents would be aware of inopportune moments to call was valued. People indicated that they would be willing to disclose locations such as home, work, school, gym, supermarket, etc., but few would keep the device turned on all of the time.

Feature set

New features that people would like included are: the ability to know who was watching you; the capacity to talk to the person observing you; a "busy scale" which could either be set manually or "smartly" by the system; the ability to provide a false location if necessary; the option to leave messages; a "general" vs. "detailed" mode indicating for example "shopping" instead of the name of a particular store; the option to request a person to turn their device on; and preventing children from turning their devices off or overriding the system with a false location.

People definitely did not want the system to include: hidden cameras; the option for people to track you without your knowledge; the possibility of hearing everything said; the option to permanently store the information on a person's movements; and for unauthorized people to get hold of this information.

People indicated that they were willing to give some location information to a few chosen people they trust, but were very concerned about being monitored without their consent and knowledge. Almost everyone (31/32) said they would take into consideration a person's location before communicating with him, and would want this courtesy to be reciprocal. We asked what other information, besides location, people would be willing to reveal. The responses received were very bimodal. Many people seem reluctant to provide more of their specific context information and prefer a more abstract "busy" or "do not disturb" label, whereas others want family trying to contact them to know that they are driving, or in a meeting, or on vacation, etcetera.

"Thinking of you"

The second scenario asked people to imagine a device that displayed a picture of whoever happened to be thinking about them. We wanted to know with whom people would be willing to share their thoughts, so to speak, and how they would respond when the device displayed a picture of someone thinking about them. Twenty-two people of the survey group said they would share this experience with a combination of immediate family, close friends and siblings. One person said it would be nice to be able to let friends and family know that he was thinking of them without having to take time to call or write a message, and that he could list at least 30 people he would regularly like to let know he was thinking about. Two people found this idea “creepy” and did not like it.

As for their response, the group who liked the concept of the device, said they would react to receiving a picture by: phoning the person if they were not too busy; have a “warm feeling”, reply by sending back a picture and perhaps phone depending on who they were; would just be happy but not do anything about it; would respond only to spouse; or would email or call them to get together.

5.2 Usability study

We conducted a small evaluation of the watch prototype, focusing on usability, choice of communication modes, and the appeal of such a technology. The 15 subjects (eight female, seven male) were aged 25 to 48, and included students, administrative staff and people not belonging to the lab. The one-on-one sessions lasted from 20 minutes to 1.5 hours. First, we explained and demonstrated the user interface. Then, subjects were given as much time as they wanted to explore the interface display and buttons; in fact, no subject spent more than two minutes doing so. Each subject was asked to perform three specific communication tasks using the device. The device logged the whole interaction and the subjects were observed while performing the tasks by the researcher. At the end of the third task, each subject filled out a questionnaire (Appendix B). After completion of the questionnaire most of the subjects felt compelled to talk about the system in general and the prototype in particular, get more detail, and offer comments. Some of these unforeseen conversations over the prototype lasted close to an hour.

The first task involved sending a text message to a specific person, the second task was to send a voice instant message to someone else, and the third task was to communicate in any modality to a third person. The first two tasks were directed at the usability of the watch, while in the third we wanted to see the utility of the context information of the remote person, and whether having that information affected the communication mode chosen. The watch prototype used was fully functional, with simulated sensor data received via the server.

Usability

Subjects were asked to grade on a 1-7 scale (1-very hard; 7-very easy) how simple the system was to use, and how well they thought they had performed. The mean and standard deviation for ease of use were $\mu = 5.67$ and $\sigma = 0.9$. For the self-reported performance $\mu = 5.6$ and $\sigma = 0.91$, although the observer considered that all had indeed managed to perform the task and everyone did so in 6-7 minutes total.

Almost all the complaints were related to the button interface, rather than to system features or functionality. People found the buttons too small, making it difficult to navigate. Some found it hard to distinguish the buttons and remember which function they performed, though this could be due to the novelty of the device. The robustness of the buttons was an issue, requiring us to re-glue them often. Some people liked the fact that it was not clear which buttons were functional, making the watch look “more like jewelry, less nerdy”. One way to reveal the buttons to the user only is to give them a slightly different texture. Clearly we will have to rethink and redesign the button interface.

A few subjects disliked the “texting” feature. Text messages are composed by choosing characters from a soft keyboard, via the navigation buttons. Each chosen character is appended to the message string. Some users added more than one space character since they had no visual feedback that it had been appended. Once composed, the message is sent by pressing a different button. These two buttons were intentionally placed next to each other to facilitate quick texting with one thumb. Several users confused the buttons, sending incomplete messages. Although most did not bother to send another message with the remainder of what they had intended to write, this could obviously be done. Perhaps only few mentioned these issues because texting on a small device is known to be problematic and hence their expectations were low.

Choice of communication mode

In the third task, the person they were to communicate with had left them two text messages and one voice message; the context data indicated that she was driving and expected to be home in 35 minutes. Sixty percent chose to give her a call with explanations such as: “she is driving so text is not a good option but she seems available”; “I called because her voice message said *give me a call*”; “it seemed urgent and this was the quickest way to reach her”. Three people left a voice message and explained that the recipient was driving and therefore a phone call was not recommended, and three left a text message since it was the easiest for them. Seven said they chose a mode based on the recipient’s convenience, four considered only their own, one person considered both, and three considered neither. Fourteen out of the fifteen correctly interpreted the remote person’s context data.

The voice message the subjects listened to indeed said “Give me a call when you get a chance”, however this was said in a casual tone. Since the messages were from a fictitious person, and not from an insider as the system is envisioned to be used, the subjects’ interpretation of the context varied. Those who thought it was urgent to get in touch with her did not believe convenience to be a relevant factor. One person misinterpreted the context data –he thought she had been home for the last 35 minutes, and not that her ETA was 35 minutes– he afterwards said that in that case he would have just waited until he saw that she had arrived home and only then phoned.

We also asked about general preferences of communication channels. Text messaging was the least preferred for sending but, significantly, what people said they preferred for receiving. Composing a text message on a small device with few buttons can indeed be tedious. The asynchronous text mode for reception is generally preferred since it can be accessed at any time and there are no privacy concerns with others listening in. It is also faster to read than to sequentially listen to audio.

Appeal

Subjects were asked how much they liked the system (1-not at all; 7-very much), what they specifically liked and disliked about it, and who they would share this type of information with. People seemed to really like the system ($\mu = 6.07$, $\sigma = 0.96$); 10/15 would share this information with their spouse or boy/girl-friend, 7 would share with other family members such as siblings or parents, and 9 would share with some close friends. Twelve, out of fifteen, indicated that they would use such a technology, however only eight would want it as a watch.

As noted before, the predominant thing said against the system was the button interface. The graphical user interface was generally found to be simple, intuitive and aesthetic. Many liked the icons and especially the information they convey. Someone noted that he would like the context data to feel more in touch with his girlfriend and other friends who are all on the other side of the Atlantic.

Comments regarding the different communication modalities were very positive: "*I really like the features in this watch. In general I hate all-in-one devices, but this one is great. It groups together things that make sense, they all have to do with communication, and in a simple way*"; "*it let's me communicate more politely*"; "*I like the blurring of the boundaries between message types*".

Others projected how such a device would fit into their own life: "*This is the perfect device for me and my brother. We talk five times a day. Ninety percent of all of my calls are to him. Often I call just to get the information I see here.*"; "*This is awesome. Don't need to phone kids 'where are you'?!, Don't need to embarrass them in front of the other kids. Don't need to freak out because the kid doesn't answer his phone, e.g. is in swim practice and doesn't have the phone in the pool!!*".

One surprising fnding was that seven of our subjects no longer wear watches. For some this is due to the physical constraints (heavy, make you sweaty, etc.), while many noted that the time is readily available, e.g. on their mobile phones, computers, or clocks in the environment. Clearly people who do not wear watches are less inclined to a technology that you wear on your wrist, but if the phone and the watch become the same gadget, this new trend may be reversed. In any case, a surprising number of people liked the technology; those who do not want it on their wrist would like to have a device you could clip to your belt or put in a pocket, or simply on a conventional mobile phone. Not having the device located in the periphery of visual attention would require rethinking the design of the interaction, perhaps relying more on auditory or tactile cues.

5.3 Features of context

We conducted an anonymous online survey to see how knowledge of different parameters of context affected how and when people chose to communicate, e.g. whether they chose a more/less intrusive mode in different situations. A URL to the survey was sent to five people unfamiliar with the project and they were asked to forward the link to others. We received 50 responses (26 female, 24 male) from seven different countries, spanning the ages 20-65. The survey (Appendix C) included three scenarios, and based on different snippets of context information, people were asked to answer how they would chose to communicate, and the reasons for their choice.

We believed that knowledge of context would have an effect on the modality chosen; that urgency of a message would, and should, override consideration of the remote person's context; and that providing people with information that someone was thinking about them would foster communication.

We found that the choice of communication mode did vary based on the context information provided. There was a lot of variability within a subject's answers across different contexts, however we did not see any strong trends of preferred modalities in specific situations across subjects. We found indication that providing "thinking of you" information could foster communication: of the 50 respondents, 88% reported that their reaction to receiving such information would be to contact the person in question; 72% of them would call.

To sum up, from these three evaluations:

- We found that people are willing to share information with family and friends. In the concept survey 28/32 indicated they would share it with immediate family. In the usability study, after using the prototype, 10/15 participants stated they would share it with their spouse/partner, seven would also share with other family members, and nine with some close friends.
- We also found that abstracted sensor data can be presented on a small interface such that it can be interpreted at a glance. In the usability study people indicated that the interface and icons we simple, intuitive and aesthetic. People readily interpreted the context data – 14/15 interpreted it correctly.
- Additionally, providing awareness information can affect the choice of communication channel. In the usability study 12/15 took it into consideration before choosing a communication modality and in the context survey people adapted their choice based on the different contexts.
- Both the surveys suggest that providing "thinking of you" information could foster communication between insiders.

An interesting evaluation to perform in the future would be a study to correlate how available for communication an individual considered himself to be, versus how available he was perceived by others based on the context information conveyed. This could be evaluated by collecting sensor data from a participant throughout a day. The data could then be played back to that same participant on the watch interface, when the events of the day were still fresh in his memory, and at random points in time the participant would be asked to rate how available he was, for example, for a phone call from a close friend. This same data could then be similarly played to some of the participant's insiders, and at the same points in time they would be asked to rate how available they thought the person was for a phone call from them. It would be interesting to see how the answers differed between people who knew the first participant, i.e. insiders, versus strangers. Certain features of context might be interpreted equally in both cases, however it is expect that insiders familiar with another's routines and preferences will be able to form a broader interpretation of the sensor data. A stranger would know, for example, that the person left home 10 minutes ago, is driving, and in a conversation. An insider however may understand from that same information that the person has already picked up his friend and that they are heading to the gym.

Another very interesting evaluation would be a longitudinal study, for example on a family with children, to see how the system: was used to infer availability; coordinate household members and activities; whether it fostered communication or helped maintain a sense of connectedness; whether communication patterns between the family members changed over time; and whether people are comfortable sharing this type of information on an ongoing basis even with their insiders.

5.4 Classifiers

5.4.1 Location learning

The location learning and route detection algorithms were validated over a period of several months with real data from two different users, each with very different travel patterns. All of the personal landmarks were found automatically from the data, i.e. although the users had the option to manually mark locations as important, the feature was not used. Personal landmarks found included: home, work, food trucks, supermarket, a friend's house, daycare, restaurants, and a couple of favourite picnic spots. The system found all of the personal landmarks the two users expected it to find. A known limitation is the learning of locations where the user spends very little time, e.g. a location where a child is frequently dropped off, however these could be manually marked. Not surprisingly, the system was unable to learn or identify locations in very close proximity, such as two adjacent stores.

Routes between the endpoints were automatically identified, and canonical route representatives created. The routes of one user are very constant, both in the trajectory traveled and the time to destination. In this case, 2-3 examples of each route are enough to create a route template that yields very high prediction accuracy of both destination and ETA. Some of the routes of the second user had a lot of variance, both in path and travel time, as he often deviates and goes for a bike ride or walk before eventually retaking his route to destination. In this case three different route templates were created to represent the different trajectories.

The route prediction was tested through an interface that enabled the user to load a file of his GPS data and progress though the 100 m route segments, at each stage seeing the system's prediction based on its alignment to the canonical routes –the prediction information was either plotted or displayed on the watch interface. For both users the alignment to the correct route was very accurate. In cases where more than one route overlap, e.g. when a user leaves home and travels the same segment before deviating either to the supermarket or daycare, the prediction is based on the route it best aligns to with the given data points, even if one destination is much more frequented than the other. This could be modified in the future to consider the priors of each different route and in cases of overlapping segments, predict based on the most frequently traveled route. The ETA prediction was found to be accurate in cases where there was a good canonical representative, i.e. built from examples with similar travel times.

When the full traveled trajectory does not align well to a canonical route, the system tries aligning the last 500 m segment to any known route. Again, this alignment turned out to be accurate in predicting the part of the known route the user was on, albeit not always the direction; the system identified the user on route AB or BA, but could not differentiate between them. The routes could be disambiguated by analysing the actual latitude/longitude coordinates.

5.4.2 Speech detection

The speech detection algorithm was validated using its Matlab instantiation, running on a desktop computer with files recorded from the iPaq version. A more formal evaluation would be required, however the algorithm was found to work adequately in an office environment.

5.4.3 Activity recognition

Many have shown recognition rates of 85-95% for ambulation from acceleration data, and rates of 80-95% without individual training [Uiterwaal98, Mantyjarvi01, Lee02]. Bao (2003) conducted a study on several subjects and various activities in semi-naturalistic settings, using multiple accelerometers on different body locations. We are using the algorithm with the highest reported accuracy from that study, and classifying only a subset of the activities examined. We have not tried to duplicate his findings and conduct an evaluation on other subjects.

The implemented algorithm, the decision tree, was validated using the 10-fold cross-validation method. This involves randomly dividing the labeled training data into 10 parts, such that the different classes are represented in proportions similar to the full data set. The algorithm is trained on nine-tenths of the samples and an error rate is calculated by testing on the remaining tenth. This is repeated for each of the 10 folds, and the 10 errors are averaged to produce the overall generalization error. The results obtained were:

Correctly Classified Instances	1131	96.5841 %
Incorrectly Classified Instances	40	3.4159 %

Figure 5.1 shows the Confusion Matrix, i.e. the actual classification versus the predicted one. The overall 96% accuracy is an upper-bound; the error rate is expected to be higher for individuals other than the trainer. However, the generalization error was calculated by looking only at one instance at a time, whereas in the real-time classifier implemented on the iPaq, the activity-state is toggled only after seeing three consecutive examples identically classified, and not just an instantaneous decision; this reduces some of the misclassifications during transitions, for example stops when biking (classified as standing) or initial pedaling movement (sometimes confused with walking).

	Predicted				
Actual	walking	standing	running	biking	driving
walking	463	4	4	6	0
standing	5	163	0	5	2
running	2	0	27	0	0
biking	7	1	1	139	0
driving	0	3	0	0	339

Fig. 5.1 Confusion Matrix showing the actual and predicted classifications.

Figure 5.2 shows the detailed accuracy per class. The recall or true positive rate (TP) is the proportion of correctly identified samples, the false positive rate (FP) is the proportion of incorrectly classified samples, and precision is the proportion of predicted positive cases.

Class	TP Rate	FP Rate	Precision	Recall
walking	0.971	0.02	0.971	0.971
standing	0.931	0.008	0.953	0.931
running	0.931	0.004	0.844	0.931
biking	0.939	0.011	0.927	0.939
driving	0.991	0.002	0.994	0.991

Fig. 5.2 Detailed accuracy per class.

6. Related work

6.1 Awareness through video and audio

The Montage [Tang94] system provided lightweight audio and video "glances" to support a sense of cohesion and proximity between distributed collaborators. It used a hallway metaphor where one could simply glance into someone's office to see if it was a good time to interact. A similar metaphor was used in Cruiser [Root88, Fish92], an earlier system, which enabled a user to take a cruise around other offices. The purpose of the system was to generate unplanned social interactions. The user was presented with a few seconds of audio and video from each office on either a planned or unplanned path. The Portholes project [Dourish92] also aimed to initiate informal communications between non co-located workers. The approach taken was to periodically present –on a user's workstation– updated digitized images of the activities occurring in public areas and offices. These systems required cameras and microphones set up in the workspace, and broadband connections to support the transmission of video and/or audio. Like them, *WatchMe* aims to provide awareness of a remote place to help a user infer an opportune moment for interaction. It however supports mobility, requiring no infrastructure in the environment, and uses low bandwidth.

Although many media spaces have utilized both audio and video to provide shared awareness, some systems have focused on awareness solely through audio. Thunderwire [Ackerman97] was an audio-only shared space for a distributed group. It was essentially a continuously open conference call in which anything said by anyone could be heard by all; users could connect, disconnect or operate in listen-only mode. It had no visual interface so the only way of knowing who was listening was by asking. *WatchMe* does not transmit the ambient audio, rather it analyses it to determine whether it corresponds to speech, and transmits only the classification result. Additionally, it enforces symmetry, preventing users from being solely in "reception mode", and moreover the user knows not only who has access to his information but also who actually accesses it. ListenIn [Vallejo03] used audio to provide awareness of domestic environments to a remote user. In order to add a layer of privacy, the audio was classified and, as in *WatchMe*, a representative icon was presented instead of the raw data; if the audio was classified as speech it was garbled to reduce intelligibility. We envisioned connecting the projects such that one of the insider icons on the watch would correspond to the user's home, and the context icons would reflect the activity as classified by ListenIn.

6.2 Location awareness

Groupware calendars have been useful tools to locate and track colleagues. Ambush [Mynatt01] looked at calendar data to infer location and availability. It used a Bayesian model to predict the likelihood that a user would actually attend an event entered in his calendar. The Work Rhythms project [Begole02] looks at location of computer activity to create a user's temporal patterns. Awareness of these patterns helps co-workers plan work activities and communication. When a user is "away", the system can predict when he will be back. Calendars and probabilistic Bayesian models have also been used in a notification system to infer a user's state of attention [Horvitz99] –based on computer activity and location (predicted from scheduled appointments in calendar, and ambient acoustics in user's office)– and calculate the cost of interrupting and delivering a message. These systems are effective provided a user's activity is confined to a computer, or detailed in a structured calendar. In life outside the workplace most of our activities are not measured by keyboard use, and household calendars are typically less structured and not usually online. *WatchMe* predicts a user's location from his current geographic position (latitude/longitude) and his learned patterns of mobility.

Location-aware systems have also used infrared or radio frequency sensors to keep track of electronic badges worn by people [Want92], or GPS [Marmasse00]. The main advantage of using GPS is that it is low-cost, requires no additional deployment of infrastructure in the environment, and is maintenance free as far as the user is concerned. Its primary drawback is that –unless it is assisted for example through the phone cellular network– it requires line-of-sight to satellites in a good geometry constellation and therefore only functions reliably outside.

6.3 Context and mobile telephony

The so-called context-awareness of computer systems falls very short of what humans can assess. As Erickson [2001] puts it: "the ability to recognize the context and determine the appropriate action requires considerable intelligence". Several systems keep the human "in the loop" by enabling the potential recipient to select a profile appropriate for the context. In the Live Addressbook [Milewski00] users manually updated their availability status and the location where they could be reached. This information, as well as an optional brief message (e.g. "urgent only"), was displayed to anyone trying to contact them. Although the updates were manual, the system prompted the user when he appeared to be somewhere other than the location stated. Another system that shares the burden of the decision between caller and callee is Context-Call [Schmidt00]. The potential recipient updates his context on his cell phone (free/meeting/working/at home/busy, or free form text). When the caller places a call he will be notified of the stated context and have the option to leave a message, place the call anyway, or cancel. These systems helped people make more informed telephone calls by providing them with context information prior to potentially interrupting at an inopportune moment. This however came at the price of having to manually update the context, and such information is only relevant if it reflects the user's current state. In *WatchMe* the context is automatically updated; the user only needs to take action if he wants to prevent the information from being sent.

Quiet Calls [Nelson01] enabled users to send callers pre-recorded audio snippets, hence attending a call quietly. The user could listen to what the caller was saying and send a sequence of standard answers, enabling him to negotiate postponing the call to a later time, or having the caller wait until he exited a meeting, for example. This system provided a means to better handle calls if they occurred at an inconvenient moment. The approach in *WatchMe* is to provide information helping the caller assess whether it is a suitable moment, as well as various communication channels to negotiate availability before the call.

6.4 Lightweight text communication

Babble [Erickson99] aimed to support communication and collaboration among large groups of people. It presented a graphical representation of user's availability, based on their computer interaction. Nardi *et al*. [2000] studied the extensive use and affordances of instant messaging (IM) in the workplace. In these settings, lightweight text communication is used for rapid exchanges of information and affect, or to negotiate communication via a different channel. We believe that it will serve a similar function in *WatchMe*.

Desktop tools for managing communication, coordination and awareness become irrelevant when a user is not near their computer. The Awarenex system [Tang01] extended instant messaging and awareness information to wireless handheld devices. Hubbub [Isaacs02] addressed awareness and opportunistic text conversations on a mobile platform. IM systems typically use sound to indicate that someone has logged on/off, however the sound is generic and does not indicate who it is. Hubbub provided awareness through musical sounds unique to each user, enabling others to know (without looking) who had just turned from idle or offline to active. A significant fraction of the communication occurred immediately after the person turned active, suggesting the usefulness of awareness information for opportunistic interaction. This system also had location information manually updated by users. In *WatchMe* we believe that the "smiles", or "thinking of you" information, will likewise foster communication at opportune moments.

6.5 Non-verbal communication systems

There are a few systems that have looked at ways to enhance interpersonal communication by adding physical feedback via actuators. ComTouch [Chang02] was designed to augment remote voice communication with touch. It translated in real-time the hand pressure of one user into vibrational intensity on the device of the remote user. The Kiss Communicator [Buchenau00] enabled couples to send each other kisses. One person would blow a kiss into one side of the device and the remote piece would start blinking. The other person could respond by squeezing the communicator causing the lights to blink on the side of the original sender. The Heart2Heart [Grimmer01] wearable vests conveyed wireless “hugs” by simulating the pressure, warmth and sender’s heart-beat as would be felt in a real embrace. In these systems it is the user who explicitly triggers the effect on the remote device. In *WatchMe*, although the user has the option to explicitly send non-verbal messages, the “smiles” or “thinking of you” information is automatically transmitted when one user thinks about another and views his context information. Paulos [2003] suggests a system with sensors (accelerometer, force sensing resistors, temperature, microphone for ambient audio) and actuators (Peltier temperature junctions, bright LEDs, vibrator, “muscle wire”, speaker for low level ambient audio) to enhance non-verbal telepresence. This system will use Intel’s Motes and will include a watch interface. It appears that the messages will also be explicitly sent.

6.6 Family communication

Technology to help maintain family connectedness is a more recent theme. To date the focus has been on communication between remote households or maintaining awareness between mobile house members and the home. The Casablanca project [Hindus01] explored ways to connect homes. Turning on a Presence Lamp in one home would cause a matching lamp to be lit in the remote household. The remote locations also shared a digital message board. The aim of the Digital Family Portraits [Mynatt01] concept was to provide awareness of the activities of elderly family members living on their own and promote peace of mind to their remote family. The prototype used icons, and manual updates simulating sensor input, to provide awareness of daily life and maintain a history of activity over a 28-day period. The ASTRA project [Romero03] connects a household to its mobile individuals. Mobile users capture and send visual cues of events in their life to a home device where they can be viewed by family members. In *WatchMe*, although we have considered connecting the mobile family member with his home environment, via ListenIn [Vallejo03], its main goal is to maintain and strengthen the connections between the different mobile household members themselves.

Go *et al*. [2000] proposed the concept of Familyware: tools to increase the feeling of connection between people who have a close relationship. Their definition includes objects that send simple signals to convey a shared feeling, however, they rule out text-based information and audio- or video-mediated communication. One of their prototypes is a Teddy bear wired to a PC. When a child manipulates the stuffed animal, her physical behaviour serves as a trigger and her photograph appears on her father's computer screen. Kuwabara *et al*. [2002] define *connectedness oriented communication* as a mode focused on maintaining and enhancing social relationships, or fostering a sense of connectedness. They contrast it to *contents oriented communication*, the goal of which is the exchange of information. As they point out, the existing communication technologies do not really support connectedness. *WatchMe* combines both content and connectedness oriented communication, two aspects of communication it believes to be fundamental to maintaining a functional and socially healthy household.

7. Conclusions

7.1 Contributions

In this thesis we have identified how powerful concepts, resulting from two decades of grounded research on the role of technology in the workplace (CSCW), are also applicable in other aspects of life. We call this niche Collaborative Living, and have described design criteria pertinent to it.

We set out to build a prototype that satisfied these criteria. The main research questions were whether it was possible to gather relevant data from sensors, present it in a meaningful way, within a user's peripheral vision, on a mobile device. And whether this information could provide some meaningful understanding of a remote person's context. Based on the rich body of literature, we believed that information presented in this manner could be assimilated by the wearer at a quick glance and, furthermore, that receiving this information would indeed foster communication. Moreover, we believed that provided with multiple channels of communication, people would choose the communication mode most appropriate to a user's context.

This thesis encompasses two main contributions. First, a watch which fuses abstracted sensor data from multiple sources, and presents it on a device easily accessible and frequently visible. The watch provides multiple verbal and non-verbal wireless communication channels and automatically alerts a remote person when someone is thinking of them and viewing their context data. Second, we have developed algorithms for location learning and route prediction. Our algorithm can pick out potential personal landmarks from GPS data, which are then presented to the user on a map interface for labeling. A user could manually provide a system with locations of interest, however this algorithm makes that unnecessary. Routes between named locations are then identified and canonical route representatives are created. The route learning algorithm enables the system to not only display a user's current location, or the elapsed time since his last known one, but also predict where he is heading and estimate his time of arrival.

In the process of building the prototype, we have designed and developed several hardware and software components. These include the design of a display in a watch form factor, as well as the design of icons to represent the information. Additionally, a stand-alone mini sound-card with its own onboard file system was created. It interfaces through a serial port to a PDA or mobile phone, from which it receives commands for audio capture or playback. We also contributed to the next generation of wireless accelerometers that are small and light enough to be used in real-life situations.

7.2 Future work

Microelectronics are becoming smaller, faster and cheaper. An important issue is what applications they will support and how these will enhance people's lives. This thesis is a first step towards building a system that keeps members of a closely-knit group continuously connected via awareness information and multiple channels of verbal and non-verbal communication. It is an example of a system to enhance Collaborative Living. We have used the watch prototype to demonstrate the functionality and as a platform to stimulate dialogue regarding what is socially desirable for applications in this area.

The next step would be to build prototypes which are small and robust enough to be used in real life over an extended period of time. As designers we have built into the system our own assumptions as to how such a technology would be used. Our evaluations and conversations have provided insight into how people think they would use such a system or how they did so in an experimental setting. However, only through a longitudinal study will we see how people really integrate such a technology into their life, and whether it enhances it.

Location information is often considered very private. Perhaps more sensitive than being seen at a specific location at a given time –especially since the user has the option of pretending he is “out of range”– are the patterns that might emerge over time. It is also very possible that people will only want to share certain locations with some of their insiders, and not have to make a binary decision to share with all or with none. It would be interesting to see how much information people are willing to share, even with their intimates, on an ongoing basis. A longitudinal study would also be needed to show whether and how communication patterns between the family members change over time, whether the system fostered communication between them, and whether the awareness information provided assisted the coordination of household members and activities.

The route prediction algorithms could be extended to include the prior probability of traveling a particular route. This would help to better predict in cases where two or more routes overlap. In the current implementation, if the route from A to B and from A to C are initially the same, before deviating to B or C, the system will predict the route it happens to better align to at a particular moment. If route AB is much more frequented than route AC, considering the priors, the system would always predict route AB for the portion they overlap on. The algorithms could also be extended to consider temporal travel patterns, such as the day of the week or the times that specific routes were traveled. Issues with data communication due to loss of signal, as well as poor GPS accuracy, in certain areas can make it difficult to track routes in progress. The watch only displays route predictions it is fairly certain of. The display could be modified to show levels of confidence, for example colour-coded into the location icon, or an extra zoom layer providing more detailed information.

If communication is the *sine qua non* of a functional and socially healthy household, as the niche of Collaborative Living develops we eagerly expect to see more technologies that truly link people-to-people.

Appendix A – Concept Questionnaire

Questionnaire to assess new communication technology

1. Age:
2. Gender:
3. Education: (please circle one)
   elementary, high school, some college (partial), undergraduate, graduate, post-graduate

4. Country of residency or citizenship:
5. Besides face-to-face, how do you usually communicate with members of your **<u>family</u>**? (in the last 3 years) Please note all that apply and how frequently you use this mode of communication, on a scale of 1-5.

method	frequency (1=never, 3=sometimes, 5=always)
1. land-line phone	|-------|-------|-------|-------| 1 3 5
2. mobile phone	|-------|-------|-------|-------| 1 3 5
3. text messaging on computer	|-------|-------|-------|-------| 1 3 5
4. text messaging on mobile device	|-------|-------|-------|-------| 1 3 5
5. email on computer	|-------|-------|-------|-------| 1 3 5
6. email on mobile device	|-------|-------|-------|-------| 1 3 5
7. letters	|-------|-------|-------|-------| 1 3 5
8. postcards	|-------|-------|-------|-------| 1 3 5
9. packages	|-------|-------|-------|-------| 1 3 5
10. other (please state)	|-------|-------|-------|-------| 1 3 5

Of the communication methods you indicated in question 5, please rank the three most used (in order of use)? ________________
____
____

How do you communicate with your **friends**? (in the last 3 years)

Please note all that apply and how frequently you use this mode of communication, on a scale of 1-5.

method	frequency (1=never, 3=sometimes, 5=always)
1. land-line phone	|-------|-------|-------|-------| 1 3 5
2. mobile phone	|-------|-------|-------|-------| 1 3 5
3. text messaging on computer	|-------|-------|-------|-------| 1 3 5
4. text messaging on mobile device	|-------|-------|-------|-------| 1 3 5
5. email on computer	|-------|-------|-------|-------| 1 3 5
6. email on mobile device	|-------|-------|-------|-------| 1 3 5
7. letters	|-------|-------|-------|-------| 1 3 5
8. postcards	|-------|-------|-------|-------| 1 3 5
9. packages	|-------|-------|-------|-------| 1 3 5
10. other (please state)	|-------|-------|-------|-------| 1 3 5

Of the communication methods you indicated in question 7, please rank the three most used (in order of use)?

____

____

____

Do you normally carry a mobile phone for communication?

scenario:

Imagine you had a key-chain on which you could see a symbol indicating where your family and friends are. For example, a picture of a house would mean that they were “at home”, or a picture of a supermarket cart might mean that they were “shopping”. The information would be reciprocal, so if you get their information, then they also have the possibility to get yours. This information would be known automatically, requiring no effort on your or their part –i.e. it would not have to be manually updated. You would always have the option to turn the gadget on or off, giving you full control.

With such a gadget, a working mom could for example see that her husband had already left the office, that her son was still at guitar practice (probably waiting to be picked up by dad), and that her daughter was already home.

Assuming you would use such a system:

6. Who are the people you would be willing to share this type of information with?
   (e.g. spouse, parents, grandma, my 3 best friends, my girlfriend, etc.)

7. Which type of locations would you be willing to share with all of them?
   (e.g. work, home, gym, supermarket, school, hospital, church, etc.)

8. Which locations would you only share with certain people? Please state which locations and which people. (e.g. my parents can’t know if I am at my boyfriend’s, etc.)

9. Would you keep the gadget turned on all of the time?
   If not, in which types of situations would you turn it on/off? Please explain why.

10. Assuming you wanted to communicate (e.g. phone call, text message, etc.) with someone who had this gadget, would you first consider where they were (location)?

11. Is knowing just the location of the other person enough to make a decision? If not, please mention why not and what other factors should be taken into account.

12. Would you like them to consider your location/situation, before communicating with you? (yes/no)

13. What type of additional features would you like the gadget to have?
(e.g. the option to pretend I was somewhere else, etc.)

14. What type of features would you definitely **not** want it to include? Please explain why.
(e.g. transmitting live video, the recording of conversations, etc.)

15. If the gadget could let others know more information about your situation (besides your location) what other information would you be willing to share?
(e.g. the fact that you were in a conversation, the fact that you were driving, eating, etc.)

16. Rank how you would like to be notified that someone has for example left you a text message.
(1-most desirable method, 3-least desirable)

audio notification ____ please explain why ___________________________________

visual notification ____ please explain why ____________________________________

tactile notification ____ please explain why ____________________________________

scenario:

Imagine that this same key-chain could let you know when one of your family or friends looked at their key-chain to know your location. Wanting information about you indicates that you are in their thoughts.

For example, if someone dear to you looked at their gadget to see where you are, your gadget would suddenly show a picture of their face.

17. Would you like to share your thoughts in this fashion? Please explain why?

18. Who are the people that you would like <u>them</u> to know you were thinking of them?

19. Who are the people <u>you</u> would you like to know that they were thinking of you?

20. If the gadget showed you the picture of someone you cared about, what would you do with this information? (e.g. just have a “warm feeling”, send them a picture back, phone them, send them a text message, make a reservation for dinner, etc.).

21. Besides receiving the visual notification (i.e. the picture), would you like audio notification too? If so, what type of audio? (e.g. a beeping sound, part of a song, their voice, etc.)

In general, regarding the gadget described in the above scenarios:

22. What <u>positive</u> social implications could such a device have? Please explain.

23. What <u>negative</u> social implications could such a device have? Please explain.

24. Would you use this device?

Appendix B – User Interface Questionnaire

User-Interface Evaluation – Questionnaire

Date: ______________

Subject Experiment ID: _______________

1. Age:
2. Gender: F / M

3. How easy was it to use the system?

very hard |-------|-------|-------|-------|-------|-------| very easy
1 3 5 7

4. How well do you think you performed the task?

very badly |-------|-------|-------|-------|-------|-------| very well
1 3 5 7

The next 3 questions refer to the 3rd communication task you performed.

5. Which communication channel did you use? text message / voice message / phone call
6. Why did you choose this channel?

7. Did you consider your convenience or the recipient's? Please explain.

8. Which communication channel would you generally prefer for <u>sending</u> messages? Please explain.

9. Which communication channel do you dislike the most for <u>sending</u> messages? Please explain.

10	Which communication channel would you generally prefer for <u>receiving</u> messages? Please explain.	
11	Which communication channel do you dislike the most for <u>receiving</u> messages? Please explain.	

12	To what extent did you like the system?	not at all |-------|-------|-------|-------|-------|-------| very much 1 3 5 7
13	What did you like about it?	
14	What did you dislike about it?	

15	Would you use such a system?
16	Who are the people you would share this type of information with? (e.g. significant other, parents, siblings, children, close friends, etc.)
17	Would you wear it on your wrist?
18	Do you normally wear a watch?

19 Why do you wear the particular watch that you do?
(e.g. was my grandfather’s, looks cool, etc.)

20 Any general or specific comments you would like to add?

Appendix C – Context Questionnaire

Online Communication and Context Questionnaire

Below are 3 different scenarios in which you are asked **how** you would choose to communicate, and **why** you would choose that method. The different communication modes are:

- text to a phone (e.g. SMS)
- Instant Messaging
- email
- a voice message
- "none", meaning that you would simply decide not to communicate

Example
Q: You want to communicate with your friend Mike to tell him about your promotion.
If you knew he was at the gym, what method would you use? and why?
A: voice message
reason: if he is at the gym, his mobile phone is probably in his locker and not on him,
so I would just leave a message saying I have great news.

1. Age:
2. Gender:

scenario 1:

It is mid-afternoon and you feel like chatting with your boyfriend/girlfriend or spouse to catch up on their day and possibly make plans for the evening.

3. Which communication mode would you use? __________ Why?

4. If you knew s/he was in a meeting.
   Which communication mode would you use? __________ Why?

5. If you knew s/he was driving.
   Which communication mode would you use? __________ Why?

6. If you knew s/he was in a conversation.
   Which communication mode would you use? __________ Why?

7. If you knew s/he was thinking about you at that very moment.
   Which communication mode would you use? __________ Why?

8. If you knew s/he had left you several messages throughout the day.
   Which communication mode would you use? __________ Why?

9. If **you** were the boyfriend/girlfriend or spouse,
   how would you prefer to be contacted? __________ Why?

scenario 2:

You need to talk to your brother about something urgent. It is the middle of the day. You know he is usually very busy, and hates being interrupted.

10. Which communication mode would you use? _________ Why?

11. If you knew for certain he was in a meeting.
    Which communication mode would you use? _________ Why?

12. If you knew he was driving.
    Which communication mode would you use? _________ Why?

13. If you knew he was walking, on his way to get lunch.
    Which communication mode would you use? _________ Why?

14. If **you** were the brother/sister,
    how would you prefer to be contacted? _________ Why?

scenario 3:

You have recently had a big argument with a very close friend. You are now wondering whether you should communicate to clarify your position.
(Imagine it is a specific friend of yours.)

15. Which communication mode would you use? _________ Why?

16. If you knew s/he was sitting in the park.
    Which communication mode would you use? _________ Why?

17. If you knew s/he was in the office, in a conversation.
    Which communication mode would you use? _________ Why?

18. If you knew s/he was at home.
    Which communication mode would you use? _________ Why?

19. If you knew s/he had left you several messages throughout the day.
    Which communication mode would you use? _________ Why?

20. If **you** were that close friend,
    how would you prefer to be contacted? _________ Why?

References

Ackerman, M., Hindus, D., Mainwaring, S. and Starr, B. (1997) Hanging on the Wire: A Field Study of an Audio-Only Media Space. *Transactions on Computer-Human Interaction (TOCHI)*, vol. 4, no. 1, pp. 39-66.

Ackerman, M. (2000) The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility, *Human-Computer Interaction*, vol. 15, no. 2-3, pp. 181-205.

Argyle, M. and Dean, J. (1965) Eye-contact, Distance and Affiliation, *Sociometry*, 28, pp. 289-304.

Bao, L. (2003) Physical Activity Recognition from Acceleration Data under Semi-Naturalistic Conditions. M.Eng. Thesis, Massachusetts Institute of Technology.

Begole, J., Tang, J., Smith, R. and Yankelovich, N. (2002) Work Rhythms: Analyzing Visualizations of Awareness Histories of Distributed Groups. *Proceedings of the CSCW 2002 Conference on Computer Supported Cooperative Work*, pp. 334-343.

Brave, S., Ishii, H. and Dahley, A. (1998) Tangible interfaces for remote collaboration and communication. *Proceedings of the CSCW 1998 Conference on Computer Supported Cooperative Work*, pp. 169-178.

Bouten, C., Koekkoek, K., Verduin, M., Kodde, R. and Janssen, J. (1997) A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Bio-Medical Engineering*, vol. 44(3), pp. 136-147.

Buchenau, M. and Fulton, J. (2000) Experience Prototyping. *DIS '00 Symposium on Designing Interactive Systems*, pp. 424-433

Chang, A., O'Modhrain, S., Jacob, R., Gunther, E. and Ishii, H.(2002) ComTouch: Design of a Vibrotactile Communication Device. *DIS '02 Symposium on Designing Interactive Systems,* pp. 312-320.

Christiansen, N. and Maglaughlin, K. (2003) Crossing from Physical Workspace to Virtual Workspace: be AWARE!, *Proceedings of HCI International Conference on Human-Computer Interaction*, pp. 1128-1132.

Clark, H. (1996) *Using Language.* Cambridge University Press.

Coates, D. (2003) *Watches Tell More Than Time. Product Design, Information, and the Quest for Elegance.* McGraw-Hill.

Danchak, M., Walther, J. B. and Swan, K. (2001) Presence in Mediated Instruction: Bandwidth, Behavior, and Expectancy Violations, *ALN 2001 7th International Conference on Asynchronous Learning Networks*.

Dey, A. (2001) Understanding and Using Context. *Personal and Ubiquitous Computing (PUC 2001)*, vol. 5, no. 1, pp. 4-7.

Dourish, P. and Bly, S. (1992) Portholes: Supporting Awareness in a Distributed Work Group. *Proceedings of the CHI '92 Conference on Human Factors in Computing Systems*, pp. 541-547.

Dvorak, J. (2003) Social Aspects of Conformables. Presented at *IEE Eurowearable '03*.

Erickson, T., Smith, D. N., Kellogg, W. A., Laff, M., Richards, J. T. and Bradner, E. (1999) Socially Translucent Systems: Social Proxies, Persistent Conversation and the Design of "Babble". *Proceedings of the CHI '99 Conference on Human Factors in Computing Systems,* pp. 72-79.

Erickson, T. (2001) Ask not for Whom the Cell Phone Tolls: Some Problems with the Notion of Context-Aware Computing. *Communications of the ACM*.

Fish, R., Kraut, R., Root, R. and Rice, R. (1992) Evaluating Video as a Technology for Informal Communication. *Proceedings of the CHI '92 Conference on Human Factors in Computing Systems,* pp. 37-48.

Foerster, F., Smeja, M. and Fahrenberg, J. (1999) Detection of posture and motion by accelerometry: a validation in ambulatory monitoring. *Computers in Human Behavior*, vol. 15, pp. 571-583.

Gerasimov, V. (2001) Hoarder Board. http://vadim.www.media.mit.edu/Hoarder/Hoarder.htm

Go, K., Carroll, J. and Imamiya, A. (2000) Familyware: Communicating with Someone You Love. *Proceedings of the IFIP TC9 WG9.3 International Conference on Home Oriented Informatics and Telematics, HOIT2000*, pp. 125-140.

Goffman, E. (1961). *The Presentation of Self in Everyday Life*. New York: Anchor-Doubleday.

Greif, I. (Ed.). (1988). *Computer-supported cooperative work: A book of readings*. San Mateo, CA, Morgan Kaufmann Publishers.

Grimmer, N., (2001) Heart2Heart. *Winner of Intel Student Design Competition 2001*. http://www.baychi.org/calendar/20010508/#1

Grinter, E. and Eldridge, M. (2003) Wan2tlk: Everyday Text Messaging. *Proceedings of the CHI '03 Conference on Human Factors in Computing Systems*, pp. 441-448.

Grudin, J. (1994) CSCW: History and focus. *IEEE Computer,* vol. 27, no. 5, pp. 19-26.

Grudin, J. (1994) Groupware and social dynamics: eight challenges for developers. *Communications of the ACM,* vol. 37, no. 1, pp. 92-95.

Gunawardena, C. (1995) Social Presence Theory and Implications of Interaction and Collaborative Learning in Computer Conferencing. *International Journal of Educational Telecommunications,* vol. 1, no. 2-3, pp. 147-166.

Hansson, R., Ljungstrand, P. and Redstrm, J. (2001) Subtle and Public Notification Cues for Mobile Devices. *Proceedings of UbiComp 2001,* pp. 240-246.

Heath, C. and Luff, P. (1996). Documents and Professional Practice: "Bad" Organizational Reasons for "Good" Clinical Records. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW'96),* pp. 354-363.

Hindus, D., Mainwaring, S. D., Leduc, N., Hagström, A. E. and Bayley, O. (2001). Casablanca: Designing social communication devices for the home. *Proceedings of the CHI '01 Conference on Human Factors in Computing Systems*, pp. 325-332.

Horvitz, E., Jacobs, A. and Hovel, D. (1999) Attention-sensitive alerting. *Proceedings of UAI '99 Conference on Uncertainty in Artificial Intelligence*, pp. 305-313.

Hudson, J., Christensen, J., Kellogg, W. and Erickson, T. (2002) "I'd be overwhelmed, but it's just one more thing to do": availability and interruption in research management. *Proceedings of the SIGCHI 2002 Conference on Human factors in Computing Systems*, pp. 97-104.

Hudson, S., Fogarty, J., Atkeson, C., Forlizzi, J., Kiesler, S., Lee, J. and Yang, J. (2003) Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study, *Proceedings of CHI '03 Conference on Human Factors in Computing Systems*, pp. 257-264.

Isaacs, E., Walendowski, A. and Ranganthan, D. (2002) Hubbub: A sound-enhanced mobile instant messenger that supports awareness and opportunistic interactions. *Proceedings of the CHI '02 Conference on Human Factors in Computing Systems*, pp. 179-186.

Kern, N., Schiele, B. and Schmidt, A. (2003) Multi-sensor activity context detection for wearable computing. *European Symposium on Ambient Intelligence (EUSAI).*

Kraut, R., Fish, R., Root, R. and Chalfonte, B. (1990) Informal communication in organizations: Form, function, and technology, *People's Reactions to Technology*, pp. 145-199.

Kuwabara, K., Watanabe, T., Ohguro, T., Itoh, Y. and Maeda, Y. (2002) Connectedness Oriented Communication: Fostering a Sense of Connectedness to Augment Social Relationships, *IPSJ Journal*, vol. 43, no. 11, pp. 3270-3279.

Laurier, E. (2001). Why people say where they are during mobile phone calls. *Environment and Planning, Society & Space*, pp. 485-504.

Lee, S. and Kenji Mase, K. (2002) Activity and location recognition using wearable sensors. *IEEE Pervasive Computing*, vol. 1:3, pp. 24-32.

Licklider, J. C. R. (1968) The Computer as a Communication Device. *Science and Technology*, April '68.

Mantyjarvi, J., Himberg, J. and Seppanen, T. (2001) Recognizing human motion with multiple acceleration sensors. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 747-52.

Markopoulos, P., Romero, N., Van Baren, J., Ijsselsteijn, W., De Ruyter, B. and Farshchian, B. (2004). Keeping in touch with the family: home and away with the ASTRA awareness system. *Extended Abstracts of the CHI '04 Conference on Human Factors in Computing Systems,* pp. 1351-1354.

Marmasse, N. and Schmandt, C. (2000) Location-aware information delivery with comMotion**.** *Proceedings of HUC2000 International Symposium on Handheld and Ubiquitous Computing*, pp. 157-171.

Marmasse, N. and Schmandt, C. (2002) A User-Centered Location Model. *Personal and Ubiquitous Computing*, vol. 6, pp. 318-321.

Marmasse, N. and Schmandt, C. (2003) Safe & Sound: a wireless leash. *Extended Abstracts of the CHI '03 Conference on Human Factors in Computing Systems*.

Marx, Karl (1867) *Das Kapital. Kritik der politischen Okonomie*, vol. 1.

McGrath, J. and Hollingshead, A. (1994) *Groups interacting with technology*. Sage Publications.

Milewski, A. and Smith T. (2000) Providing Presence Cues to Telephone Users. *Proceedings of the CSCW 2000 Conference on Computer Supported Cooperative Work*, pp. 89-96.

Minsky, Marvin (1980) Telepresence. *OMNI Magazine*, pp. 45-52.

Mitchell, T. (1997) Decision Tree Learning, in *Machine Learning*, McGraw-Hill, pp. 52-78.

Munguia, E., Marmasse, N., Intille, S. and Larson, K. (2004) MITes : Wireless Portable Sensors for Studying Behavior. *Adjunct Proceedings of UbiComp 2004.*

Mynatt, E., Rowan, J., Jacobs, A. and S. Craighill. (2001) Digital Family Portraits: Supporting Peace of Mind for Extended Family Members. *Proceedings of CHI '01 Conference on Human Factors in Computing Systems*, pp. 333-340.

Mynatt, E. and Tullio, J. (2001) Inferring calendar event attendance. *Proceedings of the IUI 2001 Conference on Intelligent User Interfaces*, pp. 121-128.

Nardi, B., Whittaker, S. and Bradner, E. (2000) Interaction and Outeraction: Instant Messaging in Action. *Proceedings of the CSCW 2000 Conference on Computer Supported Cooperative Work*, pp. 79-88.

Nelson, L., Bly, S. and Sokoler, T. (2001) Quiet Calls: Talking Silently on Mobile Phones. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 174-181.

Pascoe, J. (1998) Adding generic contextual capabilities to wearable computers. *Proceedings of ISWC 2nd International Symposium on Wearable Computing*, pp. 92-99.

Patrick, E. and Metcalf, C. (2001) Mediated Communication Between Extended Family and Friends: A Case Study. http://www.motlabs.com/user/papers/epatrick-crystam-vmc2001.pdf

Paulos E. (2003) Connexus: A Communal Interface. http://www.intel-research.net/ Publications/Berkeley/ 070220031047_135.pdf

Pedersen, E. and Sokoler, T. (1997). AROMA: Abstract representation of presence supporting mutual awareness. *Proceedings of CHI '97 Conference on Human Factors in Computing Systems*, pp. 51-58.

Picard, R. (1997) *Affective Computing*. MIT Press

Quinlan, J. (1993) *C4.5: Programs for Machine Learning*, Morgan Kaufmann.

Rabiner, L. and Sambur, M. (1975) An Algorithm for Determining the Endpoints of Isolated Utterances. *The Bell System Technical Journal*, vol. 54, no. 2, pp. 297-315.

Rabiner, L. and Juang, B. (1993) *Fundamentals of Speech Recognition*, Prentice Hall.

Rettie, R. (2003) Connectedness, Awareness and Social Presence. *Presence 2003, 6th Annual International Workshop on Presence*.

Rodden, T. (1991) A survey of CSCW systems. *Interacting with Computers*, vol. 3:3, pp. 319-353.

Romero, N., Van Baren, J., Markopoulos, P., De Ruyter, B. and Ijsselsteijn, W. (2003) Addressing interpersonal communication needs through ubiquitous connectivity: Home and away. *EUSAI 2003 – European Symposium on Ambient Intelligence*, pp. 419-429.

Root, R. (1988) Design of a Multi-Media Vehicle for Social Browsing. *Proceedings of the CSCW '88 Conference on Computer Supported Cooperative Work*, pp. 25-38.

Schilit, B. and Theimer, M. (1994) Disseminating active map information to mobile hosts. *IEEE Network 1994;* vol. 8, pp. 22-32

Schmidt, A., Michael Beigl, M. and Gellersen, H. (1998) There is more to context than location. *Proceedings of Workshop on Interactive Applications of Mobile Computing (IMC'98)*.

Schmidt, A., Takaluoma, A., and Mäntyjärvi, J. (2000) Context-Aware Telephony Over WAP. *Personal and Ubiquitous Computing*, vol. 4, no. 4, pp. 225-229.

Short, J., Williams, E. and Christie, B. (1976) *The Social Psychology of Telecommunications*. John Wiley.

Tang, J. and Rua, M. (1994) Montage: Providing Teleproximity for Distributed Groups. *Proceedings of the CHI '94 Conference on Human Factors in Computing Systems*, pp. 37-43.

Tang, J., Isaacs, E. and Rua, M. (1994) Supporting Distributed Groups with a Montage of Lightweight Interactions. *Proceedings of the CSCW '94 Conference on Computer Supported Cooperative Work*, pp. 23-34.

Tang, J., Yankelovich, N., Begole, J., VanKleek, M., Li, F. and Bhalodia, J. (2001) ConNexus to Awarenex: Extending awareness to mobile users. *Proceedings of the CHI '01 Conference on Human Factors in Computing Systems*, pp. 221-228.

Uiterwaal, M., Glerum, E., Busser, H. and Van Lummel, R. (1998) Ambulatory monitoring of Physical activity in working situations, a validation study. *Journal of Medical Engineering & Technology*, vol. 22:4, pp. 168-172.

Vallejo, G. (2003) ListenIn: Ambient Auditory Awareness at Remote Places, M.S. Thesis, Program in Media Arts and Sciences, MIT Media Lab.

Want R., Hopper, A., Falcao, V. and Gibbons J. (1992) The Active Badge Location System. *ACM Transactions on Information Systems*, vol. 10, pp. 99-102.

Weiner, M. and Mehrabian, A. (1968) *Language within Language: Immediacy, a channel in Verbal Communication*, Appleton.

Winston, P. (1992) Learning by Building Identification Trees, in *Artificial Intelligence*, Addison-Wesley, pp. 423-442.

Witten, I. and Frank, E. (1999) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann.