Intelligently Connecting People –

Facilitating Socially Appropriate Communication in Mobile and Office Environments

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

Von der Fakultät für Informatik der

Universität Fridericiana zu Karlsruhe (TH)

genehmigte

Dissertation

von

Maria Danninger
aus Starnberg

Tag der mündlichen Prüfung: 10. Juni 2008

Erster Gutachter:
Prof. Dr. Alex Waibel, Universität Karlsruhe (TH)

Zweiter Gutachter:
Prof. Dr. Cliff Nass, Stanford University

Dritter Gutachter:
Dr. Jacques Terken, Technische Universiteit Eindhoven

Abstract

As a mode of interactive information exchange, the phone is still unparalleled. As a way to initiate contact, however, it leaves much room for improvement. By making communication possible irrespective of context, the mobile phone has fundamentally changed our basic mechanics of social life. Poorly timed calls can disrupt social norms and break the flow of concentration. Turning the phone off to avoid undesirable calls at the same time poses the risk of missing an important one. A risk that is increasingly culturally problematic.

Currently, mobile phones provide passive channels, enabling but not facilitating communication. Along the iterative development of the Connector Service, this thesis will explore possibilities for technology to mitigate tensions between availability and interruption, with the goal to better negotiate between desirable communication and inappropriate disturbances.

Because mobile phones are not typically used in environments comparable to laboratories, but are rather used out "in the wild", instantiations of the Connector Service were evaluated in large-scale field experiments with real users engaged in real-life activities. This thesis will introduce a methodology that allowed testing new calling features in such large-scale field experiments, and report results from our investigations.

We will argue that there are two fundamentally different strategies when designing technologies that facilitate mobile communications, each putting different emphasis on design trade-offs such as who to give control to, how much mediation is appropriate, and what kind of contextual information is useful.

The first strategy was inspired by a virtual secretary metaphor: technology acts similar to a personal secretary that knows a person's availability and negotiates with the caller on the receiver's behalf. The ultimate goal is to free receivers from inappropriate or untimely interruptions, and reduce call management loads for both caller and receiver. A large-scale mobile phone field study with 90 participants revealed seven factors that play a role in constructing a mobile context. Among those, social factors were the most significant and consistent predictors of availability for mobile phone communication. We will review different versions of context-aware agents, based on contextual information from calendars, smart offices and smart phones. By deploying such context-aware virtual secretaries in real-world contexts, we could empirically show their capacity to enhance team performance and reduce inappropriate workplace interruptions. We will discuss major trade-offs when designing context-aware agents: mobility, privacy and a lack of control.

The second strategy puts the receiver back into the loop of negotiating with the caller the most appropriate medium and style of communication. The emphasis here, in contrast to previously outlined context-aware approaches, is on offering a greater choice of interaction style or modality to deal with untimely

interruption, and make the interruption non-interrupting, rather than actively preventing it. In an iterative process of design, we have developed two systems to explore this corner of the design space: One-way Phone and Touch-Talk. Both systems were deployed in a business context, and tested by dozens of employees of a large IT company in the Silicon Valley. Touch-Talk was widely appreciated by these busy industry employees, and used for actual business communications with real colleagues and customers over an extended period of time.

The research presented in this thesis will try to shed light on most important design trade-offs and strategies when negotiating between desirable communication and unwanted interruption, and has a significant contribution to make to the discussion of how technology can and should facilitate future mobile phone interactions: both by pushing the state of the art technologically and in socially intelligent design.

Zusammenfassung

Die Entwicklung des Mobilfunks und der Gebrauch von Mobiltelefonen in unserem täglichen Miteinander hat neue Kommunikationsformen geprägt, und unser soziales Leben wesentlich verändert. Kommunikation ist nunmehr abgekoppelt von einem Ort und somit oft auch einer Situation. Leute sind permanent miteinander verbunden, und immer und überall erreichbar. Handygespräche sind wohl nicht zuletzt deshalb allzu oft private Gesten in öffentlichen Räumen. Die neuen Praktiken und Handlungsweisen stören häufig bestehende Kommunikationsregeln, verletzten Privatespären, oder unterbrechen konzentriertes Arbeiten.

Mobiltelefone sind passive Medien, welche Telekommunikation lediglich ermöglichen, nicht aber aktiv zwischen Anrufer und Empfänger vermitteln. Derzeit entscheidet der Anrufer, abhängig von seiner gegenwärtigen Situation, wann und wie er eine andere Person kontaktieren will, wobei er die Erreichbarkeit der Empfänger bestenfalls raten kann. Die pragmatische Lösung, um unpassende Anrufe von vornherein zu verhindern, nämlich das Telefon einfach auszuschalten, birgt gleichzeitig das Risiko, einen wichtigen Anruf zu verpassen. Ein Risiko das in unserer Informationsgesellschaft, wo Erreichbarkeit oft vorausgesetzt und erwartet wird, zunehmend problematischer wird.

Diese Forschungsarbeit beschäftigt sich mit der Entwicklung eines *Connector Service*, welcher moderne Technologien einsetzt, um Telekommunikation intelligenter und sozialverträglicher zu machen. Das auserkorene Ziel ist, nicht immer dann wenn man erreichbar sein will oder muss, auch gleichzeitig unnötig gestört zu werden. Gewünschte Kommunikation und unpassende Stoerungen sollten sich nicht länger gegenseitig ausschliessen.

Die verschiedenen Versionen des *Connector Service* wurden prototypisch implementiert, evaluiert, und iterativ weiterentwickelt. Da Mobiltelefone im Alltag verwendet werden, kann man ihren Einsatz und ihre Verwendung unserer Meinung nach nicht representativ in laborähnlichen Umgebungen beobachten. Versionen des *Connector Service* wurden deshalb in Benutzerstudien und Experimenten mit echten Anwendern, in ihrem wirklichen Leben, evaluiert. In fuenf groß angelegten Studien haben hunderte Studenten und duzende Industriebeschäftigte eines grossen Softwareherstellers die hier vorgestellten Technologien verwendet und getestet.

Wir werden zwei grundlegend verschiedene Ansatzweisen diskutieren wie Technologie eingesetzt werden kann, um Kommunikation zu verbessern. Der erste Ansatz basiert auf dem Prinzip einer virtuellen Sekretärin, welche die Situation erkennen und - an Stelle des Empfängers - mit dem Anrufer Zeit und Art der Kommunikation aushandeln kann. Idealerweise könnte so ein System entscheiden, wann welcher Anruf durchgestellt werden sollen. Eine exakte algorithmische Bestimmung von Erreichbarkeit hat sich jedoch

als sehr komplex und schwierig herausgestellt. Menschen können immer noch viel besser entscheiden, ob ein bestimmter Anruf wichtig oder dringend genug ist, um eine andere Person in einer bestimmten Situation zu unterbrechen. Eine unserer Studien hat gezeigt, dass dabei soziale Faktoren (also ob, wie und mit wem wir gerade interagieren, private oder öffentliche Umgebungen, usw.) wichtiger sind als andere Arten von physischen oder kognitiven Aktivitäten.

Wir haben uns deshalb in erster Linie darauf konzentriert, dem Anrufer wichtige Informationen über die Situation des Empfängers zukommen zu lassen. Am Ende gewinnen beide. Der Anrufer muss weniger fürchten, zu einem ungünstigen Zeitpunkt zu stören, kann aber wenn nötig zu jedem Zeitpunkt einen wichtigen Anruf durchstellen. Der Empfänger wird seltener unterbrochen, da letztendlich weniger unpassende oder störende Unterbrechungen durch das System sickern. Wir konnten in verschiedenen Studien zeigen, dass dieser Ansatz tatsächlich unpassende Störungen am Arbeitsplatz verringern konnte, und Teams mit Zugang zu solchen Technologies Koordinations- und Kollaborationsaufgaben schneller lösen konnten. Als wichtigste Nachteile dieses kontext-sensitiven Ansatzes haben sich eingeschränkte Mobilität, verletzte Privatsphäre und Verlust von Kontrolle auf Seite des Empfängers erwiesen.

Die zweite Ansatzweise ermöglichte mehr Entscheidungsgewalt und Kontrolle für den Empfänger, der nun keine virtuelle Sekretärin mehr besaß, die zwischen ihm und dem Anrufern vermittelte. Das System hielt sich jeglicher Entscheidungen fern, ermoeglichte jedoch entgegennehmen von zuvor unpassenden Anrufen durch Bereitstellung einer groesseren Auswahl an Interaktionsformen oder -modalitaeten. Zum Beispiel durch Mischen von Sprache und Text, so daß man einen Anruf beantworten konnte, ohne selber laut zu sprechen.

Da diese Art von Technologien kein Kontextwissen mehr benötigen, sind sie auch nicht länger auf instrumentierte Umgebungen beschränkt, was Mobilität maximiert und mögliche Sorgen über Privatspäre minimiert. Aus diesem Grund konnten wir zwei Technologien welche auf Multimodalität anstelle von Kontext-Sensitivität setzten (One-Way Phone und Touch-Talk) in Zusammenarbeit mit einem großen IT Unternehmen in einem Kontext außerhalb der universitären Forschung testen und evaluieren. Unser Touch-Talk System wurde beispielsweise von Mitarbeitern dieses Unternehmens sehr positiv aufgenommen, und sogar über einen längeren Zeitraum hinweg für tatsächliche Gespräche mit Kunden und Kollegen verwendet.

Neben der Gegenüberstellung dieser beiden verschiedenen Ansatzweisen stellt die vorliegende Arbeit die gemeinsam verwendete Methodik vor, um neue Telefonietechnologien und -funktionen umgehend auf jedem Telefon zur Verfügung stellen zu koennen. In einem VoIP-basierte Ansatz erfolgte die Implementierung meist vollstaendig netzwerkbasiert, ohne zusaetzlich installierte Software auf den mobilen Endgeraeten. Mit Hilfe einer speziell konfigurierten und manipulierten Telefon-Vermittlungseinrichtung konnten Anrufe zwischen beliebigen Teilnehmern in Festnetz- oder Mobilfunknetzten serverseitig mitverfolgt, kontrolliert und aufgezeichnet werden. Erst dieser Ansatz erlaubte den reibungslosen Ablauf der gross angelegten Datensammlungen und Benutzerstudien.

Die Ergebnisse unserer Untersuchungen leisten einen massgeblichen Beitrag zu der Diskussion wie in Zukunft Technologie eingesetzt werden kann und soll, um Einschraenkungen gegenwaertiger Mobiltelefonie beheben oder zumindest verbessern zu koennen. Vielleicht wird bald "erreichbar" nicht gleichzeitig auch "unterbrechbar" sein. In einem interdisziplinären Ansatz stellen wir neue Technologien genauso vor, wie neue Methodiken, um solch neue Ansätze und Ideen implementieren, testen und evaluieren zu können.

Acknowledgments

The research covered by this thesis was largely funded by the European Commission through the Information Society Technologies (IST) priority in the Sixth Framework Programme under Project "CHIL - Computers in the Human Interaction Loop", and was conducted at the following institutions:

Universität Karlsruhe (TH) Research University · founded 1825

STANFORD UNIVERSITY

SAP

During the course of this thesis, I have been accompanied and supported by many people. This work would not have been possible without them.

First of all, I would like to sincerely thank Prof. Alex Waibel from the Department of Computer Science at the Universität Karlsruhe (TH), and director of the Interactive Systems Labs (ISL), for advising this thesis, and his support and trust in me during the course of this work. Moreover, he provided me with the unique opportunity to work in an extremely inspiring and international research environment, and unusual flexibility to pursue interdisciplinary research at different institutions, both in academia and industry.

Sincere thanks are also due to Prof. Cliff Nass from the Department of Communication at Stanford University who served as my secondary advisor and generously contributed numerous invaluable discussions and ideas. Cliff's experimental social science approach to studying Human-Computer Interaction, and the truly inspiring and interdisciplinary research with his Communication between Humans and Interactive Media (CHIMe) Lab, have tremendously influenced my research.

Moreover, I owe deep gratitude to Dr. Rainer Stiefelhagen, who leads the Computer Vision for Human-Computer Interaction research group at the Universität Karlsruhe (TH), for his continuous and reliable mentoring and support, numerous valuable discussions and ideas, and participation in field deployments, that has all helped me to concentrate on completing this dissertation.

I would also like to thank Dr. Jacques Terken from the Technical University of Eindhoven, who served as reviewer on my dissertation committee, and led our efforts towards service evaluation within the CHIL project.

I wish to acknowledge the contributions of my colleagues at the Interactive Systems Labs (ISL), both at the Universität Karlsruhe (TH) and Carnegie Mellon University. Special thanks are due to my fellow PhD students Keni Bernardin, Hazim Ekenel and Kai Nickel who provided various perception components and technologies for context-aware services, Gopi Flaherty who passionately programmed mobile phones, Robert Malkin who shared acoustic classification technologies, and Tobias Kluge and Lorant Toth who helped with integration and development efforts. Sincere thanks also goes to Silke Dannenmaier, Ines Himpel and Annette Römer for reliable and cheerful administrative assistance, and Dan Valsan for technical assistance.

My special appreciation goes to my colleagues and fellow PhD students from the CHIMe lab at Stanford, in particular Leila Takayama, Erica Robles, Jane Wang and Abhay Sukumaran, who collaborated on conducting large-scale user studies and evaluations, and with whom it has been an outstanding experience and pleasure to work with, professionally as much as personally. I would also like to thank all members on my 268 One-way Phone and Touch-Talk project teams, and the hundred Stanford students who took part in various user studies and evaluations.

Sincere thanks are extended to my colleagues and managers at SAP in Palo Alto, in particular Dr. Paul Hofmann, VP Research, Dr. Frankie James from SAP Research, and Dr. Jörg Beringer from Emerging Solutions, who have added a much appreciated real-life business perspective to my work. Moreover, I owe thanks to the dozens of SAP employees who volunteered to test my technologies in contexts where they were needed most: during their normal daily activities, and real business communications.

Many thanks also go to Kornel Laskowski for valuable and much appreciated last minute assistance with my thesis talk, and Wolfgang Gräther from Fraunhofer FIT for critical first to last page thesis proof-reading. Extended thanks go to Prof. Roland Vollmar, Prof. Tanja Schultz and Prof. Wolfgang Karl from my dissertation committee.

Special thanks are due to the entire CHIL consortium for great inspirations and great spirit: Universität Karlsruhe (TH) ISL and IPD, Fraunhofer IITB Karlsruhe, DaimlerChrysler AG, ELRA-ELDA Paris, IBM Prague, RESIT-AIT Athens, INRIA Grenoble, ITC-irst Trento, KTH Stockholm, CNRS-LIMSI Paris, TUE Eindhoven, UPC Barcelona, Carnegie Mellon University and Stanford University in the US.

Finally, my warmest gratitude goes to my family and friends for their constant support, encouragement and love. I dedicate this thesis to Genilson who supported and encouraged me over and over again from the first to the last day of my PhD studies (during which we mostly lived scattert across 3 continents, 9 cities and 15 different apartments).

Contents

Chapter 1

Introduction

1.1 Slamming Door vs. Ringing Phone

The following example of a door design can be found in an article on "Socially translucent systems" [26]:

> *"In the building where our group works there is a door that opens from the stairwell into the hallway. This door has a design problem: opened quickly, it is likely to slam into anyone who is about to enter from the other direction. In an attempt to fix this problem, a small sign was placed on the door: it reads, "Please Open Slowly". As you might guess, the sign is not a particularly effective solution.*
>
> *Let us contrast this solution with one of a different sort: putting a glass window in the door. The glass window approach means that the sign is no longer required. As people approach the door they see whether anyone is on the other side and, if so, they modulate their actions appropriately. ...".*

We believe that the door described here has much in common with mobile phone interaction. By making communication possible irrespective of context, the mobile phone has fundamentally changed our basic mechanics of social life. We now call people instead of places, and we can reach them anytime, anywhere. This unprecedented connectivity has tremendous benefits. However, at the same time, mobile phones have introduced the constant risk of interruptions. Callers place phone calls when convenient for them, but they have little knowledge about the receiver's situation. As such, a poorly timed phone call is like accidentally slamming the door into someone, or crashing into a meeting. People in front of closed doors face the same problems as callers: they can only guess whether the other party's context is convenient for communication.

Poorly timed calls can disrupt social norms and break the flow of concentration. Simultaneously, there has been an increase in mobile phone use during meetings, in social interactions, and in public spaces. Turning the phone off to avoid undesirable calls at the same time poses the risk of missing an important one. A risk that is increasingly culturally problematic. We have not yet developed successful strategies for dealing with the socially novel disruptions that mobile communication poses.

Figure 1.1: Examples of cell phone etiquette.

That's why our mobile phones would all need small signs (Figure 1.1) reading "Please Use Considerately". Other very useful signs could read: "Keep Business Private", "Drive Now - Talk Later", "Choose Ringtones You Won't Regret" etc. However, signs are no effective solution for phones either.

As a different solution, imagine a phone that had the equivalence of a glass window to the other end of the communication. Wouldn't that be useful? Callers could take more informed decisions about how and when to initiate contact. But this is not the only possible solution, there are many others. E.g., imagine phones that could behave more autonomously, and know when to ring, or when it is more appropriate to vibrate or mute and not disturb. Just like a door that would automatically know when it should remain locked (e.g. during a confidential meeting), or when it should swing open (image you approach the door with a huge packet in your hands).

Moreover, we could think of a fundamentally different approach to deal with limitations of current phones. New technology could introduce entire new ways of communicating via the phone, with the goal to make previously interrupting phone calls non-interrupting. What if you could answer a phone call without speaking aloud? Or if the caller could leave a voice message without the other person's phone ringing? Just like a door that does not swing, but slide, and thus cannot slam other people.

Even an everyday object as simple as a door needs careful design in order to be usable and useful for humans. In his best-selling book "The Design of Everyday Things", Norman describes how difficult it can be "even for the smartest among us ... to figure out whether to push or to pull, or slide a door" [54]. We can expect that designing new mobile phone interaction will involve much more complex *psychological* as well as *technological* investigations.

When merely thinking about a phone's glass window, many people will probably imagine some sort of self-disclosed precense information displayed next to the name in the address book, similar to the away status of many instant messaging applications. Immediately, a ton of questions come up. What information would be important that callers knew? How can we describe "availability", and how can it be best presented to callers? Where does the required information come from? Will it be too bothersome to set it manually, or can availability possibly be detected automatically? And, maybe even more importantly: how much information do we want to disclose, and how much control give to technology?

Mobile phones are undoubtedly powerful and convenient. The question is one of dealing with their own success. This thesis explores ways of better integrating the affordances of mobile technologies into human communication processes. Knowing that there may not be a universal answer or perfect solution, this thesis will try to shed light on what we believe are some of the most important limitations, design trade-offs and strategies when negotiating between desirable communication and unwanted interruption. And we believe that some of them are not obvious at first sight.

1.2 Introducing the Connector Service

Currently, mobile phones provide passive channels, *enabling* but not *facilitating* communication. The *Connector Service* is a technology that was designed to more appropriately negotiate mobile phone communications between people within the *Computers in the Human Interaction Loop (CHIL)* project[1] [17]. Within the CHIL project, we envisioned design solutions for computer services that assist people during daily interactions with others. Rather then expecting people to spend their time attending to technology, CHIL's goal was to develop computer services that are sensitive in attending to human activities, interactions, and intentions.

Along the iterative development of the Connector Service this thesis will explore possibilities for future mobile phone interactions to mitigate tensions between availability and interruption, to better negotiate between desirable communication and inappropriate disturbances.

We will discuss context-aware versions of the Connector Service that act similar to a virtual secretary that knows a person's availability and negotiates with the caller on the receiver's behalf, trying to reduce call management loads for both caller and receiver and free users from inappropriate interruptions. Other versions of the Connector Service followed a different approach. Rather than using context information to actively prevent interruptions, systems facilitate answering previously unanswerable calls by offering a greater choice of interaction style or modality to deal with untimely interruption.

Different versions of the Connector were developed in an iterative process of design. The next Section will discuss the research method in more detail.

1.3 Introducing the Research Method

If technology claims to solve a technological problem, such as tracking or identifying objects in images or video streams, or recognizing human speech in audio signals, it can be evaluated on manually annotated datasets. Performance is measured by pre-defined metrics, such as tracking errors or word error rates.

The Connector, however, tried to solve a social problem: improving communication between humans. In order to uncover the potential contribution of such an interactive system, it has to be evaluated by (a large enough number of) users, which has so far been a focus of the social sciences. Our research on the Connector Service therefore combined basic research on mobile communication with a series of design explorations and prototype systems. New approaches were implemented as proof of concept, but as well as fully functioning prototype systems, ready to be explored and tested in evaluations with real users,

[1] http://chil.server.de

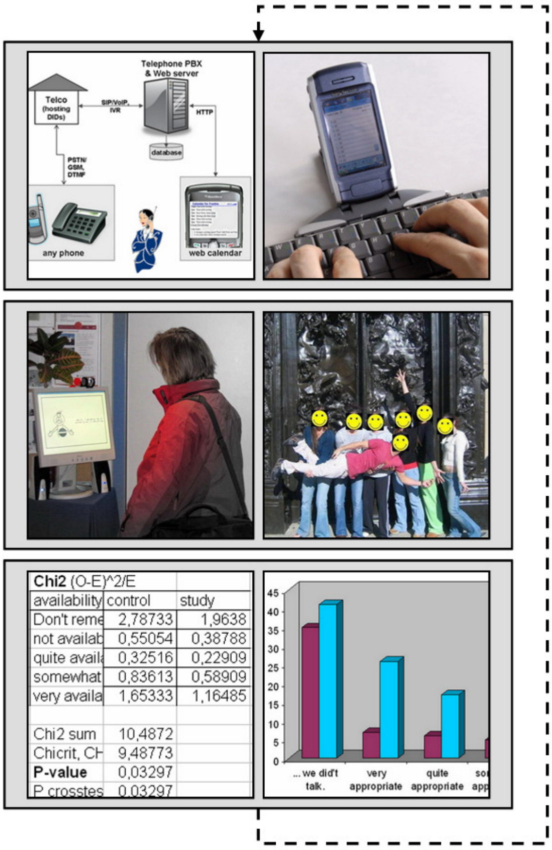


Figure 1.2: User-centered research method: Connector Prototype systems were evaluated in field studies and experiments with a large number of real users, engaged in real-life activities.

which happened in close collaboration with the Department of Communication at Stanford University. Figure 1.2 illustrates our iterative design approach.

Evaluation tests the usability, functionality and acceptability of an interactive system. There are many different possibilities to evaluate an interactive system. While some approaches are based on expert evaluation, our evaluations all involved actual user participation.

There are a number of different approaches to evaluation through user participation, such as empirical or experimental methods, observational methods or query techniques. Evaluation may take place either in the laboratory or in the field. Different versions of the Connector Service were evaluated through different methods, but always in the field and with real users. The following provides a brief introduction to user studies to help understand the challenge and benefits of studying interactive systems with real users. More detailed information on different evaluation techniques can be found in [22], or on the web[2].

[2] http://www.socialresearchmethods.net/, http://www.usabilityprofessionals.org/,

Large-scale evaluations in the field (vs. in the laboratory)

In comparison to controlled laboratory studies, field studies involve inherently different benefits and disadvantages. A laboratory can be equipped with recording or analysis facilities, such as audio/video recording, instrumented computers, etc. In addition, the participant operates in an interruption-free environment. However, exactly this lack of context accounts for an unnatural situation, as it never arises in the real world. It is especially difficult to observe several people cooperating on a task in a laboratory situation, as interpersonal communication is so heavily dependent on context. Interactions between individuals can only be observed in realistic situations.

As the Connector Service was designed to facilitate more intelligent communications between people, we had reason to be afraid that its actual usage and benefits would have been missed in laboratory studies. We therefore chose to evaluate the Connector Service in actual use in the field. Large-scale user studies were conducted with hundreds of college students, as well as in real-world business contexts with dozens of employees from a large IT company.

However, while field studies are much more realistic, they are at the same time much more difficult to carry out, especially when it comes to experimentation.

Measuring performance and user experience: Experimental vs. observational methods

There are a number of different evaluation techniques that must be carefully chosen when conducting an evaluation. A popular way is to gather information about actual use of a system is to introduce it into a natural environment and *observe users* interacting with it. However, user observation is challenging when study participants are not with you in the laboratory, and are using their own mobile phones to communicate with each other.

This thesis will introduce a methodology that allowed to immediately deploy new calling features from *any* phone, and permitted to monitor, control and record actual phone communications online from a server connected to the internet. This technique in addition to surveying and interviewing users allowed testing user experience and acceptance of different version of the Connector Service.

Observation and querying techniques provide mostly subjective information from the user or experimenter. This is why one of the most powerful methods is to use a *controlled experiment*, as this provides empirical evidence to support a particular claim. An experiment manipulates and measures variables in order to produce and compare different experimental conditions. Results from different conditions are statistically analysed in order to find differences that actually made a difference.

For example, a typical experiment design in our research would have been to split users in groups and ask each group to complete a predetermined task (such as solving a scheduling problem or a puzzle game) using phone calls only, during their normal daily activities. While groups in the study condition get access to a new calling feature, others in the control condition solve the same task using traditional calls only. Additionally, every participant filled in an online questionnaire about his attitudes towards using the system. Such a setup allows to statistically compare behavioral measures such as performance towards task completion, number and length of calls, call success rate, etc., and more attitudinal survey results on user experience.

To summarize, when evaluating different version of the Connector Service, we combined subjective evaluation methods in order to determine user experience, as well as controlled experiments to provide unbiased and comparable results. All studies took place in the field, with real users engaged in real-life activities.

1.4 How this Thesis is Organized

Part I will uncover benefits and limitations of current mobile phone use. We will motivate the need for more advanced technology that can empower humans to better negotiate between desirable communication and inappropriate disturbances, and review related work in the area. After introducing core design dimensions along which we might classify different approaches, we will introduce two fundamentally different strategies along which the Connector Service was designed.

Part II describes our explorations following the first strategy, one in which technology plays the role of an active mediator. This mediator makes use of context-aware technologies, and tries to reduce inappropriate or untimely interruptions. After introducing different designs and empirical evaluations of the Connector Service that implemented this strategy, we will discuss most important limitations and design trade-offs: privacy, mobility and locus of control.

These trade-offs will lead to a fundamentally different approach to the problem, which is discussed in **Part III**. One in which technology plays the role of a passive facilitator, that empowers humans through more flexible interfaces that afford greater choice of interaction style or modality. In the end, we will discuss lessons learned from real-life field deployments.

Part IV compares the two strategies along design trade-offs, and reviews results and main contributions of our investigations. We conclude by discussing design implications for mobile communication technologies appropriate for citizens of an ”always on world”.

The **Appendix** contains additional information on experimental design and statistical analysis that is referred to throughout this thesis.

Part I

Intelligently Connecting People: From Enabling To Facilitating Communications

As a mode of interactive information exchange, the phone is still unparalleled. As a way to initiate contact, however, it leaves much room for improvement. Technology can tackle this problem by facilitating rather then just enabling communications between sender (or caller) and receiver (or callee).

This Part of the thesis will discuss benefits and burdens that mobile phones have brought into our Always-On world, and motivate the need for technologies to better negotiate between desirable communications and inappropriate disturbances (**Chapter 2**).

Afterwards, we will discuss previous work towards designing novel mobile phone interactions, and related research on interruptibility and awareness (**Chapter 3**).

Chapter 4 introduces our approach when developing the Connector Service. We will discuss the two fundamentally different strategies along which the Connector Service was designed. The first uses context-awareness in order to more intelligently mediate communications. We will discuss design dimensions along which we might classify such context-aware communication networks: who to give control to, and what sort of contextual information is useful. We will argue that there is a second strategy, that goes beyond context-awareness, and uses more flexible interfaces that afford greater styles of communication, e.g. by mixing modalities.

Chapter 2

Mobile Communication In An "Always-On World"

2.1 Benefits and Burdens

Mobile communication technologies, especially the mobile phone, have radically changed the relationship between context and communication. Where before phone communications were tied to particular places (home, office, school, etc.) they are now tied to particular people. Where once locations bounded interactions, conversations now take place across vast geographical distances. This movement towards instantaneous, ubiquitous communication promises to empower individuals by delivering both unprecedented control and unlimited access. However, it also guarantees that we must learn to manage the associated burdens of being part of an "always-on" world [2].

The benefits to everyday life are evident. Being independent of location allows you to bring the communication device into the context where it is needed. People create for themselves small private spaces in the middle of large public ones, and they seamlessly blend social and professional engagements. Whether conducting business in cafes or receiving personal calls in meetings, having a mobile phone allows you to renegotiate with the spatial and temporal constraints of your surroundings.

These gains, however, do not go unchecked. Ironically, it is the contextual features that mobile phones disregard that represent the best guide to maximizing desirable communication while minimizing disruptions. Context provides common ground [10] from which we can assemble a shared understanding about the capacity for conversation. Fixed phones are associated to a *place* [32] and consequently a location and position in the social sense. This placement provided some ground rules for communication. With mobile phones, no such place-based guidance is present. We have not yet developed successful strategies for dealing with the socially novel disruptions that mobile communication poses.

It is perhaps for this reason that in the situations where established social protocols matter most, such as at the workplace, it is also most difficult to take advantage of this ubiquitous technology. For example, research on mobile phone use in the workplace finds that not more then half of all phone calls lead to an immediate conversation, the others fail to reach intended recipients [20, 71].

Figure 2.1: There are extended discussions about social norms of mobile phone use.

Callers place phone calls when convenient *for them*, but they know little to nothing about the receiver's situation. This corresponds with increases in mobile phone use during meetings, interactions with others, and in public spaces. It is hardly remarkable to see people pick up calls in meetings, making the self-contradictory statement, "I can't talk right now". Such mobile phone communications that occur in inappropriate contexts result in potentially disturbing others in close proximity and break the flow of concentration [12]. They can as well lead to invading personal and corporate privacy, and more broadly breaking social norms. As a results we have seen extended discussions about social norms of mobile phone use [39, 45, 51, 57]. (Figure 2.1)

2.2 Striking the Proper Balance

As a result, people develop compensatory tactics to prevent the limitations of phone calling. They turn the phone off to avoid an undesirable call, but run the risk of missing an important one. A risk that is increasingly culturally problematic. Blocking individuals at specific times is no solution, since it requires the kind of management that is so cumbersome as to be ignored completely by most users.

A current trend that can be observed in private as well as business settings is an increase in using asynchronous and text-based communication rather than traditional phones. People use instant messaging to negotiate availability for phone conversations [9], write emails about urgent issues, or listen to callers as they leave voice mails before deciding whether to pick it up. We constantly negotiate the connectivity trade-off between desirable communication and unwanted interruption.

Nevertheless, as a mode of interactive information exchange, the phone is still unparalleled. Synchronous voice conversation is still one of the richest ways in which two parties that are not co-located can communicate. It enables immediate confirmations and real-time negotiations, and has all the benefits of voice communication. Companies seem to have recognized these shortcomings of pervasive email business cultures, as some firms have experimented with mandate weekly no-emails days, e.g., Veritas[1]. As a way to initiate contact, however, the phone leaves much room for improvement.

Mobile phones are undoubtedly powerful and convenient. The question is one of dealing with their own success. The flexibility of the medium implies that there are few constraints that might guide their usage into predictable patterns. The following Chapter will review previous research on facilitating more intelligent or socially appropriate phone communications.

[1] www.marketwatch.com archive: "Fridays: Casual days without e-mail.".

Chapter 3

Related Work

This chapter reviews previous work towards designing novel mobile phone interactions, and related research on interruptibility and awareness. We will discuss related research, from estimating interruptibility in order to design context-sensitive systems, to smart phones that know when to ring, and when to mute.

3.1 Interruptibility and Awareness

A sheer breadth of work concerns contextually appropriate communication technologies and is articulated predominately in research on interruptibility and awareness.

Interruptibility

Interruptibility research focuses primarily on decreasing untimely interruptions by developing systems that manage trade-offs between the relative cost of interruption and the potential benefit of information delivered. Interruption strategies aim towards basic design guidelines of how and when a user should be interrupted. Many guidelines are based on findings from the social sciences, especially psychology and sociology, and human-computer interaction.

For example, Clark found in [11] that people in normal human-human language usage have choices of whether to allow interruptions. People have four different responses to interruption: (1) accept with full compliance, (2) accept with alteration, (3) decline, or (4) withdraw.

McFarlane presents four basic strategies for coordinating interruption in human-computer interaction: immediate (no control), negotiated (user control), mediated (agent control) and scheduled (agent control according to prearranged schedule) [48]. He found in an experimental multitask that user negotiated solutions caused the best overall user performance, but admits that agent-mediated solutions are an "attractive but controversial" alternative.

Based on results from a related study on agent-mediated communication, Katz warns against overhead costs related to user-controlled negotiating of interruptions, and advises people to be cautious that this overhead does not become more troublesome then the immediate interruption [41].

Other researchers [1, 13] could show that there are significant differences in interrupting people at different moments within task execution. These kind of strategies relate mostly to earlier findings from Miyata and Norman [50] who classify task execution in three phases: planning, execution and evaluation.

On the computer-science side, researchers have put extended efforts into automatically predicting people's in-the-moment interruptibility. Systems range from rule-based and user-driven [42] to sensor-based and system-driven [27] solutions. These systems appear across a variety of contexts ranging from the office [38, 55] to the home [52] and to the mobile user [5, 33, 34].

For example, Horvitz et al. have examined models based on calendar information, PC activity and audio and video data [36]. Fogarty and Hudson et al. have studied an office environment. They introduced an interruptibility model incorporating the activity of the user, emotional state, and social engagement. They used built-in laptop microphones, motion sensors and computer activity information to learn interruptibility with statistical methods [28, 27]. Ho and Intille [34] have presented a sensor-enabled mobile device, that uses wireless accelerometers to detect transitions between activities, and argue that delivering messages when the user is in transition between two physical activities (e.g. sitting to walking) is perceived more positively.

Awareness

Awareness - *"an understanding of the activities of others which provide a context of your own activity"* [23] - research constructs a person's availability for communication primarily in social terms. Researchers focus on the fundamental problem posed by technologies that offer none of the contextual cues that are generally part of face-to-face interactions. The goal is to create systems that allow the same richness and variety of interaction as face-to-face interactions, with distance no longer being an issue. Producing an environment which is as close as possible to "being there", e.g. by establishing audio and video channels between distant locations.

Hollan and Stornetta argue that systems must go "beyond being there", and provide evidence that some mediated forms of communication have much to offer to human-human interaction in ways that face-to-face communication never could [35]. Erickson and Kellogg [26] describe such awareness-enabled approaches "socially translucent systems".

A number of researchers have designed such technologies that provide contextual cues about the state of a desired communication partner (e.g., busy or away status), and thus facilitate socially-appropriate decision-making about conversational availability [5, 26, 56]. More examples of awareness-enabled phone will be described in the following Section.

3.2 Context-Aware Phones

There has been a significant amount of effort placed on adding context to phone communication. Schilit et al. have described context-aware communication as a subset of context-aware computing [61], and have suggested a two-dimensional space for classifying such applications. The two dimensions are based on a distinction between the "communication action" and "context acquisition" from manual to automatic

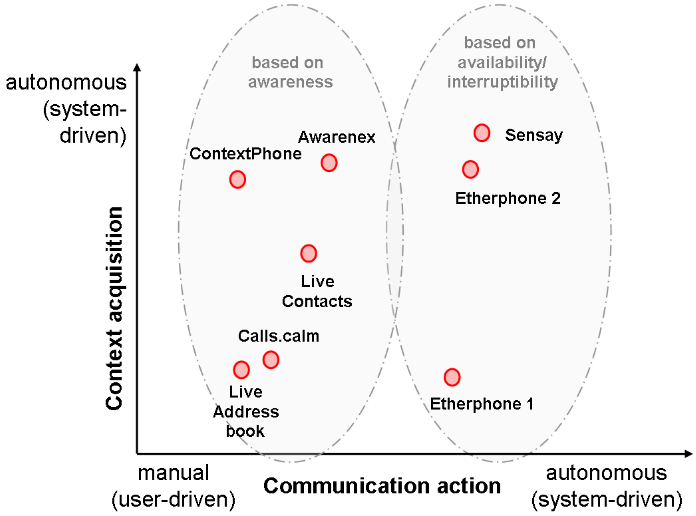


Figure 3.1: Context-aware phones along the design dimensions suggested in [62].

[62]. Measing that context can be entered manually, or sensed automatically, and the communication act itself (e.g. how and when to initiate contact) can be achieved manually by the user, or autonomously by an agent.

In the following, we will give examples and place systems along these two dimensions. Figure 3.1 shows an overview. We will distinguish between two different approaches, that broadly follow the previous notions of interruptibility and awareness. Awareness approaches require people to take final decisions, based on contextual cues provided by a system, while interruptibilily-based approaches rely more on a system deciding about the outcome of a call based on its knowledge of the person's availability.

Systems based on awareness

Brown and Randell [7], in their essay on context sensitive telephony, discuss the possibility of an automated agent that blocks calls on the behalf of users. They conclud that a better solution would be to provide the receiver's context information to the caller, to let the caller make a more informed decision about whether or not to initiate contact. The designs of a number of context-aware phones align well with this *user-driven* approach.

"ContextPhone" [55, 59] is a smart phone application running on the Series 60 Smartphone platform. ContextPhone enables users to share their context with others who use the same application. Cues that promote awareness are a composition of social ("with John"), cognitive ("busy"), positional ("city center"), interactional ("has been on the phone for 4 min") and communicative ("2 unread SMS").

A number of systems on mobile devices allow users to see others availability status in the address book, with an interface similar to today's instant messaging buddy lists. "Live Addressbook" [49] allows users to manually update their availability and best phone contact (e.g. office, cell, home) from a Palm device. However, they discovered that users never seemed to remember to change their availability state to

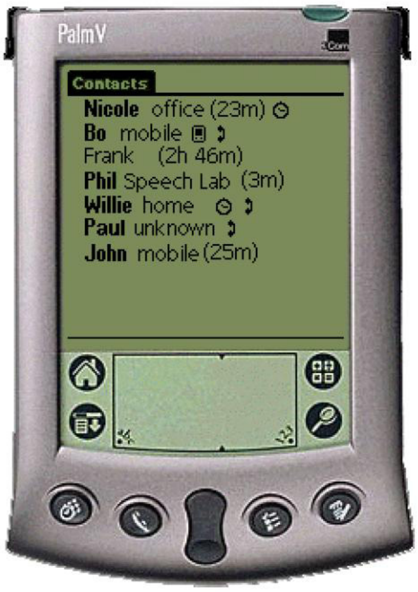


Figure 3.2: Screenshot of the "Awarenex" phone that shows context information directly in the address book (taken from [69]).

render the previous ineffective, and that availability is too irregular to be reliable. "Calls.calm" [58] also let receivers set their own context information manually and provided callers with information about the receiver's current situation. Both systems relied on callers to make reasonable choices of how and when to initiate contact. In "calls.calm", the receiver could specify caller's access to situation information.

"Awarenex" [69] can also suggest the most appropriate communication channel according to a simple set of rules. Moreover, the system automated context acquisition by using calendar information as well as desktop activity and phone usage. Working prototypes ran on Palm and RIM Blackberry platforms. Figure 3.2 shows a screenshot of the Awarenex phone, where one can see that colleague Nicole is currently in her office, has been idle on the PC for the past 23 minutes, and has a meeting scheduled in her calendar.

The "Live Contacts" system [70] also provides preferences for communication channels and is linked to a mainstream IM client. Awareness cues include basic location information (office, home, mobile), calendar information, IM status, and user-defined traffic-light-style availability indicators.

"Enhanced Telephony" [8] is a desktop-based design of an enhanced PC-phone, that was designed to study PC-phone user experience. Results suggest that the most valuable features of such a PC-phone may be those that help people manage their incoming calls.

In all of these systems, users must either manually update their availability state or context information is inferred automatically from sources such as login time, personal calendars, messenger status, idle time of computer input devices, and engagement in communication activities.

Systems based on availability and interruptibility

Other systems were built to act more autonomously and *system-driven*: e.g. the phone knows when and when not to ring. Such systems base their decisions on contextual features that describe a person's interruptibility, and that can be detected by current sensor technologies. Previous research suggests that a few key sensors placed in the environment, or on the device, may provide enough information to improve interruptibility detection [27, 33, 34].

The "Etherphone 1" system [67, 74] was an early context-aware application developed at PARC in the 80s, before the widespread adoption of mobile phones. It allowed users to tell the system where in the PARC office building they were located, e.g. by logging into a desktop computer interface, and would then route their incoming phone calls to the current location, manipulating the phone to play a distinct ring tone so that people could recognize their calls. "Etherphone 2" later integrated active badges in order to automatically detect people's locations. Automating Etherphone had a number of consequences. Being dependent on unreliable sensor information, the system sometimes rang phones in the nearby offices when people were walking down the active badge enabled hallway. Also, cases when call routing was not desired required user action, rather that being the default.

Mobile settings cannot make use of sensors in the environment, and thus pose additional challenges to both designers and developers. They must use sensors worn or carried by the user. Hinckley and Horvitz [33] have built sensors into a mobile computing device. They used an accelerometer to measure tilt, a touch sensor to detect whether the user was holding the device and an infrared proximity sensor.

"SenSay" [63] is a mobile phone that adapts to changing user states by manipulating ringer volume, vibration, and phone alerts for incoming calls. SenSay uses a number of wearable sensors including accelerometers, light and microphones mounted on the user's body to provide context information. This device was tested using a Wizard of Oz protocol simulating sensor behavior.

Chapter 4

Towards Designing More Appropriate Telephone Communications

Traditionally, how telephones worked is that a caller initiated the communication by dialing a receiver's number, and if the receiver picked up, the telephone channel allows both to talk and listen at the same time (see Figure 4.1). As we have seen in the previous discussion of related work, mediation technologies can exhibit a wide range of possibilities for more intelligently facilitating communication.

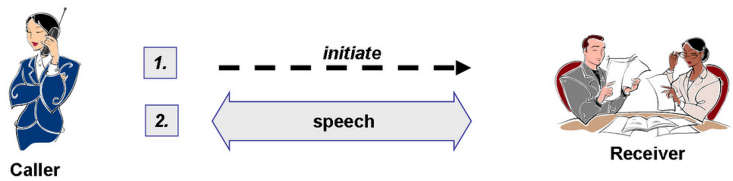


Figure 4.1: Traditional telephones allowed caller and receiver to talk and listen at the same time.

In this Chapter we will argue that there are two fundamentally different strategies when designing technologies that go beyond enabling communications and towards more appropriately facilitating or mediating mobile phone interactions: one facilitates the initialisation phase and is described in Section 4.1, the other augments the subsequent communication phase and is introduced in more detail in Section 4.2.

4.1 Augmenting Knowledge: Using Context-Awareness

When initiating a phone call, the caller can only guess the other parties availability. This became even more apparent with the introduction of the mobile phone, which is no longer associated to a location, such as someones office or home, which provides at least some place-based guidance or common ground for both caller and receiver. Currently, callers place phone calls when convenient *for them*, because they know little to nothing about the receiver's situation.

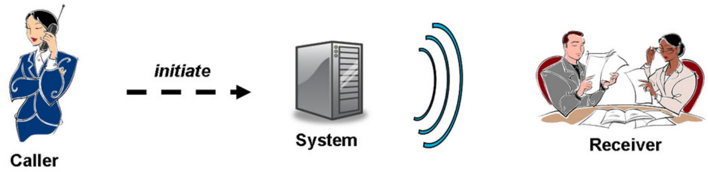


Figure 4.2: An intelligent agent augments knowledge (e.g. about the receiver's situation) in order to facilitate more appropriately timed communications.

A system that knows the receiver's context can be placed between caller and receiver to facilitate more intelligent decisions on how and when to initiate contact. Such systems act very much like a virtual secretary that uses contextual knowledge in order to help sender and receiver to connect more socially appropriately (Figure 4.2).

Using context-awareness is a widespread approach to better mediate time or place of a communication. Context-aware communication was previously defined by Schilit et al. as

> "*the class of applications that apply knowledge of people's context to reduce communication barriers*" [62].

The authors suggest a two-dimensional space for classifying such applications, based on a distinction between "communication action" and "context acquisition" from manual to automatic (see Figure 3.1). In the following, we will describe these two dimensions in more detail.

4.1.1 Locus of control: system or human?

Systems that make use of contextual information to actively intervene must choose to what extent to place decision-making responsibilities - the locus of control - with the technology, or with the humans. The locus of control describes the "communication action"-dimension, and might be controlled manually by the users or autonomously by the system. The resulting systems focus either on interruptibility, as it is judged autonoumusly by sensors, or awareness, as judged by humans.

In case control falls on the technology, the system uses contextual knowledge to actively and autonomously take decisions on behalf of people. Such an agent's behavior can be manually set by the user, e.g. by explicitly defining rules or self-reported feed-back. Unsupervised learning approaches take that burden off users by automatically and implicitly observing the user and learning from his actions and reactions. Machine-learning approaches, for example, forestall interruptions by screening potential communications based on inferences about the current state of the receiver. Or we can imagine a phone that automatically controls incoming call alerts, such as switching to silent, vibrate or ring. Or the system can even go a step further, and proactively initiate a call as soon as the context of both parties is convenient for communication. By assigning in-the-moment judgments to the technology, these systems decrease the receiver's need to disrupt their current context in order to decide whether or not to accept a call.

Technology-driven decision-making can take decision-making burdens of humans, and thus reduce unwanted interruption. Where it fails is in taking advantage of human capability for social judgment.

The awareness-approach tries to leave decision-making responsibilities in the hands of the humans by empowering them with context information so that they can better decide when and how to initiate communication.

An artificial agent can communicate relevant context information to users, and let them make sense of it ("glass window strategy"). E.g an agent could provide additional contextual information to the caller or receiver. So that the caller can take a more informed decision about whether or not to initiate contact, or so that the receiver knows what kind of conversation to expect before deciding whether to pick up the call (e.g. caller ID). Moreover, the agent can transmit various details of personal information, or simply collapse complex information to a single indicator, e.g. available or not?

We will argue later that availability for communication is extremely complex and hard to predict. Even if we had perfect sensors which provided an autonomous agent with information about the person's situation, the relationship between caller and receiver, and the urgency or importance of the particular matter, human common sense is still a quality that Erickson has described as "(at best) awfully hard to implement"[25].

4.1.2 Source of contextual knowledge

The "context acquisition"-dimension deals with the source of contextual knowledge. Systems that actively take part in the decision making process need to understand the current context, and know how to behave in it. This is because human behavior is context-dependent. An appropriate behavior makes reference to a *context* in which the behavior takes place. Especially if technology is involved in the decision-making process of what is appropriate, it needs to take context into account. E.g. smart phones need to know what context they are in in order to decide whether it might be appropriate to ring.

Context-aware systems need a description of the context, as well as information about how to behave in it. Useful contextual knowledge in the domain of mobile phone communication could be a composition of *(1)* the current situation of the two parties involved (e.g. location, activity, others nearby, private or public, physical engagement, mental engagement, capability of phone device etc.), *(2)* their relationship to each other (e.g. manager, co-worker, family, acquaintance, etc.), as well as *(3)* the nature of the communication itself (e.g. its urgency, importance, confidentiality, etc.).

The possible sources of contextual knowledge that can be exploited by an artificial agent to mediate a phone call are either *user-driven* or *system-driven*, *explicit* or *implicit*, and *planned* or *situated*.

Mediation technology can get input manually from users. Explicit user input can include manually updating a presence status. Implicit user-driven information can be acquired from personal calendars or can be inferred from other data sets, such as company directories specifying relationships between employees (Who is who's manager, co-worker, etc.).

As such, manual input is always user-driven, can be explicit or implicit, and is mostly planned. Relying solely on planned information, such as from calendars, can be an unreliable source of information for in-the-moment judgments. There are unattended appointments in everyone's calendar, as well as spontaneous meetings in the hallway. Moreover, manually updating a presence status is not an ideal solution. Users need to spend time attending to technology, instead of interacting with other humans. And the additional overhead may even become more troublesome then the immediate interruption.

Machine perception is the ability to use input from sensors (such as cameras and microphones) to deduce aspects of the world. A mediator can use contextual knowledge from perception technologies placed in the environment (e.g. smart rooms) or on mobile devices. E.g. an audio sensor could make use of a phone's built-in microphone to classify noise level of the environment, and adjust the ring volume accordingly. While such automated sensing of context allows acquiring in-the-moment information of the current situation, it can raise critical privacy concerns.

4.1.3 Design objectives

Context-aware systems, while focusing on a range of different problems, suggest a general set of design objectives, as summarized in [62]. These are (1) improving relevance by deciding when a communication is relevant to the person's current context, (2) minimizing disruption by deciding how and when to notify people that they have a communication, (3) improving awareness by deciding which information and mechanisms can help people make more intelligent decisions on when and how to communicate, or (4) reducing overload by filtering out communications that don't apply given the current context.

The overall goal is to actively mediate time or place of the communication in order to free people from unwanted or inappropriate interruptions. Different instantiations and prototype implementations of the Connector Service that aim towards augmenting knowledge through context-awareness are introduced in Part II.

4.2 Augmenting Modality: Using More Flexible Interfaces

The Connector Service was designed as a mediation service, facilitating more socially appropriate telephone communication. We have seen how context-awareness is a promising path towards facilitating more intelligent decisions on how and when to initiate contact. However, there are a range of alternative possibilities that may have potential to reach the same goal without being reliant on context and challenges and problems that might occur when buiding and deploying context-aware agents (we will discuss some major trade-offs in Chapter 9).

A different strategy is to make an interruption non-interrupting, by offering new paths of interaction that extend traditional phone usage. The traditional telephone network enabled two persons to talk and listen at the same time. As such, the telephone network supported synchronous, full-duplex, point-to-point voice communication. We have seen a number of alternative modalities of how two people can interact via the phone: sending text or multimedia data instead of voice (SMS, MMS), asynchronous voice communication (voice mail), half-duplex walkie-talkie style voice communication to multiple receivers (Push-To-Talk), etc.

Future directions could lead towards further integration of text, voice and multimodal data, and seamless switching between synchronous and asynchronous modes within the same communication. With wireless internet access available on more and more phones, and advances in voice-over-IP telephony, we will most likely see extended integration of traditional phone calls with other services available over the Internet.

The goal of such multimodal systems is to empower both caller and receiver to deal with the communications in a way that is most convenient given their current situation. The underlying technology should

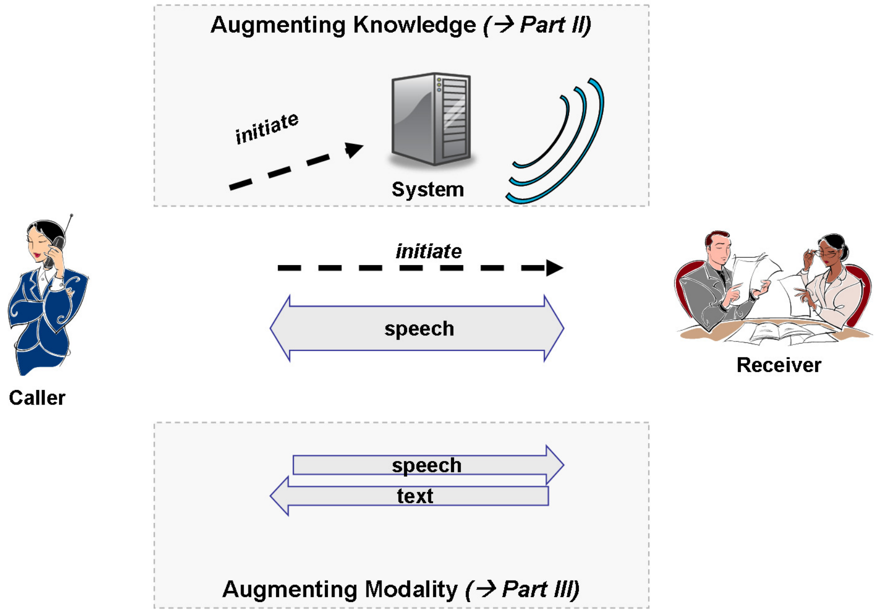


Figure 4.3: The Connector Service explored two fundamentally different strategies (augmenting knowledge and augmenting modality) of how to facilitate more intelligent mobile phone communications.

no longer be a limiting factor that lets receivers answer all phone calls in one and the same way. For example, one can imagine a technology that allows users to switch a call from speech to text, so that one can read what the other person has to say and type back, which would potentially allow people to answer previously unanswerable calls, e.g. in meeting situations. As such, the user is provided with tools to cleverly deal with communications in socially inconvenient situations. Systems following this strategy follow a more circumvention-based, rather than mediation-based approach, and aim at facilitating the communication, rather then mediating the interruption.

Instantiations and prototype implementations of the Connector Service that aim towards facilitation more appropriate communications by using more flexible interfaces are introduced in Part III.

In contrast to the previously discussed contect-aware approach, such systems emphasize "design" over "knowledge". Rather than understanding the situation and actively prevent the interruption or try to deflect it to a more appropriate time or place, this approach relies on building smarter, more flexible (i.e. multimodal) interfaces that afford greater choice of interaction style or modality. As such, the two different approaches align with opposing viewpoints of how humans and computers should interact in the fields of Artificial Intelligence (AI) and Human Computer Interaction (HCI) [72]. Figure 4.3 illustrates the two different strategies.

Part II

Augmenting Knowledge: Mediating Interruptions Through Context-Awareness

This Part describes efforts towards developing a context-sensitive and intelligent communication network that helps people make contact at the right time. Systems act similar to a virtual assistant by providing cues about the receiver's availability, and negotiating with the caller on the receiver's behalf. The goal was to reduce untimely or inappropriate interruptions on the receiver's side.

Different versions of the Connector that implemented this strategy were based on contextual information from calendars **(Chapter 5)**, smart rooms (**Chapter 7**) and smart phones (**Chapter 8**).

In addition, we discuss an exploratory field study on creating a scientific model for mobile phone availability (**Chapter 6**), and conclude by summarizing major limitations, challenges and design trade-offs of such context-aware systems (**Chapter 9**).

Chapter 5

A Calendar-Based Connector

Our first design, which was partly published in [18], exhibited a relatively straightforward approach: mediating phone calls by offering availability information from receivers' calendars to callers at the time of the call.

This relatively simplistic approach (reliant solely on pre-planned calendar information) was intended as an early proof-of-concept of the Connector idea. A major challenge was to implement a prototype system that could be deployed by a large number of users, so that we could study effects of such a system on user experience and team performance in a large-scale field experiment. We needed a system that could track and control people's interactions with each other during their normal daily activities, using their own mobile phones.

We will first discuss the idea of this calendar-based Connector system, then its implementation, and next results of deploying it in a field experiment with 90 users.

5.1 Mediating Personal and Group Calls

Employing a "virtual personal assistant" metaphor, the system encouraged senders to call only when receivers were available (according to their calendars), thus increasing the chances for successful connections and minimizing the risk of inappropriate interruptions or missed calls.

> Connector[individual]: *"Hello, this is Bob's Connector agent. Bob is currently busy. He will be available next time at 3:30pm. Please call back later. To still connect your call, you may press 1 now."*

In addition to offering availability information about individuals, the system featured information about entire groups, in which case the whole group would have a unique dial-in number:

> Connector[group]: *"Thanks for calling the reading group. The following persons are currently available: Anna, Peter, Jane. To call Anna, please press 1, for Peter press 2, for Jane press 3."*

5.2 Implementation as VoIP-based Network Service

This relatively simplistic approach was intended as an early proof-of-concept of the Connector idea. That's why we decided to explore this system in a mobile phone field study with a very large number of participants. Since mobile phone communications typically do not occur in contexts comparable to controlled laboratory environments, but are rather used “out in the wild”, we wanted people to explore this system during their normal daily activities, and using their own mobile phones.

5.2.1 Designing for large-scale user testing

Therefore, the system was designed entirely on the network, so that users could immediately send and receive calls from any phone capable of sending DTMF signals (i.e., touch-tones). No modification or installation on users' mobile handsets was required. Interactive voice responses (IVR)[1] were used to interact with users. IVR systems scale well to handle large call volumes.

People using the system were assigned a toll-free phone number, which was a Direct Inward Dialing number (DID)[2] hosted by a telephone company, and connected to our telephone system, a custom private branch exchange (PBX) network set up to act as Voice over Internet Protocol (VoIP) router. Asterisk™ (www.asterisk.org) was used as an open source software implementation of such a telephone PBX.

We configured the telephone system to facilitate interpersonal and group Connector calls. As calls were presented to the telephone system, the number that the caller dialed was given, so that the telephone server could decide who the intended receiver was and route the call to her actual cell number, stored in a local database. The Asterisk Gateway Interface (AGI) was used to add Java programming functionality, such as database access. Personal Outlook calendars were hosted by an Exchange Server, and could be easily accessed via a web interface.

This setup allowed calls connecting two parties in traditional PSTN or GSM networks to be (1) controlled, (2) monitored, and (3) logged by a server in a VoIP network.

5.2.2 In the network vs. on-board phone

The following list contains benefits of this network-based setup, for both users and researchers:

- *Usability*: people could use Connector features from their own mobile phones. IVR interactions are a familiar concept, as many business applications, e.g. call centers, telephone voting systems, etc., employ this technology.
- *Privacy*: people using the service did not need to share their actual phone numbers with other participants, which could have raised privacy concerns.
- *Low cost*: all hardware involved were two standard PCs with broadband internet access, one PC running the telephone system, and another one running the MS Exchange Server.

[1]IVR is a phone technology that allows a computer to detect voice and touch tones using a normal phone call. The IVR system can respond with pre-recorded or dynamically generated audio to further direct callers on how to proceed.

[2]Direct Inward Dialing (DID) is a feature offered by telephone companies for use with their customers' PBX systems, whereby the telephone company (telco) allocates a range of numbers all connected to their customer's PBX.

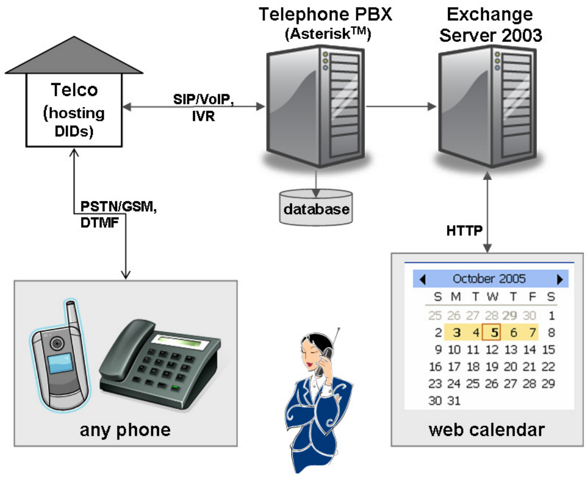


Figure 5.1: System architecture. All calls were routed through a custom telephone system. Outlook calendars were hosted by an Exchange Server, and could be accessed via a web interface.

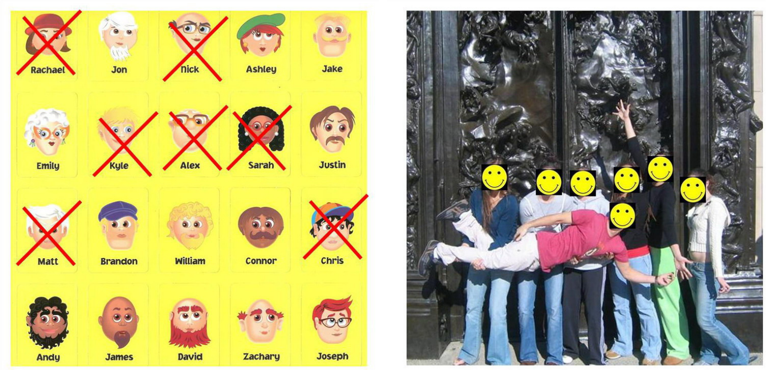
- *Easy maintenance*: problems could be resolved on the server (and as such as well remotely), rather then on each mobile device. E.g., our final user study was run at Stanford, while the server was maintained from Karlsruhe.
- *Simplicity*: software running on-board phones depend on the underlying mobile platform (Java ME, Windows Mobile, Symbian, Palm OS, etc.), while our network-based approach could be immediately deployed from any phone.
- *Flexibility*: Connector features could be activated on a per number basis, by simply changing a database entry. The feature was available with the next incoming call to this number.
- *Robustness*: the way the system was set up, output (including potential problems) with one call did not affect following calls.

Ultimately, locating the agent on board of the phone would enable direct access to phone behavior and the graphical display, which would provide a range of additional possibilities. E.g., custom graphical interfaces are much easier to use (and especially browse) than voice interfaces. For this calendar-based Connector system, people would likely have preferred to see directly in the address book whether another person was available.

However, if possible, a network-based approach is preferable for testing prototype systems in large-scale field deployments, especially in early stages of system design.

5.3 A Large-Scale Field Experiment

Over the course of two weeks, 90 students participated in a between-subjects two-stage field experiment at Stanford University to assess the system's potential contribution towards effective and enjoyable



(a) "Guess who puzzle" as collaboration task

(b) Picture of a successful group meeting at the Gates of Hell

Figure 5.2: 90 college students were split in teams to each solve a collaboration and a coordination task.

coordination and collaboration. This study was run in collaboration with the CHIMe lab at Stanford University.

5.3.1 Teams solved coordination and collaboration tasks

We divided participants into eight teams, each tasked with collaboratively solving a "Mystery Person" puzzle and then arranging a face-to-face meeting at the Gates of Hell, a well-known Rodin sculpture located on Stanford campus (see Figure 5.2).

The *collaboration task* asked participants to contact their teammates to exchange clues (such as facial features) in order to collectively identify, through a process of elimination, their team's Mystery Person. For the *coordination task*, group pictures taken in front of the sculpture served as proof that the team had successfully coordinated the meeting over the phone.

Both tasks were designed with some redundancy, so that approximately two thirds of the team members needed to contact each other in order to solve it. This allowed teams to solve the task even in cases when some team members did not cooperate. Teams were asked to submit results of the task as soon as they were finished.

All participants were asked to mark out, via an online calendar, all times when they would be busy during their participation in the study. We provided web-based access to Outlook calendars.

5.3.2 Study design

In a 2x2 study design, we divided the eight teams into four experimental conditions, two teams of 11 to 12 persons in each. The conditions differed according to team access to different Connector features, as defined in Table 5.1. All teams used specially allocated toll-free numbers to contact their team members.

Table 5.1: Overview of the 2x2 experimental design that tested the calendar-based Connector Service.

Experimental conditions (2 teams of 11-12 students each)	**Individual Connector feature**	**Group Connector feature**
Control condition (traditional calls only)	No	No
One-to-One Connector	Yes	No
One-to-N Connector	No	Yes
All Connector	Yes	Yes

1. The *control condition* featured regular phone calls only. All calls were immediately routed to the receiver.

2. The *"One-to-One Connector" condition* gave teams access to the individual Connector feature. The caller was encouraged to call only when the receiver was currently available according to his personal calendar.

3. The *"One-to-N Connector" condition* gave teams access to the group Connector feature. Each team was assigned one toll-free group number that members were asked to call. The caller could decide to connect the call to any team member who was currently available.

4. In the final *"All Connector" condition*, teams were allowed access to both kinds of Connector capabilities, and were assigned numbers for each person, and a group number to call the team.

Participants' attitudes towards Connector features were assessed via online questionnaires at the end of each task, asking participants to rate their level of frustration, team cohesion, interest in the task, etc. The telephone server automatically logged behavioral information about caller, receiver, call mode (One-to-One or One-to-N), time of call, length of each call, and audio recordings of all calls.

We statistically searched for differences between the different types of calls (traditional or Connector-mediated), and analyzed overall team performance as well as individuals' experience between the four experimental conditions. If not explicitly mentioned differently, all results reported here are statistically significant[3]. The different statistical techniques employed (Chi-square and ANOVA) are explained in more detail in the Appendix.

5.3.3 Connector-mediated calls were more successful

Overall, 710 calls were dialed using this Connector system, and participants spent 15 hours 38 minutes on the phone trying to solve the two tasks. The average length of a phone call was 1:19 min. Only two of the mystery person's could be identified, while six group pictures were taken.

Table 5.2 lists all calls that were dialed in the course of the 2-week long field experiment. Out of the total 710 calls, 267 (37.6%) were successfully picked up. There were 305 (43.0%) missed calls, the vast majority of them reaching the voicemail box, and only 16 calls were not answered at all. Calls specified

[3]by a confidence level of 0.05

Table 5.2: Connector-mediated vs. traditional phone calls in the field experiment.

Type of call	successful	voicemail	not picked up	discouraged	other
Interpersonal call	145	151	7	53	35
Group call	51	40	3	0	31
Traditional call	71	98	6	0	19
Total	*267*	*289*	*16*	*53*	*85*

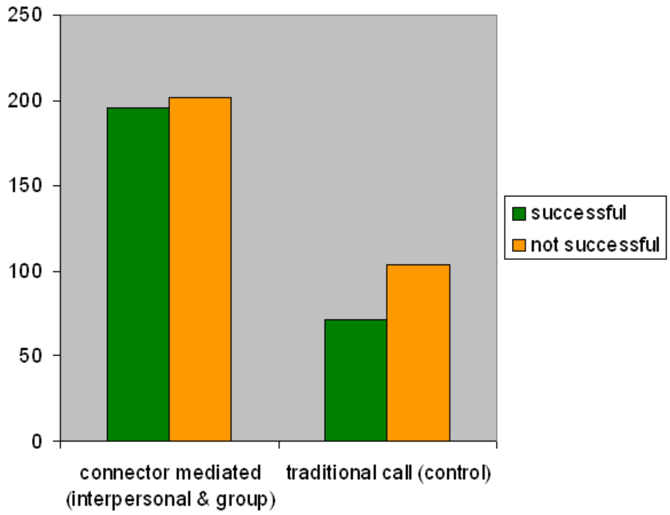


Figure 5.3: Distribution of success of phone calls with and without the Connector.

as "other" include calls that experienced technical difficulties, or where the caller had hung up before the call could be connected.

We assumed that people would be less available when they had marked out the time as busy in their calendars. This was confirmed when comparing the success rates of Connector-mediated calls (either interpersonal or group calls) to traditional calls (see Figure 5.3). Connector-mediated calls were only connected to person's who were available according to their calendar. Connector-mediated calls had a higher chance of being successfully connected than traditional calls. In a Chi-square analysis on frequencies of calls, this difference was almost significant ($\chi^2 = 3,78$, df=1, p=0.05).

When dialing an interpersonal Connector call the caller was informed in case the receiver was currently busy, which was the case in 14% of the time, and when the person would be available again according to her schedule. Such calls that were discouraged by the system were not counted as either successful nor unsuccessful, since the caller would ideally see this information right in his address book and not attempt to call in the first place. Group calls were only connected to persons who were available according to their calendar.

When using the interpersonal Connector feature to contact a team member who was currently busy, we left callers the option to connect the call anyways, and found that people did not shy away from using it. Approximately every third person called even though he was informed that the other person was currently busy. But only a single one of those calls was ultimately successful.

In the condition where people had access to both Connector features, and could freely chose which one to use, 86.6% of calls were dialed to specific persons and 13.4% were group calls.

5.3.4 Team performance and user experience

After finding out that information from personal calendars could be used to facilitate more successful phone calls, we were interested in whether this had an effect on how well the teams performed in completing the two study tasks: solving the Guess-Who puzzle and coordinating a group meeting at the Gate of Hell. We assumed that team performance towards task completion would depend on the number and success of actually dialed calls, and thus as well on people's motivation and attitudes towards the system they were using.

Collaboration Task

In the collaboration task, teams with both Connector features were the only ones who managed to find their mystery person, and thus finished the collaboration task in first and second places. It seams that access to both Connector features facilitated team performance in our study. This result is statistically significant given the odds of this outcome are .04[4].

There were significant differences in the overall number of calls dialed. Teams that had access to the interpersonal Connector features, among them the winning teams with access to both Connector features, dialed significantly more calls then their counterparts in the control condition, or with the group Connector only. Moreover, The number of dialed calls seemed to be correlated with user experience and attitude towards the system.

Regarding user experience, individuals with access to any of the Connector features reported greater interest and lower frustration than their counterparts in the control condition. We evaluated frustration based on an unweighted, averaged factor that combined annoying, complicated, and interrupting adjective items in the questionnaires. Based on 10-point Likert scales, the presence of the interpersonal Connector feature created significantly greater interest in the task[5]. This pattern, statistically significant, also results for presence of the group Connector[6]. Participants with access to the interpersonal Connector reported less frustration than their counterparts[7]. Similarly, participants with access to the group Connector were less frustrated than those without[8].

Coordination Task

Analysis of the subsequent coordination task showed that both teams with access to One-to-One Connector capabilities exhibited the best performance, able to complete the task with a minimum of seven members present. Overall, six teams had sent their group pictures via email, but the other four groups had less then 7 members in the picture.

[4]The odds that two teams in the same condition finish first and second out of eight teams is $(\frac{1}{8})(\frac{1}{7})+(\frac{1}{8})(\frac{1}{7}) = \frac{2}{56} = .04$, which is $< .05$, and thus statistically significant.

[5]with interpersonal Connector feature: (M=4.9, SD=2.2), without: (M=3.9, SD=2.2), (F(1,66)=4.03, $p<.05$).

[6]with group Connector: (M=5.2, SD=1.9), without (M=3.5, SD=2.3), (F(1,66)=9.77, $p<.01$)

[7]with interpersonal Connector feature: (M=4.3, SD=1.8), without : (M=5.3, SD=1.6), F(1,46)=10.32, $p<.01$

[8]with group Connector: (M=4.5, SD=1.9), without (M=5.1, SD=1.5), F(1,46)=7.24, $p<.05$

Individuals in the One-to-one condition also reported the greatest sense of team cohesion[9]. One-to-N Connector produced the opposite effect[10]. Teams with access to the group feature felt less connected to their team members compared to teams without the group feature. Moreover, cohesion was not significantly affected by the success covariate, p=.65; ruling out performance differences as the cause for perceptions of team cohesion.

5.4 Discussion: Social Dynamics and Mobile Phone Use

The user study deployed Connector features to teams and allowed to observe their impact on team performance and user experience for coordination and collaboration tasks. In both conditions, the presence of a feature had significant effect on each of these reactions. Access to both features enhanced collaboration while access to One-to-One service enhanced coordination. All in all, Connector-mediated calls had a higher chance of successful connection then traditional phone calls.

These differences confirm a tight relationship between social dynamics and mobile phone use. Collaboration and coordination require different interpersonal skills, thus calling for access to differential features from technologies that wish to better facilitate these processes. Moreover, the capacity for Connector features to significantly affect both user frustration and perceptions of team cohesion points towards the importance of appropriate design for mobile communication services.

Given the system's limited grasp of contextual information (reliant solely on participants' calendar entries), these results were extremely encouraging. All analyses supported the initial insight that transmitting information about the availability of receivers is a mechanism for enhancing mobile communications.

[9] with interpersonal Connector feature: (M=5.0, SD=1.4), without: (M=3.8, SD=1.8), $F(1,52)=4.63$, $p<.05$

[10] with group Connector fetaure: (M=3.3, SD=1.4), without: (M=5.0, SD=1.6), $F(1,52)=16.53$, $p<.01$

Chapter 6

Understanding Availability

The next logical step was to expand beyond using information from calendars as the only indicator of the receiver's context. Expanding on previous location and sensor-driven approaches (see Chapter 3), we investigated what factors might play a role in the construction of a "mobile context". Using a large-scale, week-long, field study of 90 college-aged mobile users, seven key dimensions of mobile availability were uncovered. We will discuss how these insights can be applied to design and theory of context-aware mobile communication technologies, and how they informed the subsequent design of the Connector Service.

6.1 A Large-Scale Mobile Phone Field Study

A large scale field study was conducted in collaboration with the CHIMe lab at Stanford University. 90 college-aged students at Stanford University (52 males, 38 females) participated in this study as part of an undergraduate course. All but three participants owned mobile phones, which they used for this study. Mobile phones were provided for the remaining three participants.

6.1.1 Detecting ”availability in the wild”

We collected three different availability measures from users: (1) prospective calendar entries, (2) in-situ experience sampling, and (3) open-ended retrospective questionnaires (Figure 6.1). By assessing availability with multiple methods, it was hoped to produce predictive and generalizable results.

Calendaring Participants were asked to mark out, via an online calendar, all times they would be busy during their participation in the study. We provided web-based access to Outlook calendars.

Experience Sampling The experience sampling phase of the study deployed a telephone server that randomly placed calls to participants throughout the day (between the hours of 10am and 10pm). The purpose of these calls (so called ”pings”) was to detect in-the-moment availability for communication; no actual phone conversations took place. The system randomly ”pinged” each participant four times each

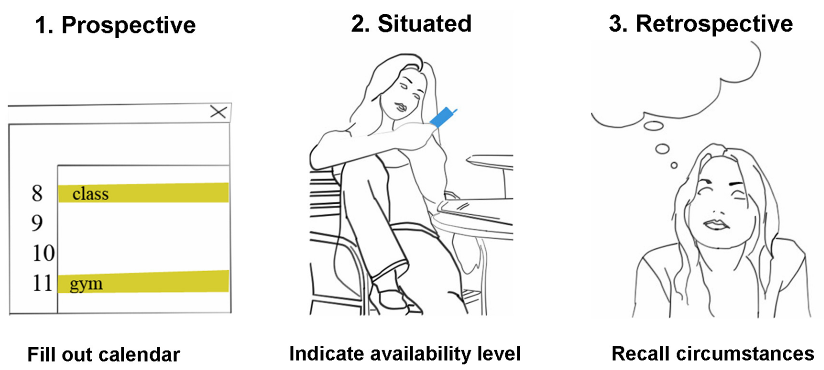


Figure 6.1: In a large-scale field study, availability was assessed with multiple methods.

day, for two consecutive days. If the call was answered, the system played a recorded voice prompt asking the receiver to indicate their current availability for a conversation. Participants responded by pressing 1 through 9 on their mobile phone keypad.

> Prompt of a server-generated phone call: *"Hello. This is the Comm 1 mobile service calling. Please press a number between one and nine on your phone to report how available you are for a phone conversation right now. One indicates not available at all. Nine indicates very available. After pressing the button, you may hang up the phone."*

Post-questionnaires At the close of each day, participants completed an online questionnaire where they provided open-ended descriptions about their situation and decision to communicate or not at the time of each call. These open-ended qualitative reports of activity and context generated a wide range of responses such as:

> - *"Watching a movie, by myself, not very busy but interested in the movie."*
> - *"I was at work, but I was packing up to head home for the night. I was not busy. My boss was in the room and asked what the call was about."*
> - *"I was talking with a friend, so it was a bit inconvenient. But I was still willing to quickly talk."*
> - *"My phone was on vibration mode on my desk, and I was coming back from the bathroom, and by the time I got there, it hung up.".*

These three different measures formed the basis of subsequent content analysis. Analysis was based on call logs correlated to calendar entries and open-ended responses.

6.1.2 Dimensions of "mobile context"

Seven dimensions accounting for people's attitudes and behaviors in mobile contexts were derived through iterative content analysis of open-ended questionnaire answers. Questionnaire answers were first associated with pings to ensure that subsequent analysis was based on some verifiable comparability between elements in the dataset. Any questionnaire entries that did not seem to correspond to real pings were eliminated.

This content analysis was carried out by multiple independent coders (annotators) on the open-ended questionnaire responses. A pilot test of the initial coding scheme was conducted with two coders and fifty randomly selected diary entries. Based on coder feedback, categories were refined. A second test was conducted with four new coders and a new set of fifty randomly selected diary entries. This second test revealed sufficient inter-coder reliability. The same coders then categorized the remaining data.

Seven contextual dimensions could be derived, and are described in more detail in Table 6.1: (1) Typicality, (2) the current activity, (3) physical, (4) mental and (5) social engagement, the (6) presence of others and whether (7) the location was a private or public space. Each dimension was reliable concerning intercoder reliability[1] as reflected in the Krippendorf's α scores in Table 6.1.

The latter three dimensions combined described the *social context*. Due to high correlation (social engagement factor loading=.85, presence=.92, private vs. public=.81), we collapsed these factors to avoid violating assumptions for subsequent regression analysis.

6.2 Predictors of Availability

We subsequently analyzed the power of each of the remaining five dimensions describing a person's mobile context (typicality, activity, physical and mental engagement and social context), and the calendar entries, to predict receiver availability. We distinguished between four different classifications of availability, from whether or not participants actually answered our call, to how available they reported to be.

Using the logged data from the experience sampling phase, we classified availability in four distinct ways.

- Answered phone provided a straightforward classification of whether or not participants answered our call.
- Extremely unavailable classifies the lower extreme of responses (37th percentile), by combining self-reports of extreme busyness (response=1) with no answer.
- Extremely available classifies the upper extreme (80th percentile) by combining self-reports of extreme availability (responses=8 or 9).
- General availability refers to self-reports ranging from 2 through 7.

We subsequently analyzed the power of each of these dimensions to predict receiver availability, using predictive regression models of each dimension.

[1]Intercoder reliability measures the extend to which independent coders reach the same conclusion, and is a critical component of content analysis. Intercoder reliability is reflected in the Krippendorf's α scores. More details can be found in the Appendix.

Table 6.1: Seven contextual dimensions of availability for mobile phone communication.

Dimension	Description	Intercoder reliability
TYPICALITY	"Choose how typical it would be for a participant to answer a phone call in their current situation?"	($\alpha = .66$, $n = 341$)
ACTIVITY CATEGORY	"Choose which of the following categories best describes the primary activity in which the participant was engaged at the time of the call. Please choose the one that best describes the activity. • *Basic* (e.g., eating, showering, sleeping, bathroom). • *Transportation* (driving, using public transport, cycling, walking) • *Required* (doing homework/research, at job, at meeting, in class, at sports practice) • *Alone/Personal* (Internet browsing, Entertainment Media, leisure reading) • *Social* (socializing with friends, partying, dining out, smooching)"	($\alpha = .83$, $n = 336$)
PHYSICAL ENGAGEMENT	"Indicate the degree of physical engagement required by the activity that the participant is engaged in (e.g., active use of ears, hands, mouth, eyes, legs)"	($\alpha = .66$, $n = 333$)
MENTAL ENGAGEMENT	"Indicate the degree of mental engagement required by the activity that the participant is engaged in."	($\alpha = .70$, $n = 333$)
SOCIAL ENGAGEMENT (-> Social context)	"Indicate whether or not the participant was engaged in a social interaction. A social interaction refers to the participant communicating or acting directly with others, whether or not those others are remote e.g. a cellphone conversation, working with a team on a problem set, etc."	($\alpha = .74$, $n = 292$)
PRESENCE OF OTHERS (-> Social context)	Indicate whether or not others are present in the immediate environment (i.e., capable of observing the participant) that the participant is not socially engaged with?	($\alpha = .90$,$n = 288$)
PUBLIC VS. PRIVATE CONTEXT (-> Social context)	Indicate whether the participant is in a private or public environment. Private environments are personal spaces (home, room, car), i.e., not public. Public environments are impersonal spaces (on a sidewalk, in a building), i.e., not private.	($\alpha = .74$, $n = 310$)

Table 6.2: Predictive regression model of typicality of availability.

Availability measures	*Typicality as pedictor of availability*		
	Model's power	β	p
Answered phone	$R^2 = .30$	1.13	$< .001$
Extreme unavailability	$R^2 = .35$	-1.22	$< .001$
Extreme availability	$R^2 = .31$	1.31	$< .001$
General availability	$AdjR^2 = .14$	.98	$< .001$

6.2.1 Typicality as proof of concept

It was not surprising that *typicality,* or how typical it would be according to social norms to answer a phone call in a given situation, significantly predicted whether or not participants chose to answer their phone. However, this result was important as proof-of-concept, since it provided evidence that situational judgments about someone's availability can be reliably derived from third-person reports based on reading others' anonymous diary entries.

Table 6.2 shows results of the regression analysis[2]. The R^2-value describes the goodness of fit of the regression model, in our case how much of the variability in the availability measures could be predicted by the measure of typicality, as extracted from content-analysis of open-ended questionnaires. The p-value suggests that typicality predicted all classifications of availability significantly better than chance.

6.2.2 Influence of activity on availability

Activity categories significantly distinguished differences between expected and observed availability, as was tested in a post hoc Chi-square test[3]. Someone's current activity, as described by our five availability categories, significantly influenced whether this person had picked up the call, or not ($\chi^2 = 26.50$, $df = 4, p < .001$). Figure 6.2 shows the frequency distribution of phone pick-ups across activity categories. Moreover, the current activity influenced all other three classifications of availability: extremely unavailable ($\chi^2 = 30.38$, $df = 4, p < .001$), extremely available ($\chi^2 = 26.55$, $df = 4, p < .001$) and general availability ($\chi^2 = 45.24$, $df = 28, p < .05$).

Table 6.3 shows a more detailed post hoc[4] Chi-square analysis that identifies exactly which categories contained a frequency distribution significantly different from random chance. Knowing whether someone was engaged in alone/personal activity was particularly useful for predicting mobile phone availability. Participating in social activities seemed to influence fine-grained levels of availability (i.e., general availability), while required and transportation activities influenced decisions of binary availability, i.e., available or not.

[2]Regression analysis can be used to compare relationships between continuous variables, or in other words whether a variable can be predicted by one or multiple others. More information on regression analysis can be found in the Appendix.

[3]Chi-square tests are used to compare whether two categorical variables are different (i.e. independent) from each other. More information on using Chi-square tests can be found in the Appendix.

[4]Analysing the data, after the experiment has concluded, for patterns that were not specified a priori.

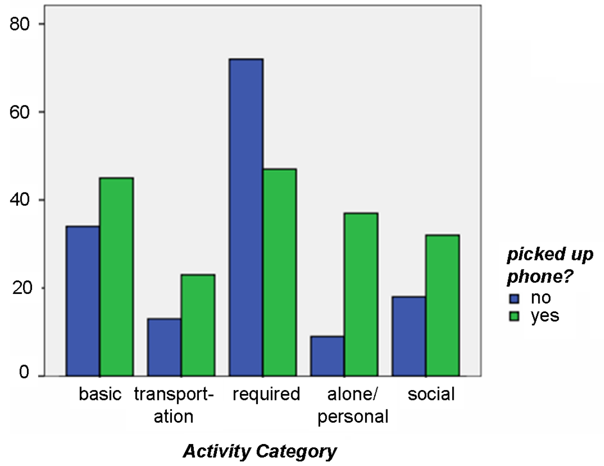


Figure 6.2: Frequency distribution of phone pick-ups across activity categories.

Table 6.3: Post hoc Chi-square results for each activity category.

	Basic	*Transpor-tation*	*Required*	*Alone/personal*	*Social*
Answered phone (df=1)	0.06, (n.s.)	3.86, p<.05	4.17, p<.05	20.86, p<.001	5.56, p<.05
Extremely unavailable (df=1)	1.08, (n.s.)	7.52, p<.01	4.30, p<.05	25.79, p<.001	2.87, (n.s.)
Extremely available (df =1)	1.80, (n.s.)	0.18, (n.s.)	0.84, (n.s.)	30.53, p<.001	0.71, n.s.)
General availability (df =7)	8.37, (n.s.)	9.11, (n.s.)	2.48, (n.s.)	11.00, (n.s.)	18.48, p<.05

These findings suggest that especially social activities influence availability. People choose to renounce engaging in mobile communication on behalf of those with whom they are already engaged.

6.2.3 Comparing social, cognitive and physical dimensions

A subsequent analysis used regression to weigh social activities against more cognitive and physical ones in terms of their ability to predict variance in availability. Table 6.4 not only identifies significant predictors of availability (the p-value in the table), but also provides weighted scores of each variable in relation to the others (standardized β), as well as the full model's predictive power (R^2) for each operationalization of availability. All the models reported here were significant. For example, we can read from the table, that calendar entries, social context and physical engagement were about as important when predicting whether a person would answer the phone, and that the combination of them could predict 15% of the variance in the data. For predicting extreme unavailability (picking up the phone and pressing 1), the

Table 6.4: Predictive regression models of social, cognitive and physical measures of availability.

Availability measures	*Availability predictors* Model's power	*Calendar entry* β	p	*Social context* β	p	*Physical engagem.* β	p	*Mental engagem.* β	p
Answered phone	$R^2 = .15$	-.33	.01	-.32	.01	-.33	.01	–	n.s.
Extreme unavailability	$R^2 = .12$	.36	.00	.23	.06	.46	.00	–	n.s.
Extreme availability	$R^2 = .13$	-.72	.00	-.50	.01	–	n.s.	–	n.s.
General availability	$Adj R^2 = .02$	–	n.s.	-.15	.04	–	n.s.	–	n.s.

current physical engagement mattered most. The percentage of variance in the availability measures that could be predicted may not sound huge, but was significant given the large level of noise in the data.

It was found that people's **social context**, the presence of others and whether they were in a private or public location, consistently and significantly predicted their availability. Most interestingly, this social factor was the only variable that significantly predicted all four measures of availability. This means that knowledge about the social context allowed us to not only predict whether the person would pick up the phone or not, but as well whether she would be somewhat available. Similar to the previous analysis, the social factor was the only variable significantly related to the more nuanced measure of general availability (picked up and pressed 2 through 7).

Calendar entries were another predictor of availability, indicating that there is a relationship between people's plans and their actual in-the-moment behaviors [66]. This confirmed our previous success when using calendar information to mediate mobile phone calls between people (see Chapter 5). However, plans alone were very weak predictors of availability. If calendar entries were the *only* predictor in the model, little variance could be accounted for in terms of: answering the phone ($R^2 = .02$), extreme unavailability ($R^2 = .03$), extreme availability ($R^2 = .05$), or general availability (Adj $R^2 = .01$). As shown in Table 2, predictions became at least twice as powerful when coupled with other situated dimensions of availability such as social context and physical engagement.

Physical engagement was another common predictor of availability, showing particular influence over extreme unavailability. But not particularly useful to predict nuances of general availability or situations where people are extremely available. This demonstrated the limitations of sensing physical engagement in activities (e.g., playing sports as measured by accelerometers).

However, **mental engagement** was not a significant predictor in any of the models. This may be an indicator that this aspect of context might be better measured via behavioral methods (like secondary task reaction time). Possibly, being in mental states of flow [12] might not be as important to availability as the other variables in these models. Moreover, purely cognitive models of availability could be insufficient for understanding the social and physical aspects of mobile phone use.

6.3 Implications for Design

This Chapter discussed investigations into the nature of people's decisions about their availability for conversation. A series of basic dimensions accounting for people's attitudes and behaviors in mobile contexts were derived.

6.3.1 The power of social context

Among those, the social context, such as interacting with others, was the most consistently significant predictor of availability. Social factors were the only ones that were significantly related to the more nuanced measure of general availability, and thus could not only reliably predict whether someone would pick up the phone or not, but as well how available the person reported to be. Such information could be used in order to decide whether or not a particular call is important or urgent enough to be put through.

The social context is often overlooked in sensor-based approaches, that try to model a person's physical activity with accelerometers, PC activity information, or location information. Our findings back up previous research regarding the usefulness of speech detection for predicting office worker's availability for interruption. Fogarty et. al obtained their best results when additionally using laptop microphones to count the number of human voices in close proximity to the user additional to other sensors [27].

The current activity, physical engagement, and planned calendar information were also useful predictors of whether one would pick up a phone call.

6.3.2 Predicting availability is difficult

Even with this predictive model, we should not forget that reality may be much more complex. In the experience sampling phase of our study, participants were pinged by a server, and indicated their level of availability by pressing between 1 and 9 on their keypad, but were not involved in actual communication. In reality, availability will likely vary based on the particular instance of communication, including nature, urgency, and relationship with conversational partner.

An ideal mediator understands which calls to facilitate at what time. It would both free the receiver from unwanted interruptions and reduce the decision-making and call management loads for both sender and receiver. Achieving this goal requires a technology that accurately models the receiver's availability, the importance or urgency of the call, and the relationship between the two parties.

However, even with state-of-the-art technologies that reliably detect contextual information like level of activity, accurately predicting availability remains a very difficult task. The best result we could achieve in our efforts to use machine learning techniques to predict people's availability for communication was 58% accuracy on a scale of one to four [16].

This is perhaps not surprising, since it has been shown that even *humans* show variations when judging people's availability in common office situations [40]. Johnson and Greenberg have asked a large number of people to look at snapshots from typical office situations and judge how available those persons were for interaction. People displayed very different assessments of availability when interpreting particular situations.

6.3.3 System or human: who decides?

Accurately and automatically predicting availability will most likely be error-prone for all but the simplest situations. This should be taken into account when systems try to automatically infer a measure of availability (a sort of "availability-meter"), instead of displaying what the person is actually doing. Special care is necessary when mistakes by the system would be critical, e.g. for business communications.

In the following, we will introduce a subsequent version of the Connector Service. This system passed on relevant contextual cues to callers, while ultimately leaving the final decision of how and when to initiate contact to the humans, based on their judgment of the importance and urgency of the call

Chapter 7

Connector In Smart Offices

The next version of the Connector was informed by insights from studying people's availability for communication, as discussed in the previous Chapter. It extended its knowledge from planned calendar information to situated contextual knowledge that had proved useful in forecasting one's availability, such as social engagement and current activity. This required a more instrumented environment - able to sense situations, recognize people and track activities.

Employing the Connector in a "smart office" environment was motivated by this need, as well as the realization that people in front of closed office doors face the same problems as callers: they can only guess whether the other party's context is convenient for communication (Figure 7.1). Many communications at the workplace - via the phone as well as face-to-face - occur in inappropriate contexts, disturbing meetings and conversations, invading personal and corporate privacy, and more broadly breaking social norms.

This section describes a context-aware virtual assistant for the workplace, that mediated both phone and in-person interactions, and was published in [19]. This virtual assistant used perception technologies installed in smart offices to automatically detect the current situation. The virtual secretary then broadcasted this information to potential callers and visitors in order to inform better human decisions about how and when to contact, with the goal to free office occupants from inappropriate interruptions. The system provided a web interface that allowed browsing most recent and past office activities.

This virtual secretary was installed in three office in our Karlsruhe research lab. A senior researcher tested the virtual secretary over a period of several weeks, during which this digital assistant mediated all his actual phone calls and in-person office meetings. We wondered whether the Connector as virtual secretary could help to reduce inappropriate and untimely workplace interruptions, e.g. during meeting or phone conversations.

7.1 Perception Technologies in Smart Offices

Detecting situated contextual knowledge required an instrumented environment - able to sense situations, recognize people and track activities. This Section will introduce perception technologies that were

Figure 7.1: People in front of closed office doors face the same problems as callers: they can only guess whether the other party's context is convenient for communication.

installed in three smart offices in our research lab. These technologies could automatically track activity in smart offices, such as whether the person was currently present, working by himself, or involved in a meeting or phone conversation. Moreover, the system detected when someone was standing in front of the office door, and could automatically recognize all members of the research lab.

7.1.1 Detecting office activity

Detecting activities could imply tracking all users in the office, characterizing their interactions, analyzing their speech, and any number of other complex perceptual procedures. We have previously presented a system that used a multi-level Hidden Markov Model to recognize a number of typical office activities, such as paperwork, discussion, meeting, etc [73]. For this system, offices were equipped with cameras and omni-directional microphones. Since this system required offline processing, we have developed a simpler vision-based version that worked in an online fashion.

The cameras installed in each smart office were used to detect activity in special regions of interest - around the work desk, by the door, or where visitors usually sit. Activity detection is done by adaptive background modeling using Gaussian mixture densities, as described in [65]. It is assumed that most of the foreground segments are caused by moving people, with adaptation making sure that displaced objects such as chairs or notebooks are slowly integrated into the background. If a certain amount of activity is observed over a fixed period of time, the presence of a person is hypothesized. Parameters to control background adaptation and sensitivity of the different regions were adjusted by hand for each office.

Figure 7.2 shows the camera views of two smart offices, with the office occupants present. A red bounding box indicates that activity was detected in the region of interest. Figure 7.3 shows snap shots of different situations in the single desk office of a senior researcher.

This rather simple but robust system could reliably run 24/7, and was used to detect basic social situations (see Section 7.1.3): whether the office was empty, if the occupant was alone, or in a face-to-face meeting with others.

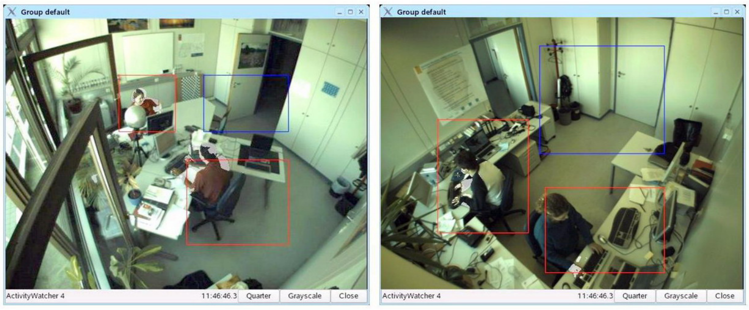


Figure 7.2: Camera snap shots from two of the smart offices with occupants present. Red bounding boxes indicate activity in areas of interests - around the work desk or where visitors usually sit.

It turned out to be most difficult to detect meetings situations, as our system required the visitor(s) to remain in ”meeting areas”, by the door or around the meeting table. The system could for example not correctly recognize a meeting when the visitor stood to close to the office occupant, e.g. right by the work desk. In order to distinguish situations where e.g. the visitor stands next to the occupant by the work desk, the system would need to actually track and distinguish single people, such as in [6], which is a much harder computer-vision task, and not within the scope of this work. Since the technology relied on background adaptation, temporarily unreliable information could be caused by changing lighting conditions, darkness or motionless people. However, we will see that even given these limitations the system performed very well in a later evaluation (see Section 7.1.4).

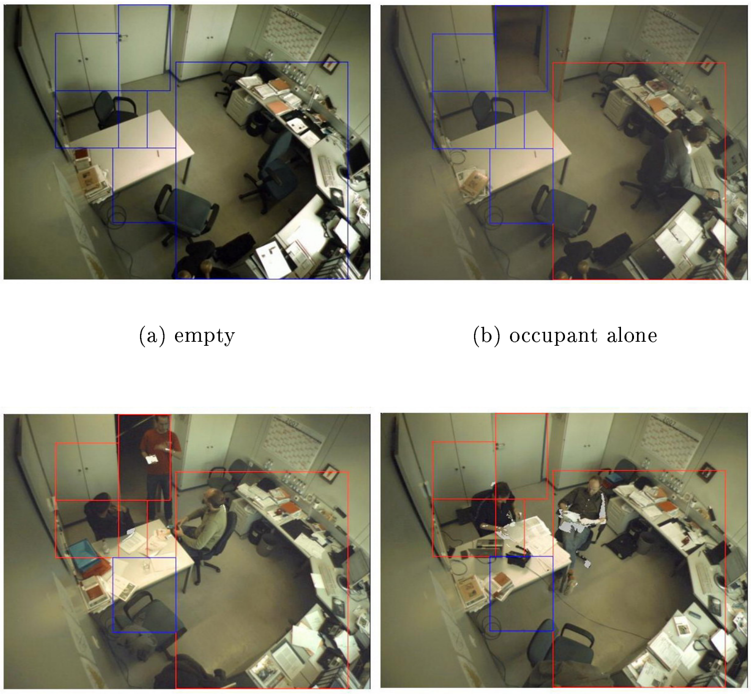
7.1.2 Detecting and identifying people

In order to detect and identify people in front of the office door, a small webcam was installed just outside of the smart office. This face recognition system was developed by Hazim K. Ekenel and Lorant Szasz-Toth.

The face recognition system used a local appearance-based algorithm that extracted features using discrete cosine transform, and is described in more detail in [24]. To be able to also distinguish unknown persons, it was necessary to enhance this system to open-set face recognition. We therefore extended this system to accurately identify all members of our research lab, approximately 15 persons, and distinguish them from unknown persons, using Support Vector Machines [68].

The system worked very robustly, and could handle varying illuminations and face sizes (Figure 7.4). A standalone evaluation of the face recognition system with five known and 20 unknown persons reported 87.2% accuracy on a per-frame basis, and as high as 99.5% accuracy when accumulating frames over time [68].

People in this set-up looked directly into the camera. When using the system for the virtual secretary, we made sure people looked into the camera by placing the camera on top of the screen from which the virtual secretary greeted visitors, which will be described in Section 7.2.2.



(a) empty (b) occupant alone

(c) occupant in meeting (d) occupant in meeting

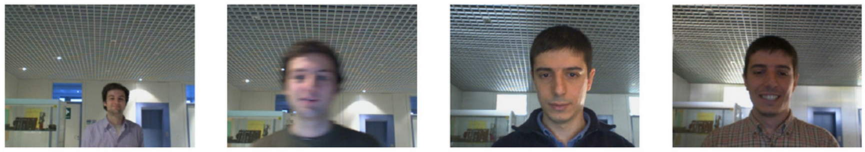
Figure 7.3: Camera snap shots show a third office in different situations. Red bounding boxed indicate activity in areas of interest.

7.1.3 Modeling office activities

We will now describe how observations from perception technologies were used to model the current office situation. Detecting activity in specific regions of interest was used to infer the current social situation: whether the person was currently present and if the person was alone or in a meeting situation. In case a person was recognized at the office door shortly before the start of a meeting, the system hypothesized that the meeting would be with this person. Moreover, since all calls were mediated by the virtual secretary (Section 7.2.1), the system additionally knew when the office occupant was speaking on the phone.

With that knowledge, a rule-based system was used to determine the most likely office state. Possible transitions between office states are illustrated in Figure 7.5. A smoothing function managed transitions between office states. This was necessary in order to deal with unreliable information from perception technologies. Office states were assigned confidence measures. Confidence measures increased over time to prevent transitions to unreliable states.

Allowing too easy transitions could cause the system to observe state changes when in reality there were none (false positives), such as changing lighting conditions. However, when transitions were too difficult, the system failed to observe a new state (false negative), such as quickly leaving the office to get a coffee.



(a) Artificial light, far away (b) Artificial light, motion blur (c) Daylight, brighter (d) Daylight, darker

Figure 7.4: An open-set face recognition system could recognize everyone in our lab (approx. 15 persons) and distinguish them from unknown people (images taken from [68]).

Figure 7.5: This finite state machine illustrates office activities and possible transitions between them.

The only exceptions were transitions to and from phone states. Beginning and end of phone conversations could be detected accurately by the telephone server, and became active immediately.

7.1.4 Evaluating technology performance

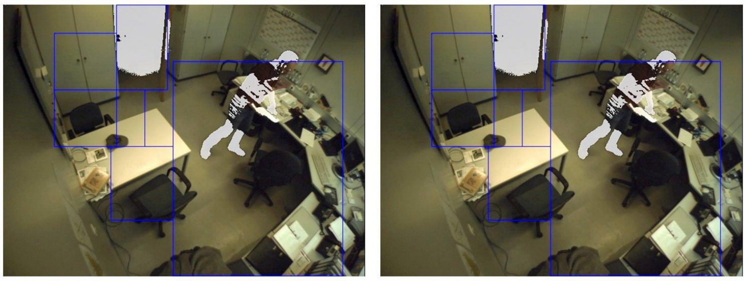
We evaluated the performance of the perception technology that tracked activity in one of the smart offices in our research lab, to decide whether the current situation was ”empty”, ”alone”, or ”meeting”. This was the office of a senior researcher, where we intended to later on test our system on real interactions. An automated script recorded the current office situation as detected by our system. Simultaneously, a snap shot of the camera image was taken, so that a human judge could afterwards easily assess whether the situation had been classified correctly.

Probes were collected every 10 minutes from approximately 8am to 6pm for 8 consecutive days. A total number of 443 data points were collected. During this time the office was empty in 54.7%, in 33.7% was the occupant alone, and in the remaining 11.4% a meeting was taking place in the office. Overall results showed that the correct office state could be determined in 90.6%.

Table 7.1 shows that results varied only slightly across different activities. The system could detect 90.8% of situations when the office was empty. When the person was present, and alone, the system correctly

Table 7.1: Recognition results of the office activity detector.

Description	*Recognition rate*	*False positive rate*	*Percentage of data*
Office empty	90.78%	13.5%	54.7%
Alone	91.1%	4.4%	33.7%
Meeting	88.0%	22.2%	11.4%



(a) snapshot from right after entering the office.

(b) snapshot from a meeting outside of typical meeting area.

Figure 7.6: Problems occured during transitions from one state to the other (a), or when meetings took place outside off typical meeting areas by the visitor table or the door (b).

detected the situation as alone in 91.1%. Very rarely did the system detect that the person was alone, when he actually wasn't (false positive rate). Detecting meetings was most challenging. Meetings could be recognized in 88.0%. However, it also happened that the system detected a meeting, when the person was actually not in a meeting, but alone (22.2% false positive rate).

Some of the mistakes occurred during transitions between states, at the very beginning or towards the end of an activity. As discussed previously, a delay was necessary before a new state could be reliably detected. Detecting meetings was especially troublesome, since meetings could only be detected if the visitor remained in common meeting areas by the door or meeting table. In case the visitor moved into working areas, e.g. by the work desk, the meeting could no longer be detected. Figure 7.6 shows snap shots of situations in which problems were likely to occur.

Given the relative simplicity of the perception technologies involved to detect activity in offices, these results were very encouraging. In the following, we describe how the virtual secretary mediated contact attempts in smart offices.

7.2 A Virtual Secretary: Mediating Interruptions in Smart Offices

We developed a virtual secretary that mediated phone calls as well as face-to-face interactions in smart offices.

7.2.1 Mediating phone conversations

The goal of the virtual secretary was to add context to phone communication to let the caller make a more informed decision about how and when to initiate contact. All calls to the office phone were directly routed to the virtual assistant. The virtual secretary informed callers about the receiver's current situation at the time of the call, e.g. whether the person was currently present, whether a meeting was taking place and with whom, or whether the person was engaged in another phone conversation. Since callers are differently related to the office occupant, the level of detail of disclosed information depended on whether the caller was known and trusted, as identified by the caller ID number.

The caller could then decide whether to put her call through. Alternatively, when the receiver was currently not present, she had the option to be directed to the mobile phone. Another feature allowed the caller to speak directly to voice mail without the phone ringing (and potentially disturbing a meeting). The virtual secretary could also assist with re-scheduling the call at a more convenient time. Here is an example of such a mediated phone call:

> Virtual secretary: *"Hello Jane. Thank you for calling "Bob Smith* [recorded]*". This is his digital secretary. Bob is currently in a meeting with Eric. Your call is important. Please hang on, or press 1, to leave a message. Instead, you may now press 2 to be connected to the office phone. To schedule a call at the next available time, please press 4."*

Pilot tests had shown that some callers hung up as soon as they heard a synthesized voice. That's why we had the person record his own name ("Bob Smith" in the example above).

Direct voice mail In case the caller decided to leave a direct voice mail (by pressing 1), the receiver's phone did not ring. Instead, he immediately received an email with a .wav attachment including the message.

> Email from the virtual secretary: *Dear Bob: just wanted to let you know you were just left a 0:15 long message (number 12) in your mailbox from <07216085929>, so you might want to check it when you get a chance. Thanks!*
> *–Your Digital Secretary.*
> [Attachment: msg_2007-11-05_07216085929.wav]

Proactive scheduling of phone calls In case the caller opted to schedule a call, the virtual assistant stored the pending call. As soon as the receiver became available, the system automatically initiated a call back, and said:

> Virtual secretary: *"Hello Jane, this is an automatic call back from Bob's virtual assistant. Bob is now available. Please hold while your call is being connected, or you may hang up the phone."*

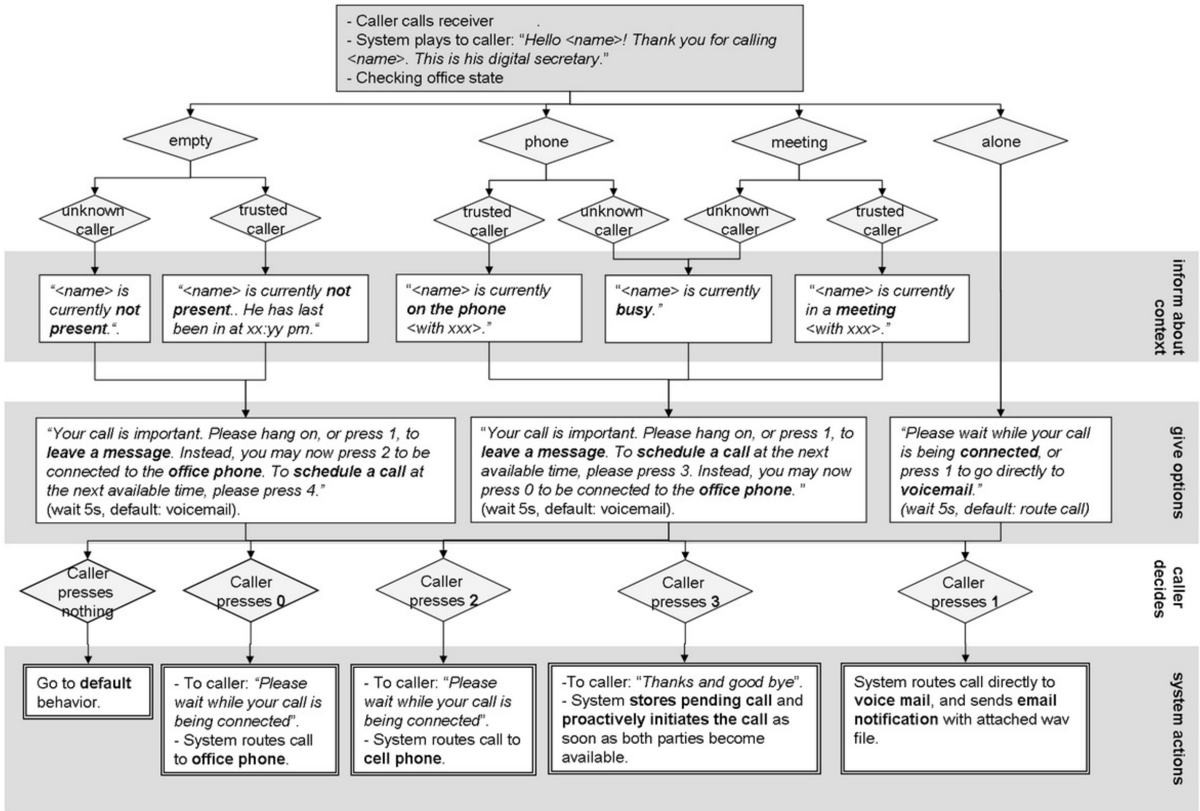


Figure 7.7: Flow diagram showing how the Connector Service mediates calls to the office phone.

In contrast to traditional call back features from telephone providers that initiate the call back as soon as the person ends the current phone conversation, the context-aware virtual assistant initiated the call back only when the receiver actually became available again. This could be after the end of a meeting, or when the person returned to the office. Moreover, if the caller was also located in one of the smart offices and using her own virtual assistant, the scheduled call was initiated only when *both parties* were simultaneously available.

Details of possible interaction flows can be found in Figure 7.7.

7.2.2 Mediating face-to-face interactions

People in front of closed doors face the same problems as callers: they can only guess whether the other party's context is convenient for communication. For an in-person interaction, one might assume that a closed door suggests that the person does not want to be interrupted, just as an open door suggests availability and less concern about privacy. Unfortunately, things are rather more complicated. In many locations, office doors are closed because of outside noise, e.g. after lectures when many students linger to chat in the hallway. More generally, it has been shown that simple physical artifacts such as open office doors are indeed used to regulate interaction, but that there are no fixed social norms [43]. Quite a few modern office designs employ glass windows on doors, so that visitors can easily check the situation inside.

Our virtual secretary took a more nuanced and privacy-sensitive approach. Visitors were automatically recognized from a small webcam when approaching the office door (the face detection system was described in Section 7.1.2). Figure 7.1.2 shows the screen and speakers which were placed in front of each office. The webcam was placed on top of the screen, so that people automatically looked into the camera while reading the welcome message. A few seconds were enough for the face identification system to detect and identify the person.

Once a visitor was detected in front of the office, the virtual secretary provided her with important contextual information about the situation inside the office (Figure 7.9). The goal was to prevent untimely disruptions from taking place, such as during meetings or phone conversations. Again, the level of detail of disclosed information depended on whether the visitor was known and trusted, as identified by the face recognition system. When the office occupant returned to his office, he saw a list of persons who had missed him (Figure 7.10).

7.3 Web-based Office Activity Diary

The virtual secretary provided a diary as web application. Users of the system could securely log into the web site where they could browse office activities: checking missed calls, finding out who came to the office while they were away, and generally seeing how their day or week or month looked like.

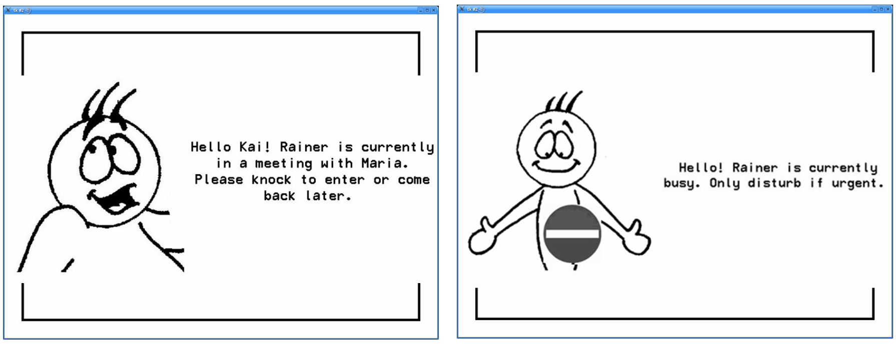
This diary was presented as interactive time line widget[1], which is shown is Figure 7.11. The time line consisted of two bands, which were synchronized and could be scrolled indefinitely by dragging with the mouse pointer. This allowed browsing through hours (larger upper band) or days (smaller lower band) etc. The bands showed visual elements for interruptions and office activities, such as when a meeting had taken place. These visual elements were automatically updated throughout the day.

Visual elements displayed in the time line could be clicked in order to display more detailed information, as shown in Figure 7.12. E.g., clicking on one of the orange meeting elements opened an info bubble with details about the meeting, including a snap shot of the situation. This snap shot helped to remind people what the meeting was like, or who it was with, and could be enlarged to full resolution by clicking on it. Similarly, clicking on a call element displayed more information on the context of the phone call, including the caller ID, the situation at the time of the call, what the caller had decided to do, and success of the phone call. The info bubble of a visit displayed an image of the visitor's face, which was especially useful for unknown persons that could not be identified by the face recognition technology.

To insure privacy, snap shots taken from office situations were only accessible via the password protected web application. The snap shots were securely stored in the back-end database, instead of on the web server machine.

[1] http://simile.mit.edu/timeline/

Figure 7.8: A visitor consults the virtual secretary at the office door. A small webcam was placed on top of the screen from which the virtual secretary greeted visitors.



(a) message for known visitor (b) message for unknown visitor

Figure 7.9: The virtual assistant in front of the office door informed visitors about the situation in the office. Known visitors got more details and personal context information, such as why the person is busy or who is in a meeting.

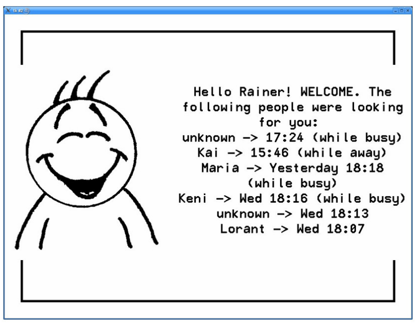


Figure 7.10: When the office occupant entered the office, the virtual assistant displayed missed connection requests.

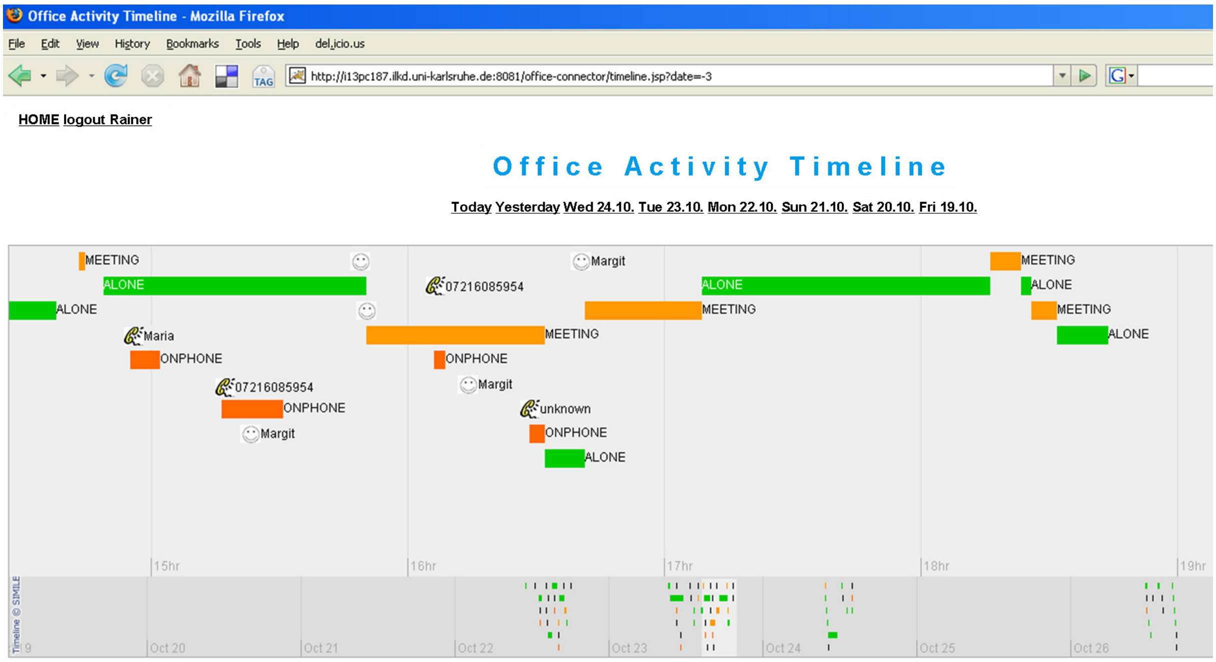


Figure 7.11: The Office Activity Diary was shown as interactive time line widget. The two bands are synchronized and can be scrolled indefinitely by dragging with the mouse pointer.

7.4 System Architecture

In our smart offices, multiple perception technologies, actuators, filters and reasoning components ran over a distributed network and had to interact with each other in an online fashion. Thus, a modular and dynamically reconfigurable architecture was of great importance. In our implementation, the problem was addressed at two levels: low-level high bandwidth data streaming and high-level message passing and control.

For managing low-level video streams, the NIST Smart Flow System[2] middleware was employed. Smart Flow allows fast transfer of high quantities of sensory data over a network of components or clients running on several real machines. Each client produces a uniquely named output flow, which can be accessed by one or more receivers over the network or on the same machine. This makes the system extensible, quickly reconfigurable and easy to use via a graphical interface.

At the higher level, a socket-based message passing scheme developed in our lab is used for sending messages and events between perception components and service modules. A central message server is responsible for registering all interacting modules and redirecting messages at runtime to registered recipients. This allows for complete separation between application semantics and message-passing logic. Besides flexibility and modularity, another main advantage is fault-tolerance, as attempts to pass messages to nonexistent or nonfunctional components cause no further problems for the remaining parts of the system.

[2]http://www.nist.gov/smartspace/nsfs.html

Figure 7.12: When clicking on any element in the activity time line, an info bubble displayed further details, such as snap shots of the situation, or image of the visitor.

The telephone system was implemented entirely on the network, and worked similar to the system described in Section 5.2. The virtual secretary had a specially assigned direct inward dialing number (DID), so that no software needed to be installed on-board caller's or receiver's phones. This aspect was particularly important since we planned to deploy the virtual secretary in a field experiment where it was supposed to mediate actual phone calls from a large number of different callers, and we could not even expect to know who these callers would be. Thus, we could not ask people to call a new number. Instead, we integrated the virtual secretary into the actual phone number by simply forwarded all calls from the normal office phone to the virtual secretary's direct inward dialing number, and placed an additional phone in each office which would ring when calls were ultimately put through.

The overall system architecture is illustrated in Figure 7.13. The system could be easily extended. e.g. by other technologies on the sensor layer.

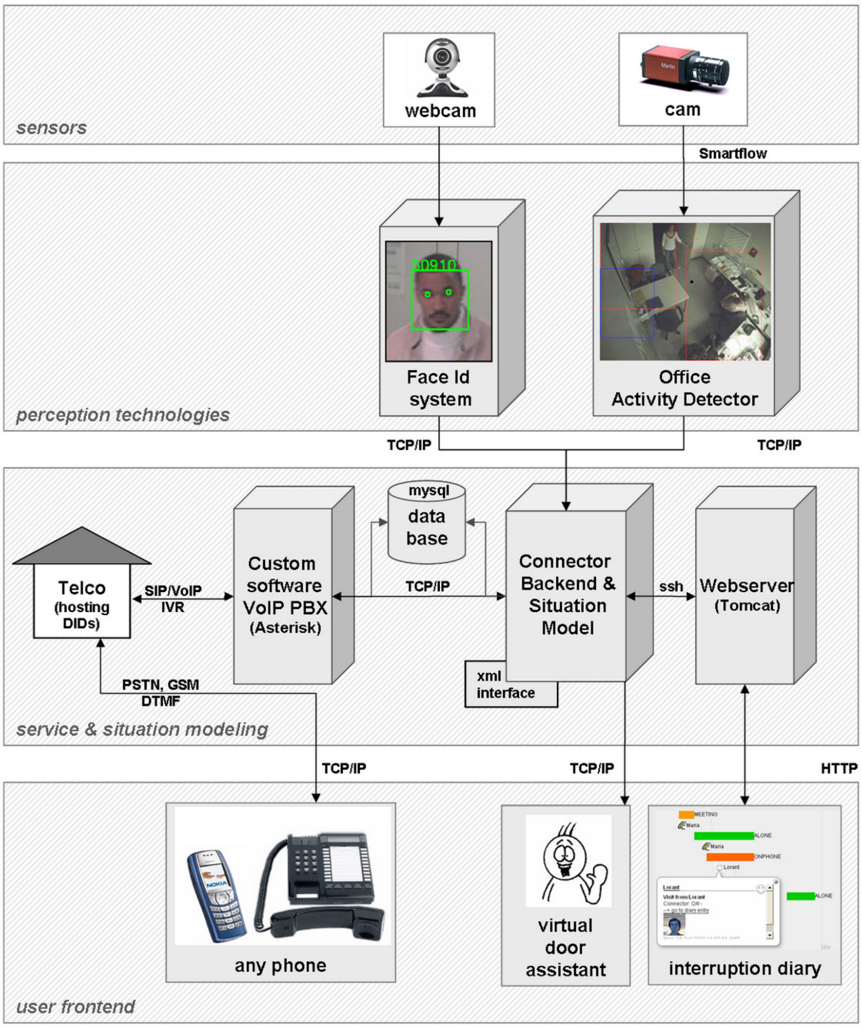


Figure 7.13: Overall architecture of the Connector Service in smart offices.

7.5 Evaluation in a Real-Life Field Experiment

The virtual office secretary was installed in the office of a senior researcher. This person was coordinating a large international research project, teaching a class, and supervising students, and was thus very frequently occupied by meetings and phone calls. We hypothesized that deploying the system in his office would decrease the number of untimely and inappropriate interruptions. But we were as well interested in general office activities, e.g. ratio of interactions via the phone compared to in-person, how many calls failed to reach him, etc.

Our participant was using the virtual secretary during his normal daily activities, mediating his actual phone calls and in-person meetings. Most of the people contacting him did not know about the system. Figure 7.8 showed how a visitor consults the virtual secretray at the office door, and is informed that the person she was trying to see was currently not present.

7.5.1 Study design and data collection

In order to explore whether the virtual secretary was able to make a measurable difference we decided to run an experiment with two conditions: a *control condition* without the digital secretary, and a *study condition* in which the digital secretary mediated phone calls and in-person interactions. Apart from that we tried to make the two conditions as similar, and thus comparable, as possible.

Each condition lasted seven work days. All phone calls to the office phone were mediated by the virtual secretary. In the control condition, the virtual secretary immediately routed all calls to the office phone, just as if it were normal phone calls. In the study condition, it gave information about the current office situation. The caller could then decide whether to put the call through, or instead leave a voice mail or call the mobile phone if the person was not currently present (as was described in Section 7.2.1. The proactive call back feature was disabled for the duration of the study.Only the proactive call back feature was disabled for this study[3].

In both conditions visitors were detected in front of the office door. In the control condition, the virtual secretary greeted people, and always asked to knock and enter just as they normally would. In the study condition it again gave information about the current situation.

All office activity was tracked by the system. Our telephone server automatically logged who was calling when, as well as length and outcome of each call. Similarly, the system logged information about visitors and times and of visits. A final in-person interview was held with the participant, and with a few callers.

Additionally, at the end of each day, the participant was asked to browse the web-based diary and annotate each interruption, according to how available he had been, how appropriate the interruption was, and some other details about the conversation (Figure 7.14). Diary entries could be opened directly from the time line widget. Three snap shots of the office helped remember the situation, one from the time when the interruption occurred, and two others that were taken ten and twenty seconds later. These snapshots were useful to remember the situation, such as whether a visitor had actually entered the office.

7.5.2 Experiencing life with a digital secretary

In the post-study interview, our participant reported that he did not at all feel disturbed by the system. The virtual secretary had operated reliably and without technical difficulties. Even though this person was a researcher in the computer vision community, it was important for him to know that the camera video stream was only used to detect the current activity, and not stored on hard disk. Snap shots were only accessible via the password protected web interface. He reported that he would have had much greater concerns if audio data had been collected.

He found the office activity diary to be very professional and useful. The web-based interface allowed easy accessibility from work as well as from home. He would have liked to keep it after the duration of the study. Our participant as well liked the fact that he could be informed about missed calls and visits (which usually occur unchecked) via the same interface.

A major drawback of the system was that caller ID information got lost for most phone calls. Unfortunately this was based on the fact that most local telephone providers disable caller ID transportation

[3]This was because caller ID information was not available, and it seemed too bothersome to ask callers to punch in their call back numbers.

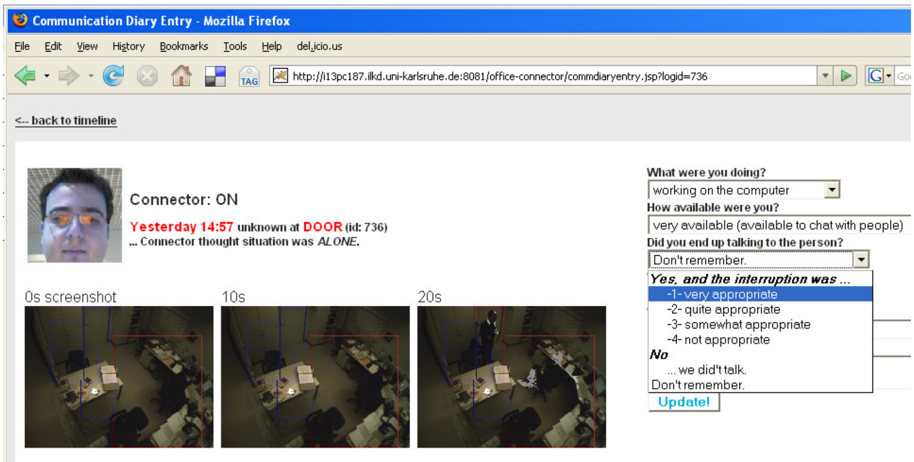


Figure 7.14: At the end of each day, the participant annotated all interruptions according to how available he was, and how timely and appropriate the interruption had been, etc.

across networks on call forwarding. Thus, this problem would not have occurred if callers would directly dial the virtual secretary's DID number, which was impractical for a real-life field study. However, since the University's telephone provider enabled such caller ID transportation all internal calls could be identified correctly.

While many visitors were colleagues who knew about the digital secretary in front of the office door, most callers were completely unaware of the fact that they could reach a digital secretary. However, we found that people had no problem to interact with the system. Many callers left voice messages when informed that the person was in a meeting or on the phone. Some callers decided to let the system route their call to the person's cell phone when informed that he was not currently present. Only very few callers were confused and hung up as soon as the virtual secretary answered. One confused caller called a colleague because she did not understand this new "answering machine".

It was difficult to get explicit feed-back from the many people trying to reach the participant during the duration of the study, especially from callers. The little feed-back we received was very positive. One of the persons who had decided to call the cell phone mentioned at the end of the phone conversation that he found it great to get this kind of assistance when trying to reach someone.

All in all, the interface could be used intuitively by the vast majority of people. Our participant was very enthusiastic about his experience with the virtual secretary, and mentioned that he would have liked to keep it permanently after the duration of the study, if only the issue with caller IDs could be resolved.

7.5.3 Performance of the virtual secretary

Throughout the duration of the study, a total number of 150 contact attempts were protocolled by the system. Out of these, 76 were phone calls to the office phone, and 74 were in-person visits.

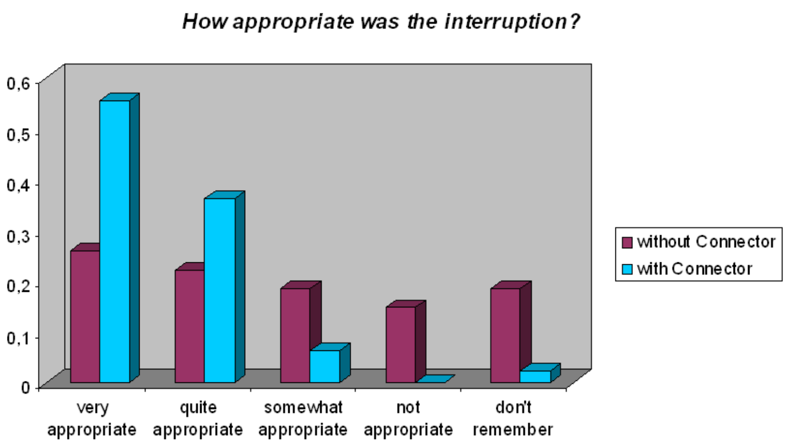


Figure 7.15: Frequency distributions of rating how appropriate interruptions were with and without the Connector.

Comparing appropriateness of interruptions

We hypothesized that the virtual secretary would free our user from inappropriate interruptions, such as incoming calls during meetings, or people knocking at the door while the person was in the middle of a phone conversation.

After the end of the study, our participant reported that he could not tell for sure whether he was less often inappropriately interrupted when he had the virtual secretary. However, throughout the study our participant had updated his activity diary and assessed each interruption according to how appropriate it was in the current context. In the post interview, he reported that the snap shots provided by the system (as can be seen in Figure 7.14) were extremely useful, and helped him remember the exact situation when the interruption ocurred.

Figure 7.15 shows distributions of how appropriate interruptions were. In the control condition, about half of all interruptions were either "very appropriate" (25%) or "quite appropriate" (22%). One third (34%) of all interruptions were perceived as "somewhat appropriate" (19%) or even "not appropriate" (15%). When the virtual secretary mediated connections, the number of "not appropriate" interruptions drooped to zero, and only 6% of interruptions were assessed as "somewhat appropriate". The vast majority of interruptions (92%) were perceived as either "quite appropriate" (36%) or "very appropriate" (56%).

We used a t-test to find out whether the different ratings of appropriateness with and without the digital secretary could as well have occurred by chance. Results showed that the observed ratings of appropriateness of interruptions were significantly different from each other ($t = -3.93$, $df = 65$, $p < .001$). Thus, using the virtual secretary in smart offices did in fact help to decrease the amount of inappropriate interruptions in our study.

Success of contact attempts

Figure 7.16 shows success of phone calls. Overall, only half (49.4%) of all contact attempts led to a conversation, the other attempts were not successful: calls did not get picked up, or ended up in the

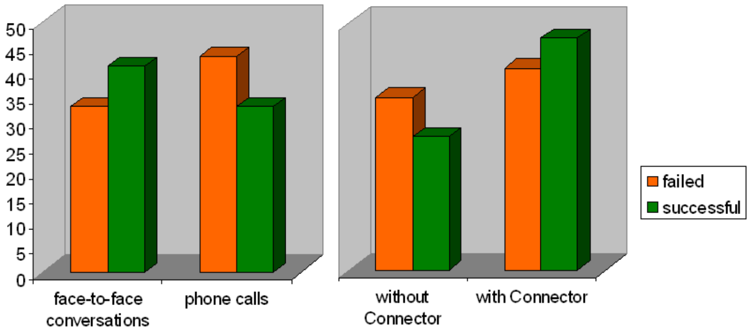


Figure 7.16: Success of contact attempts across modalities and study conditions.

voicemail box, and visitors found the door locked or encountered the person busy, with no time to talk. In-person visits were a little bit more likely to be successful (55%) then phone calls (43%). Moreover, there were more successful connections with the virtual secretary (53%) than in the control condition (43%). However, neither of these differences concerning the success of connection attempts were significant.

Moreover, there was no significant difference concerning the success of contact attempts between the conditions in our study. Actually, even more contact attempts were successful in the study condition, which means that potentially more interruptions could have occurred when the virtual secretary mediated interruptions. This supports the conclusion that the virtual secretary could actually help people to place more timely calls, instead of merely decreasing call success.

Comparing general availability?

Next, we looked at how available the person had reported to be at times when interruptions occurred. Our participant had rated his general availability on a scale of one to four: "not available", "somewhat available", "quite available" and "very available". Since interruptions that took place when interactions were mediated by the virtual secretary were rated as significantly more appropriate, we assumed that our participant would as well report that he was more available for communications. However, there were no significant differences in how available our participant was when interruptions occurred between the different conditions. This means that interruptions were more appropriate, even though the person did not report to be more available.

We looked at ratings of availability in more detail. Distributions of ratings of availablity can be seen in Figure 7.17. The vast majority of answers were either "somewhat available" (64%) or "quite available" (25%). The participant revealed in the post-study interview that it had been difficult for him to provide a general assessment of availability, which explains as well the large number of "don't remember" answers. Especially in the office, one is always to some extend busy, and never "very available" for interruptions. Also, one is never categorically un-available, not even during meetings or phone conversations, as one expects interruptions when at the workplace.

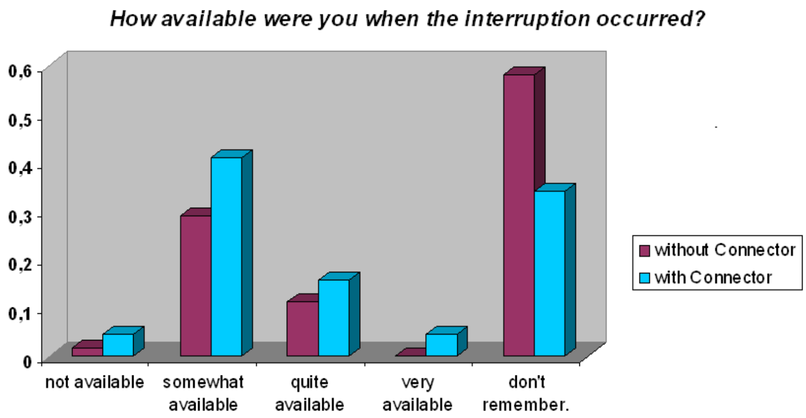


Figure 7.17: Frequency distribution how available the participant was with and without the Connector.

7.6 Discussion: From Office to Mobile Environments

Deploying the Connector as virtual secretary in the office of a senior researcher showed that the system was running robustly 24/7 over an extended period of time, and was easy to use, even without prior training. Moreover, we could empirically proof that there were significantly fewer inappropriate workplace interruptions when the virtual secretary mediated phone calls as well as in-person interactions in our smart offices.

Over the course of the study, a total of 150 real-life interactions were mediated by the virtual secretary. Analysis of collected data revealed that only 49.4% of all contact attempts led to a successful conversation, with no significant differences between connection attempts via the phone and face-to-face.

After learning that the virtual secretary had significantly reduced the number of untimely or inappropriate workplace interruptions, we had expected our participant to also report that he had been more available at times when these interruptions occurred. However, our participant found it difficult to report general availability, and mostly reported to be "somewhat available". There were no significant differences in how available our participant felt when he was in a meeting, or working by himself.

This suggests that availability for communication, at least in the office, may depend a great deal on the particular instance of interruption (who is calling, how urgent or important is the call, etc.), and is thus extremely complex and hard to predict. When designing our virtual secretary, we had assumed that humans would be much better at gauging the urgency and importance of a particular instance of communication, and had thus decided to follow a user-driven approach. Our system added context to communication, but let the final decision of when and how to initiate contact to the callers. After deploying this system in a field study we believe that this user-driven approach was the correct decision.

The Connector as virtual assistant helped to increase common ground [10] for office communications, and decreased the number of untimely interruptions, via the phone as well as in-person. However, this system was placed in a smart office environment, and mediated mostly business conversations. Office phones have the great advantage of being associated to a location - someone's office - which provides at least some place-based guidance for both caller and receiver. E.g. the participant in our study did never feel extremely available or unavailable when in his office, since he expected interruptions.

As our previous research suggests (Chapter 6), this may be very different in *mobile contexts*, where we can be anywhere from "waiting for the bus" to "presenting to the CEO". As such, in mobile contexts where traditionally no common ground exists between caller and receiver, activity plays a much more important role and the problem of inappropriate interruptions is much more apparent.

Ideally we would like to build intelligent systems mediating interruptions in mobile contexts, where communication is most critical, and potentially most disruptive. The following Chapter will describe our efforts towards deploying the Connector in mobile environments, by placing the Connector agent on-board smart phones.

Chapter 8

Connector On Smart Phones

Context-aware systems are limited to instrumented environments where perceptual technologies can keep track of situations and availability. The problem of inappropriate interruptions is much more apparent in mobile contexts, where traditionally no common ground exists between caller and receiver. To capture the context of mobile users outside of fixed locations, such as our smart offices, a mobile sensing platform was needed.

This Chapter will describe explorations of deploying the Connector as intelligent agent, on-board smart phones (Figure 8.1). This work was partly published in [14, 15], and demonstrated at the International Conference on Multimodal Interfaces in 2006.

8.1 The Phone as Sensing Platform

Researchers have used different sensors to detect physical activity of mobile users (e.g. accelerometers, light and pressure sensors) [63, 34]. Our investigations of availability for communication have confirmed that physical activity measures can provide valuable context cues to predict whether someone will pick up

Figure 8.1: Connector was implemented to run as intelligent agent on-board smart phones.

Figure 8.2: Initially, acoustic signals were captured by a wearable microphone.

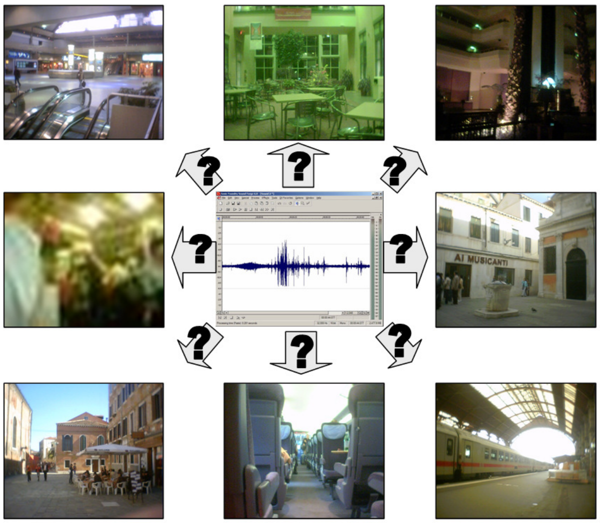


Figure 8.3: Different environments were classified from acoustic signals from the environment.

the phone. However, social context, such as interacting with others, was the most consistently significant predictor of availability.

Acoustic environment detection

In collaboration with Robert Malkin at Carnegie Mellon University we explored systems that combined wearable audio sensors with smart phones. To identify the user's current environment, acoustic signals captured by a wearable microphone were analyzed (8.2). This system could distinguish nine different environments, such as airport, bus, gallery (loosely any non-travel-purposed large indoor space), park (any outdoor, non-urban setting), plaza (any open-air, urban, non-street setting), restaurant, street, train, and train platform (Figure 8.3). Using five-second audio segments to detect environment, the system error rate was 15% [46].

We attempted to port this system to the smart phone, so that all recording and processing could be done on board the phone. But we found that the audio quality and background noise cancellation significantly limited our ability to make fine distinctions. Intensive efforts to use built-in microphones of various phones were mostly unsuccessful, as cellphone microphones are optimized for speech and often cancel out background noise, the very data we need to analyze user context. Therefore, we had to limit our environment set to three classes: inside, outside, and vehicle. Using five-second audio segments to detect environment, the system error rate was 16%.

Acoustic availability assessment

We also developed a system which did not explicitly attempt to model the environment but instead used acoustic evidence and a learned activity model in order to estimate whether or not the user was interruptible at a given time. We evaluated this system using both high-quality audio[1] [46] and cellphone audio[2] [47]. Neither of these interruptibility systems were ported to on board processing, as they were still relatively computationally expensive.

8.2 Smart Connector Phones

Smart phones allow other possibilities than simply serving as a sensing platform. They can display custom graphical interfaces, which are much easier to use (and especially browse) than voice interfaces. Thanks to Gopy Flaherty, who was responsible for programming the phones, Connector Smart Phones ran on three different platforms (Figure 8.4).

Connector Phones communicated with the back-end system via an XML-protocol over TCP/IP, GSM or GPRS, and were able to seamlessly automatically switch between networks while communicating with the back-end system. This provided greater robustness, and saved costs when a free wireless network was available.

The smart phones, running a custom Connector GUI, let the user browse contacts, set preferences (e.g. in what situation switch to vibrate), and informed callers of presence information and availability of potential contact persons before placing a phone call. In the following, we will describe a few of these features in more detail.

Controlling phone alerts Connector Phones could control the type of phone alerts, such as automatically switching to silent mode in meetings, or extra loud in noisy environments. The user could set her own rules regarding how the phone should behave in different situations, depending on the social relationship to the caller. In early versions, the phone enabled a smart notification service: if the person was busy, the phone blocked the call and automatically sent an SMS message back to the caller, and let the call go through on the second attempt.

> SMS: *The person you just called is currently busy. If this call is urgent, please call again within the next 2 minutes, and your call will be connected.*

The phone displayed blocked calls, as shown in Figure 8.5.

[1] Single-user study, total misclassification rate of 6.5%, an improvement over the 11% error rate resulting in using only the prior interruptibility distribution.

[2] Two-user study, total misclassification rate of 16% (prior 31%) and 27% (prior 44%) respectively.

Calendar integration Connector Phones integrated personal calendars. Calendars were again hosted by an Exchange Server, and could be accessed and updated directly on the phone. People could opt to share their calendar information with others via the Connector GUI.

Smart calls In order to prevent calls in inappropriate situations, the system facilitated ”smart calls”. When calling another person, the system suggested channels on how to contact the other person. Contextual information was either implicitly sensed on Connector Phones, as described in Section 8.1, or extracted from person's calendars. E.g. if the other person was in a meeting, the caller did not get the option to call, but could instead leave a voice message, write an email, or schedule a call for the next available time (Figure 8.6).

Proactive scheduling of phone calls Connector Phones provided a feature that let people schedule a phone call with another person. This was because very often people want to talk to another person, but it is not important that the conversation takes place immediately. The system allowed scheduling of phone calls, and took care of initiating a call at a convenient time. Via the GUI, people could either schedule a call before a specific day (e.g. ”I need to talk to Jeff sometime today or tomorrow”), or before an event in their actual calendar (e.g. ”I need to talk to Jeff before this meeting in Prague”). The phone displayed in the center of Figure 8.4 shows a screenshot of such an interaction. The system proactively initiated the call (both phones rang) as soon as both parties were available. Scheduled calls were as well added to personal calendars, and could thus be re-scheduled by both parties.

Connector Speech Interface In order to demonstrate a more human-like user interface to Connector Phones, Thilo Köhler helped to implement speaker independent automatic speech recognition (ASR) on our mobile devices. Speech recognition could be used to issue simple commands to the device, including dialing and sending short predefined messages. The ASR system is a PDA-optimized port of the Janus Recognition Toolkit (JRTk) [64]. This feature worked on an HP iPAQ with maximal processing power.

While our smart Connector phones worked well in the laboratory, the difficulty to sense context on-board the phones made it difficult to use the phones in a user study, or in real-world situations. This and other design trade-offs are discussed in the subsequent Chapter.

Figure 8.4: Three generations of smart Connector Phones: Sony Ericsson P900, i-mate PDA2K GSM/GPRS Pocket PC, and HP iPAQ h6315 .

Figure 8.5: A smart phone running the Connector GUI displays a blocked call.

Figure 8.6: Connector Phones provided different channels on how to best contact the other person.

Chapter 9

Discussing Design Trade-offs

The previous Chapters have introduced various prototype implementations of context-aware Connector Systems, as listed in Table 9.1). The first design (Chapter 5) extracted pre-planned contextual information as they are widely available in personal calendars. Subsequently, we have extended the system's knowledge to situated information, as detected by perception technologies which were installed in an instrumented environment, the office (Chapter 7). Finally, we have discussed our efforts towards capturing the context of mobile users (8).

This Chapter discusses most critical design trade-offs that have emerged during our efforts towards designing and deploying context-aware Connector agents facilitating more appropriate and better timed human communication in mobile as well as office environments.

9.1 Dealing with Privacy

Privacy is a major concern with system that exposes personal information, and our phone and office mediators are no exception. Such systems must balance the need to protect user privacy with providing sufficiently meaningful social cues.

Versions of the Connector Service implemented privacy rules according to the contextual knowledge they had, ideally based on who should be able to see what information when. Each rule specified when it will fire depending on the time of day (free time or work time) and the location of the user. Such privacy rules could be created for single contacts or entire groups in the contact list; default settings were used for unknown persons. E.g.

> *"I want my colleagues to see the room I am in, but only during working hours. My family should always see where I am. The default should be specifying what country I am in, etc."*

Figure 9.1 shows a screen shot of a Connector web interface that was described in more detail in [16]. The detailed location and picture of a person (highlighted red) was not visible to everyone. Maintaining and generating appropriate settings for all callers in all situations was still difficult, and an additional overhead for users of the system.

Table 9.1: Overview of context-aware Connector prototypes.

	system	context info
	*Calendar-based Connector*	scheduled / planned context
	*Connector in Smart Offices*	detected but static context
	*Connector on Smart Phones*	mobile context

Privacy is a critical issue inherent to context-aware systems. Designers must balance the trade-offs between making information public (e.g. social translucence approach [26]) and protecting the user's senses of privacy and information ownership [29]. A mediator such as the Connector can to some extend enable privacy decision making to be abstracted away, by transmitting various details of personal information, or simply collapsing complex ensemble of personal information points to a single indicator: available or not?

However, we have seen that algorithmic prediction of availablity is extremely complex. For example, when deploying our virtual secretary in a smart office, our participant found it difficult to report general availability, and mostly reported to be "somewhat available". Availability for communication depended a great deal on the particular instance of interruption (who is calling, how urgent or important is the call, etc.).

When designing context-aware systems there will always be a trade-off between relying on the intelligence of a system or taking advantage of human decision-making skills by broadcasting private context information to others.

9.2 Designing for Mobility

Another major trade-off when designing context-aware systems concerns mobility. Ideally we would like to build intelligent systems mediating interruptions in mobile contexts, to establish some common ground wherever we are. This is because the problem of inappropriate interruptions is most apparent in mobile contexts, where traditionally no common ground exists between caller and receiver.

Figure 9.1: Screen shot of a Connector web interface, where the detailed location of a person is displayed in a map. Information with red bounding boxes was not made public to everyone.

However, we experienced that it is expensive and technically very difficult to develop and deploy sensing technologies in everyday mobile environments. Connector agents running on smartphone platforms tried to sense contextual information on-board smart phones, instead of installing sensors in an instrumented environment, such as our smart offices.

While Connector smart phones worked well in laboratory conditions, the limitations of phone audio input systems made them difficult to use in real-world situations. Increased battery consumption is another limiting factor. Aside from these problems, not many people other than corporate users have expensive smart phones capable of supporting expensive computations.

And maybe even more significantly, privacy becomes an even more critical issue if sensing technologies are no longer limited to dedicated locations such as offices, but are everywhere you go. Unfortunately this is where we would need them most.

9.3 Empowering the Receiver

The locus of control is another critical factor to account for when designing a context-aware communication service such as the Connector.

While the caller can decide how she wants to communicate, and initiates the connection when her context is convenient for communication, the receiver is the one who faces most of the problems that came with mobile communication technologies. That is the reason why many context-aware agents try to relief receivers from unwanted or inappropriate interruptions by placing all the decision making either with the system or the caller, or shared between the two.

However, receivers may not *want to* give up control over their interruptions. Even though people often wish they had more uninterrupted time to finish tasks, many kinds of interruption may be welcome, and even seem to be an important part of how people manage their lives. E.g., corporate managers in a previous study have claimed that they desire to remain control over their interruptions, since filtering them is an important part of their job [38].

These considerations lead us to a fundamentally different approach to the problem, which goes away from context-aware technologies and puts the receiver back into the loop of negotiating with the sender the most appropriate medium and style of communication.

Part III

Augmenting Modality: Facilitating Communication Through More Flexible Interfaces

A fundamentally different approach to the problem of appropriate and timely communication is to focus on facilitating new ways of communication rather than managing or mediating interruption. Thus, rather than prevent the interruption and try to deflect it to a more appropriate time or place, this approach relies on building smarter, more flexible interfaces that afford greater choice of interaction style or modality.

The emphasis here, in contrast to previously outlined context-aware approaches, is on empowering the user with better tools (interfaces) to deal with untimely communication, and make the interruption non-interrupting. A direct consequence of this is that the locus of control is placed on the user, rather than on a mediating technology, with the advantage of having humans make all critical decisions. Given sensitive and careful design the system plays the role of a passive mediator that facilitates communications to the best modality or medium for each recipient.

Two systems are described, One-way Phone (**Chapter 10**) and Touch-Talk (**Chapter 11**), that explore this corner of the design space. With deploying these experimental systems in a business context, we will ask questions such as: what if people could deal with previously unanswerable phone calls in a socially appropriate manner? E.g., what if people could answer a call without talking aloud?

Chapter 10

The One-Way Phone

Our first design explored the concept of a phone that enabled one-way phone calls. This system was developed as part of Prof. Nass's experimental research practicum course for Interface Theory and Design at Stanford University, and in collaboration with SAP Labs in Palo Alto. The project team was led by the author and Qianying Wang. The involvement of an industry company allowed us to test the system with industry employees. The One-way phone was published in [20, 21].

10.1 Who Gets To Hit The Receivers Mute Button?

Telephones are traditionally viewed as a two-way medium. But communication can be either a one-way notification or a two-way conversation. When getting a phone call from a sender, the receiver has no knowledge of whether to expect a long conversation and time commitment. This could be one of the reasons why so many phone calls fail to reach intended recipients.

What if the sender could indicate to the receiver whether she plans to engage in a conversation or only wishes to send out a notification? What if the receiver could indicate her availability for one-way or two-way communication, and let the sender structure his message accordingly? What if the sender could hit the receiver's phone's mute button?

To empirically examine these questions, we implemented a one-way telephone system that let users communicate either in traditional two-way or a new one-way-only manner.

The one-way telephone system was designed entirely on the network. Negotiation of the type of call was realized using interactive voice responses. People using the system were assigned both a one-way and a two-way toll-free phone number, which were ideally both stored in callers address books, e.g. as "John 1-way", and "John 2-way". By pressing the digit '1' or '2' the caller could indicate whether the call is a one-way notification or two-way conversation, respectively.

Our system routed the call to the receiver and provided the appropriate caller ID so that the receiver could identify the caller and type of call *before* answering the phone (see Figure 10.1). Having the caller indicate the type of call instead of inferring it from the dialed (one-way or two-way) number is necessary, because returning a missed call to "John-one-way" happens in a very different situation and should not automatically be routed as a one-way call.

Receivers could join the negotiation process by choosing to receive a two-way phone call, or if their situation did not allow a conversation, they could opt to switch to a one-way call. Under this scenario, the caller was prompted for the change so that she could then scope her message accordingly or call back later. We have implemented this system. An interaction flow diagram is shown in Figure 10.2.

One-way calls ensured that only the caller's communication could be heard. The receiver's line was automatically muted in order to both preserve privacy for the receiver and to remove any social obligation for conversation.

10.2 Implementation as Network-based Calling Feature

The one-way phone system was designed on the network, so that one-way calls could be immediately sent and received from any phone capable of sending DTMF signals (i.e., touch-tones), similar to the previous Connector systems. Figure 10.3 shows an overview of the main components. No software needed to be installed on board of the phones. This allowed us to deploy and evaluate our system in an everyday business context rather than remaining confined to a controlled laboratory environment.

We used Asterisk^{TM} as PBX, and the Asterisk Gateway Interface (AGI) to add Java programming functionality, such as database access. The Cepstral[1] text-to-speech engine was used to dynamically generate voice responses on the server.

Configuring the telephone system to facilitate one-way calls was much more challenging then setting up the previous systems. This was because interactive voice response (IVR) is a phone technology that allows a computer to detect voice and touch tones in a normal phone call. An automated voice system allows callers to navigate through phone menus and be directed to the correct extension by pressing a series of numbers on a touch-tone phone. (i.e., "Push 1 for sales, push 2 for support, etc."). However, once the call is connected to a phone extension or third party phone number, in our case the actual mobile phone of the receiver, no further IVR is possible since the computer is out of the loop.

To facilitate one-way calls, we needed to configure the telephone system to go beyond such an interaction, and

1. negotiate the type of call with caller *and* receiver *before* finally connecting the call.
2. mute the receiver's line in a one-way communication.

Both problems were solved by taking advantage of Asterisk^{TM} built-in conferencing feature. The underlying implementation was transparent to the participants, for whom it feels just as a normal phone call. After the caller pressed '1' or '2' to indicate her preference for a one-way or two-way call, she was entered into a new conference, with music-on-hold playing a repeated ring tone as long as she was the only member in the conference.

[1] http://cepstral.com/

Figure 10.1: An incoming one-way phone call can be recognized from the caller ID, on any mobile phone.

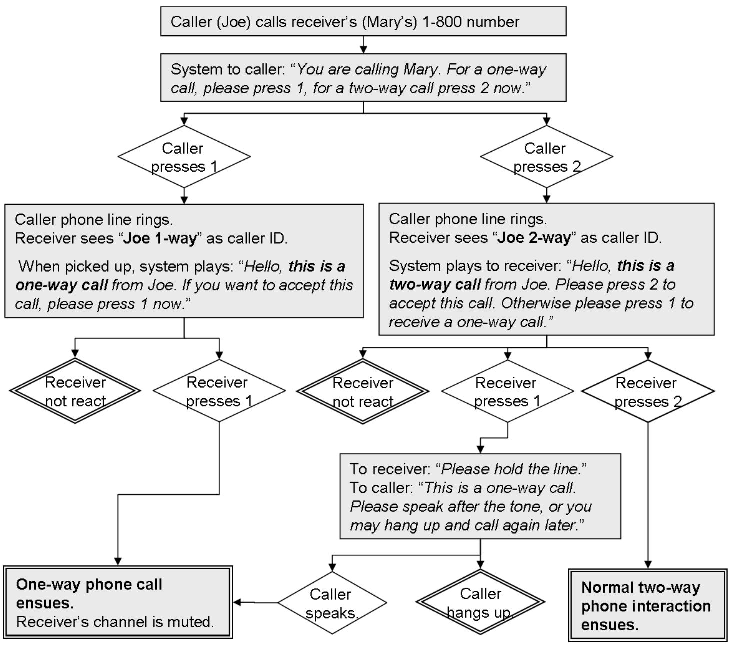


Figure 10.2: Interaction flow diagram of the one-way telephone system, where caller and receiver negotiate the type of call.

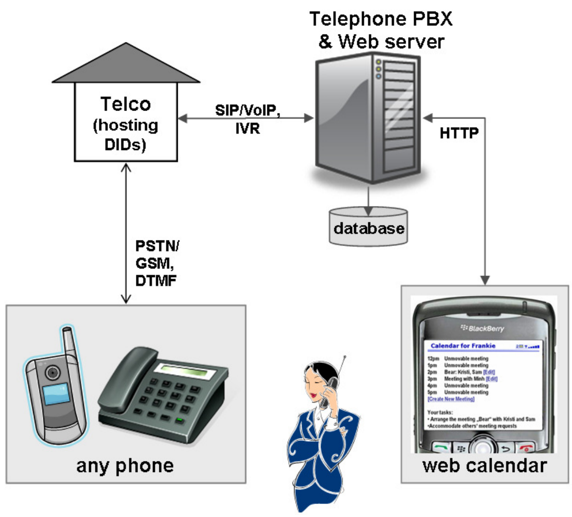


Figure 10.3: One-way calls could be immediately deployed from any phone.

Meanwhile, the telephone system dialed a new outgoing call to the receiver, setting the appropriate, one-way or two-way, caller ID. When the receiver answered the phone, the system asked her to either accept the call, or switch it to one-way (for interaction details see Figure 10.2). Only then was the receiver added to the same conference, the ring tone stopped and made the caller believe that the call just got connected.

Moreover, it is possible in conferencing to set a 'monitor-only mode', in which a user can listen only, and not be heard, which solved the second problem with muting the receiver's end of the conversation in a one-way call.

10.3 Controlled Field Experiment in a Real-Life Business Context

To investigate the usefulness and acceptance of one-way calls, we deployed the one-way telephony system in a field study with SAP employees.

10.3.1 Ten busy industry employees solved group scheduling tasks

Ten SAP employees volunteered to participate in a two week long study during their normal work days. These users are frequently occupied by meetings, conferences and discussions, and have less free time for full-blown conversations. Participants were employees from different departments within the company (finance, sales, human resources, research, and product strategy) and had different job responsibilities, ranging from assistants to managers. Each participant used his or her own mobile phone.

Using a within-participants experiment design, we examined three randomly assigned experimental conditions. Each condition lasted for two days for each participant.

Figure 10.4: Each day, participants were asked to solve a group scheduling task via a web-based calendar, which was easily accessible via PCs or Blackberry^{TM}.

- *Condition 1*: The caller could control the type of phone call and indicate whether she desired a two-way conversation or one-way notification.
- *Condition 2*: This control condition involved traditional two-way calls only.
- *Condition 3*: Receivers could join the negotiation process, accepting a two-way call or opting to switch it to a one-way call.

Each day, participants were asked to solve a group scheduling task, as this activity occurs very often in real life, especially in business contexts. Participants were given artificial web-based calendars and a phone directory with their team members' toll-free one-way numbers. Participants needed to find out the availability of other team members for a meeting by phone calls. No voice mails, emails or IM could be used. The artificial calendars were accessible via PCs or Blackberry^{TM} (Figure 10.4). Many of the managers used Blackberry^{TM} as main communication device.

We hypothesized that one-way calls would be useful to confirm a meeting, to reach someone in a meeting, or to provide requested information.

Participants' attitudes towards one-way calls were assessed via online questionnaires. Our telephony server automatically logged behavioral information about caller, receiver, call mode (one-way or two-way), time of call, length of each call, audio recordings of all calls, and all updates on participants' calendars towards task completion. Personal interviews were held with each participant after each condition, face-to-face or by phone if the participant was not a local employee.

10.3.2 Behaviors and attitudes towards one-way calling

There were significant differences in frequency of successful calls between the three conditions ($\chi^2(2) = 7.25$, $p < .05$). There was a higher frequency of task completion in the first condition (57%) than in the third condition (33%), ($\chi^2(1) = 7.08$, $p < .01$). However, in the third condition, when receivers could turn calls into one-way, participants' motivation dropped towards the end of the study, and significantly fewer phone calls were dialed overall (55 calls in the last condition, compared to 61 and 86 respectively in the first two conditions). There were no other task completion differences.

Eight out of ten participants found one-way phone calls very natural and easy to send and receive on a ten point scale ($M = 9.3$, $SD = 1.2$). The other two participants reported in interviews that they found it awkward listening to someone speak without being able to respond.

Although people found one-way easy to use, not many one-way calls were placed overall: only 12% of possible one-way calls were ultimately connected as one-way. Self-reported attitudinal data showed that participants claimed that they were more likely to pick up a one-way call than a two-way call when they were busy, ($t(9) = 5.48$, $p < .01$). Behavioral results showed that there was no significant difference in the percentage of calls that were answered. The average success rate of calls in our experiment was 52%.

Interviews with our participants left a general impression that job responsibility and corresponding communication style seemed to matter a great deal. Managers generally disliked one-way calls; some of them even found one-way calls to be arrogant and rude. Participants mentioned that multi-tasking, such as answering emails via Blackberry™ or via laptop is becoming increasingly socially appropriate during meetings, at least in the technology-savvy IT industry. During our interviews, participants mentioned that it would be extremely useful if someone was calling and one could just switch it to an IM conversation.

Our participants reported that they perceived normal two-way calls to be more polite, and as callers in most cases they wanted to have at least a quick confirmation, and felt awkward not to get any feedback on whether the receiver had understood the message. Also, as callers, they did not feel the urge to send a one-way call since callers can easily pick convenient times to initiate conversations. At the same time, participants reported that they were more likely to receive one-way calls than to send one-way calls.

10.4 Discussion: Feed-Back is Necessary

Observations from the field study indicated that one-way calls were especially useful for receivers. But had the limitation that the receiver's interaction was limited to listening only, with no way to give feedback or confirmations. People desired synchronous confirmations during communication, and found it rude in a one-way call to just give instructions and leave the other party with no opportunity for feedback. This is consistent with existing literature in linguistic pragmatics literature, e.g., adjacency pairs [30, 60] and maintenance of common ground.

Compared with asynchronous media, phone calls have the advantage of immediate response. We have learned that synchronous telephone communication is inherently two-way and must remain this way. These insights informed the subsequent design of Touch-Talk.

Chapter 11

Touch-Talk

Based on experiences with the one-way calling system we developed a novel Connector feature, "Touch-Talk". Touch-Talk enabled people to pick up previously unanswerable calls and deal with simple issues quickly, quietly and privately even in public contexts.

Touch-Talk was developed within our ongoing collaboration with the CHIMe Lab, and SAP's Research and Emerging Solutions departments. The system and results from an exploratory field deployment study were published in [20].

11.1 Quick, Quiet and Private Business Communications

With Touch-Talk, receivers could answer incoming calls without speaking aloud by pressing keypad buttons, which would literally make the phone speak for them. Each pressed number corresponded to a prerecorded voice prompt (Figure 11.1). We assumed that Touch-Talk could be especially useful in meeting situations where one cannot talk aloud, or in public spaces where others could overhear or feel disturbed by the conversation. As such, Touch-Talk allowed receivers to choose and mix text and speech input modalities.

Touch-Talk imposed a substantial burden on the caller, who needed to structure the conversation and ask clear questions. In exchange for caller's efforts to use Touch-Talk, they would ideally be able to communicate with the Touch-Talker at times when they would have otherwise been forced to voicemail. We designed Touch-Talk such that people could balance the cost of their time and efforts against the benefit of getting the most urgent issues resolved in a Touch-Talk interaction.

11.2 Touch-Talk Design

The overall goal for this iteration of our system design was to create a system that people would be willing to use in real situations, for their real interactions at the workplace. Because eliminating existing phone features (e.g., voice mail) in the previous study (Section 10.3 on page 96) caused problems, we decided to preserve existing phone features.

Figure 11.1: With Touch-Talk, receivers could answer incoming calls without speaking aloud by pressing keypad buttons, which would literally make the phone speak for them.

11.2.1 Switching a call to Touch-Talk

We designed the system so that every call started out as a traditional phone call. Each call could be rejected or answered, routed to voice mail, or switched to Touch-Talk mode. It was possible to switch between Touch-Talk and normal phone modes at any time throughout a call. See Figure 11.2 for the entire interaction flow.

At any time in a normal phone communication, the receiver could switch a call into Touch-Talk mode by pressing '0' on the phone's keypad. The system would then play a message announcing to the caller that the receiver had switched the call into Touch-Talk mode, and encouraged the caller to start asking yes/no questions. The receiver's channel was muted in Touch-Talk mode.

In Touch-Talk mode, pressing any of the digits 1 to 9 played the associated voice prompt, to both the caller and the receiver. Not using pound (#) or star (*) was a constraint imposed by our technology setup, which will be described in the next section.

11.2.2 Touch-Talk voice prompts

Selecting voice prompts

We selected a set of default voice prompts inspired by extended discussions and piloting with potential Touch-Talk users (Figure 11.3). *Yes* and *No* were chosen to enable users to accept or reject a question. We assumed that after some practice with Touch-Talk, callers would learn how to ask useful questions to find out what they wanted to know, similar to the "Twenty Questions" game. The receiver and initiator of the Touch-Talk conversation, might either require more detailed information and use *More information, please*, or decide that this issue cannot be dealt with in a Touch-Talk conversation, and use *Let's talk later* instead.

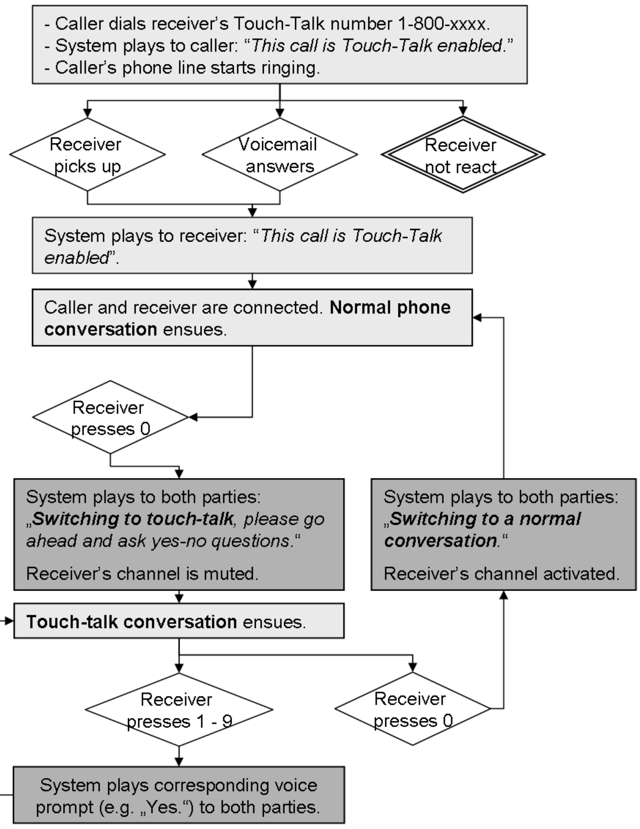


Figure 11.2: Interaction flow diagram of the Touch-Talk system.

There are long-lived expectations derived from social conventions about language use, such as creating and maintaining common ground while conversing [4]. Similarly, one should not simply end conversations; instead, a person is supposed to use an interactive close [60]. We decided to use *Thank you, bye.*

Assigning a separate key for *Okay* was inspired by findings suggesting that people do use linguistic feedback such as Mmhmm or Uh-huh to show agreement or confirmation of understanding [3], often combined with nodding in a face-to-face interaction. We assumed that users would otherwise heavily use the *Yes* for this purpose, suggesting acceptance; acceptance is different from just understanding.

We thought that *Please repeat* would be necessary in noisy environments, or if the Touch-Talker was temporarily distracted. *Remember, Touch-Talk only enables yes-no responses* gives the receiver a way to remind the caller that he cannot answer open-ended questions in Touch-Talk mode. *Sorry, the person had to leave the conversation* and will call back later was designed to be a quick escape-key, e.g., when someone is addressed during a meeting, and needs to immediately drop his quiet Touch-Talk interaction.

Mapping voice prompts

We aimed at mapping voice prompts to keys on the keypad as intuitively and easily memorable as possible. Our initial approach was to map voice prompts to the alphabetical letters assigned to each key, such as

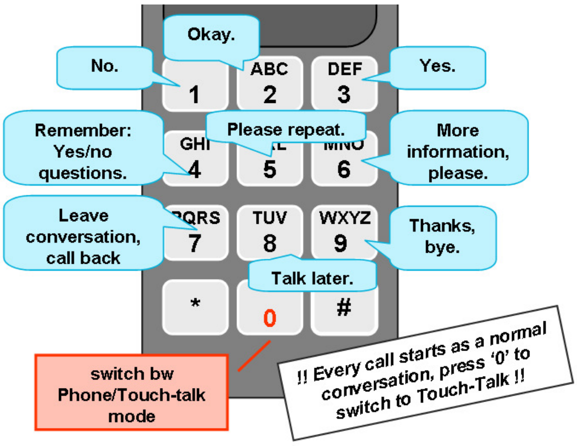


Figure 11.3: Touch-Talk keypad to voice prompt mapping.

Yes on the 9, *No* on the 6. However, this approach was abandoned after realizing that this mapping does not work for smart phones, such as Blackberry™, that have QWERTY keyboards.

We then decided to place positively valenced answers toward the right side of the keypad, and the more negatively valenced answers to the left, as this mapping of negative to positive onto left to right fits this cultural context (Figure 11.3):

- The *first keypad row* contains acceptance, confirmation and rejection.
- The *second keypad row* contains prompts when still undecided and trying to find out more, ranging from least to most polite.
- The *third keypad row* contains different closings, also ranging from least to most polite.

Creating voice prompts

We tried to keep voice responses brief, especially the ones that had potential for very frequent use. We decided to use synthesized voices rather then pre-recorded human voices, even though human voices are often easier to understand. This was because we did not want to confuse the caller who might think that there is the actual person talking, as happens sometimes in the beginning of voicemail introductions. We were also trying to match voice quality with Touch-Talk agent ability as recommended by [31], lowering user's expectations of the Touch-Talk agent by using computer-generated voices because the agent could ultimately only convey nine different pre-scripted voice prompts.

Since humans respond to voice technologies as they respond to actual people and behave as they would in any social situation, Nass suggests in [53] that computer generated voices should not use personal pronouns such as I or my. Therefore, we used *More information, please* instead of *I need more information*, etc. We believe that this avoidance of the first person is appropriate even when a human triggers the voice responses.

Moreover, to match the gender of Touch-Talkers with the gender of voice prompts, female participants were assigned a synthesized female voice; male participants a synthesized male voice.

In order to conform to turn-taking nature of human dialogue, distinct warning sounds were played to announce any Touch-Talk prompt or switching between modes to prevent the caller from feeling inappropriately interrupted while speaking [4]. Error sounds were played to the receiver when a key was pressed that was not assigned to an action.

11.3 Implementation as Network-based Calling Feature

We adapted and extended the original One-way telephone system so that it supported Touch-Talk conversations, considering design goals and requirements mentioned in the previous section. Main challenges were:

1. the interception of DTMF signals from the receiver's channel during the phone conversation (phone conference).
2. playing Touch-Talk voice prompts to both channels.

In order to keep complete control over the course of a phone call, we reused the idea of simulating a phone call through a conference call. However, it was technically not possible to directly intercept DTMF signals in the conference[1].

We bypassed this constraint by using a functionality that allows a party to exit the conference by entering a valid single digit extension. We thus allowed the Touch-Talker to escape the conversation by pressing any key (between '0' and '9'). In this case the call is bounced back to the dial plan[2]. From the dial plan, the Touch-Talker was immediately re-added to the conversation. This transition worked surprisingly fast and was not noticeable to the users.

It was then necessary to play the voice prompt associated to the key press to *both* partners in the conversation. Since caller and receiver are connected to each other in the conference call, both channels are out of reach for interactive voice responses.

The issue was ultimately resolved by adding a "third party" to the conversation. This was a computer program acting very much like a proxy that can speak for the Touch-Talker. This concept is illustrated in Figure 11.4. This proxy was another IVR script, which was created by the server through dialing a direct outbound call to the proxy's DID number, and thus to itself, and adding the proxy to the conference.

Now, whenever the Touch-Talker pressed a key, the telephone system could immediately send a message[3] to his Touc-Talk proxy. The proxy consulted the database to find the corresponding action, created the synthesized voice prompt, and played it to the conference, so that both parties could hear it.

This setup worked surprisingly well. Setting the proxy to "speak only mode" significantly reduced background noise in the conference call, so that it sounded just as a normal phone call. The system could handle pounding in keys in rapid succession, and handle multiple calls at once.

[1]At least not without connecting hardware telephone lines to the server, since this feature only works with Zap channels.
[2]The dial plan is where one can define what actions the telephone system will take when calls are answered
[3]All Java programs called by the telephone system via the AGI interface operate in the same address space, and can communicate via simple method calls.

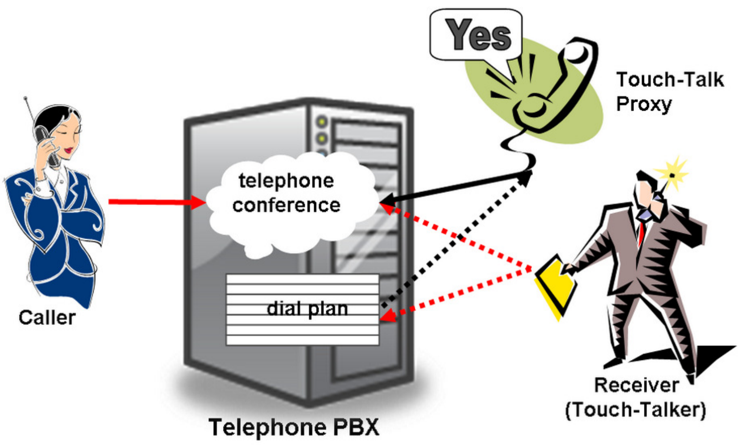


Figure 11.4: A Touch-Talk conversation consisted of a phone conference including caller, receiver and a proxy talking on behalf of the receiver.

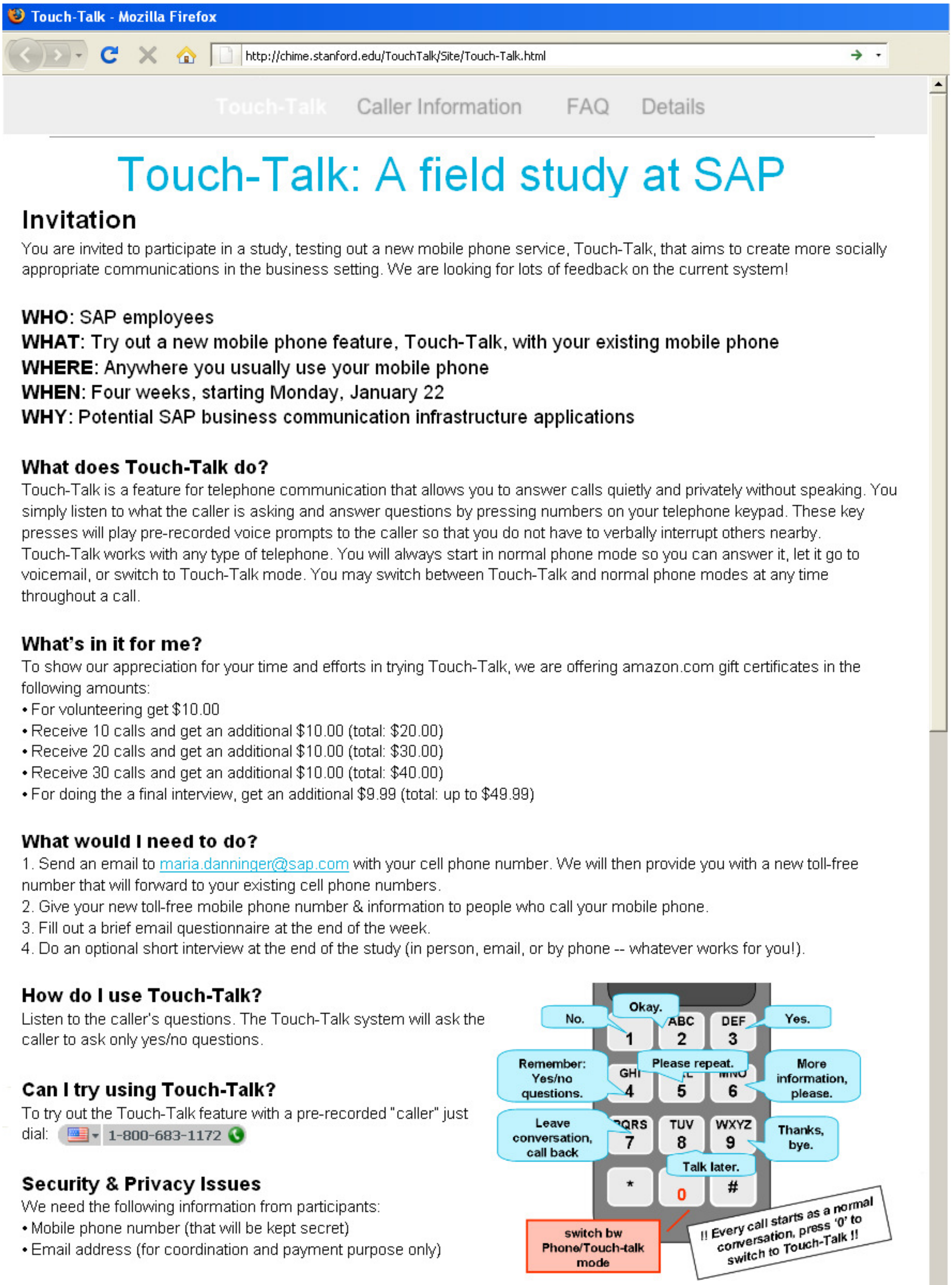
11.4 Exploratory Field Study in a Real-Life Business Context

We chose to do an open-ended exploratory field study by introducing the technology into a real world business setting, maximizing external validity, rather than doing a controlled experiment with random assignment to conditions.

11.4.1 Recruiting busy industry employees

Recruiting busy industry employees to voluntarily use a new system for their real-life business communications over a several week period was a daunting challenge. A web site was set up, including a training number to test the system (Figure 11.5). Emails advertising Touch-Talk and potential business benefits were sent to hundreds of members of SAP Research and Emerging Solutions in the US. However, most participants only signed up after attending in-person presentations, where the system was actually demonstrated. Ultimately, fifteen employees volunteered to use Touch-Talk for their real business conversations. The participants were mostly researchers and managers. Their participation in this project was an encouraging sign for the future of Touch-Talk.

The fifteen volunteers, three female and twelve male, were urged to practice Touch-Talk by calling a special training number so that they could practice with the system before they received any live Touch-Talk calls. After familiarizing themselves with the system, each participant was assigned a toll-free number and distributed it to whomever they chose, including colleagues, friends, and family. They sent out an e-mail to these potential callers which included a link to the Touch-Talk website. This website explained, in detail, what the caller could expect from a Touch-Talk conversation.



Touch-Talk - Mozilla Firefox

http://chime.stanford.edu/TouchTalk/Site/Touch-Talk.html

Touch-Talk Caller Information FAQ Details

Touch-Talk: A field study at SAP

Invitation

You are invited to participate in a study, testing out a new mobile phone service, Touch-Talk, that aims to create more socially appropriate communications in the business setting. We are looking for lots of feedback on the current system!

WHO: SAP employees
WHAT: Try out a new mobile phone feature, Touch-Talk, with your existing mobile phone
WHERE: Anywhere you usually use your mobile phone
WHEN: Four weeks, starting Monday, January 22
WHY: Potential SAP business communication infrastructure applications

What does Touch-Talk do?

Touch-Talk is a feature for telephone communication that allows you to answer calls quietly and privately without speaking. You simply listen to what the caller is asking and answer questions by pressing numbers on your telephone keypad. These key presses will play pre-recorded voice prompts to the caller so that you do not have to verbally interrupt others nearby. Touch-Talk works with any type of telephone. You will always start in normal phone mode so you can answer it, let it go to voicemail, or switch to Touch-Talk mode. You may switch between Touch-Talk and normal phone modes at any time throughout a call.

What's in it for me?

To show our appreciation for your time and efforts in trying Touch-Talk, we are offering amazon.com gift certificates in the following amounts:

- For volunteering get $10.00
- Receive 10 calls and get an additional $10.00 (total: $20.00)
- Receive 20 calls and get an additional $10.00 (total: $30.00)
- Receive 30 calls and get an additional $10.00 (total: $40.00)
- For doing the a final interview, get an additional $9.99 (total: up to $49.99)

What would I need to do?

1. Send an email to maria.danninger@sap.com with your cell phone number. We will then provide you with a new toll-free number that will forward to your existing cell phone numbers.
2. Give your new toll-free mobile phone number & information to people who call your mobile phone.
3. Fill out a brief email questionnaire at the end of the week.
4. Do an optional short interview at the end of the study (in person, email, or by phone -- whatever works for you!).

How do I use Touch-Talk?

Listen to the caller's questions. The Touch-Talk system will ask the caller to ask only yes/no questions.

Can I try using Touch-Talk?

To try out the Touch-Talk feature with a pre-recorded "caller" just dial: 1-800-683-1172

Security & Privacy Issues

We need the following information from participants:

- Mobile phone number (that will be kept secret)
- Email address (for coordination and payment purpose only)

Figure 11.5: A web site was set up, including a training number to test the system without bothering real people.

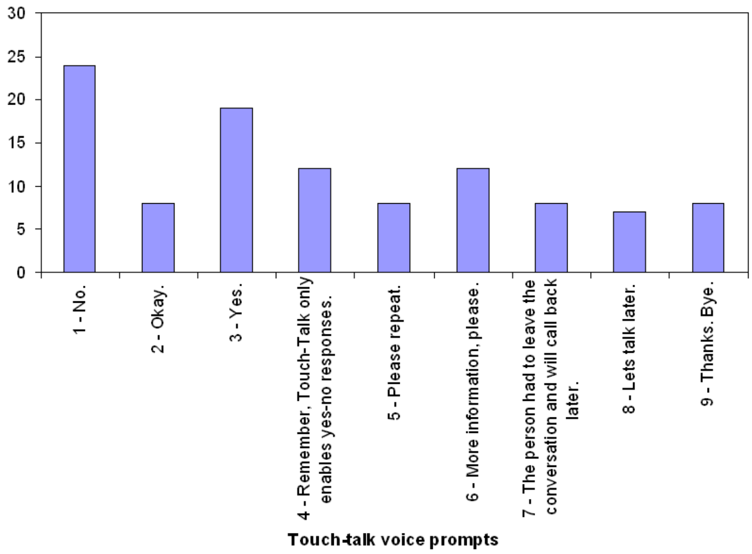


Figure 11.6: Frequency distribution of voice prompt use.

11.4.2 Touch-Talk in actual business conversations

The biggest challenge of the study, and the primary reason why most people did not receive as many Touch-Talk enabled calls as they would have liked, was the lack of potential callers to call the Touch-Talk numbers. A few participants reported that they had sent out the number to their whole group, but received no calls. We hypothesized that callers were reluctant to enter the temporar Touch-Talk numbers into their phone contact lists, this would have made the process simpler and more automatic. Searching for an old email and manually dialing the number was apparently too much of a burden.

Nevertheless, over the four-week duration of the study, 58 calls were made to Touch-Talk phone numbers and 34% of those were switched into Touch-Talk. To respect the privacy of our participants, we did not record any phone conversations, but we did receive consent to log the key presses of all calls. 105 Touch-Talk voice prompts were used in a total of 20 Touch-Talk conversations.

Voice prompt usage

When creating voice prompts, we worried that users would shy away from using *No* because it was somewhat abrupt and could have been perceived as impolite. However, behavioral data showed that *No* and *Yes* were the first and second most popular voice prompts, respectively. There were nearly significantly more *No* and *Yes* responses than all others combined, ($\chi^2(1) = 3.77$, $p = .05$). As can be seen, there was clearly an uneven distribution of key selections across all 9 choices, ($\chi^2(8) = 23.91$, $p < .01$) (see Figure 11.6).

The average Touch-Talk conversation lasted 37 seconds from switching to Touch-Talk mode to the end of the call or switching back to a normal conversation.

Investigating endings of Touch-Talk conversations provided interesting insights about the conversation. *Thanks, Bye* was used at the end of 41% of all Touch-Talk conversations, suggesting that a conclusion could be reached. 24% of calls were closed with *Let's talk later* or *The person had to leave the conversation and will call back later*, suggesting that the issue needed a follow-up conversation. Another 18% of calls were switched back to a normal conversation before the end of the call.

11.4.3 Attitudinal results of Touch-Talk use

Touch-Talk users were able to give feedback about the system in biweekly questionnaires and a final interview. Generally users were very enthusiastic about their Touch-Talk experiences, and reported that they would have liked to use it more, and would like to use it after the duration of the study.

They mentioned using Touch-Talk mainly in meetings and public spaces. Surprisingly, a few participants used Touch-Talk while driving and especially in city traffic. One user explained that he found Touch-Talk very useful commuting to work using public transportation.

Some of our users called their colleagues' Touch-Talk numbers as well. They said that although calling was awkward at first, the more they used the service, the more useful it became. For example, one user reported that she called her colleague who was in a meeting and actually got agreement on several urgent issues.

Participants utilized Touch-Talk for a variety of reasons. Many used Touch-Talk to let the caller know that they were busy, and negotiate an alternative time for a phone conversation. A nonnative speaker explained that he was excited about Touch-Talk because it allowed him to communicate more clearly in the English language.

Users also commented that they would feel comfortable changing all conversations into Touch-Talk mode, though a few mentioned they would not do so with an authority figure such as their manager. One user also added that he would not use Touch-Talk with his spouse because the service might be too impersonal.

While these results were encouraging, it was clear that the system needed to be improved with new features to fully maximize the benefits of Touch-Talk. Users would have liked Touch-Talk to be integrated into their real phone number instead of routed through a different number. Next, even though each Touch-Talker was equipped with a "reminder card" depicting the keypad and corresponding voice prompts, the users still complained about forgetting the mapping of voice prompts. Head-sets were useful, especially in meeting situations. Finally, although all users reported that the default voice prompts were very useful, the Touch-Talkers also wanted an option to customize their own prompts.

11.5 The Next Iteration

Informed by the results and user feedback from our field study, we added two more features to the existing system: personalization and a voice prompt preview feature.

11.5.1 Personalization (Touch-Talk 2.0)

We implemented the first generation of a web-based personalization interface where users could log in and customize their voice prompts from any web browser (Figure 11.7). For improved user experience,

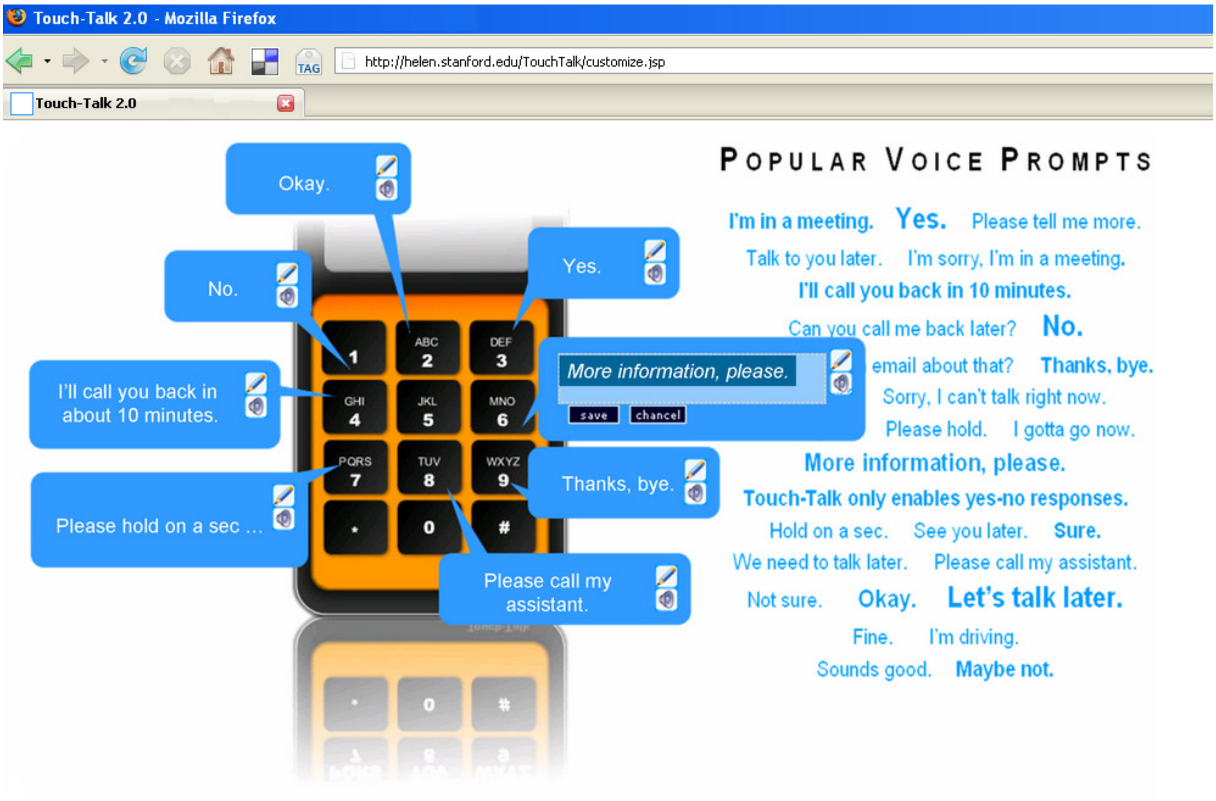


Figure 11.7: Screen shot of the Touch-Talk personalization web interface. On the right are popular voice prompts from other people.

the interactive website was created using Ajax (Asynchronous Javascript and XML). Users could create new prompts and listen to how they sounded as synthesized speech, before assigning them to keys.

Users could even opt to share their voice prompts with others. All shared prompts were displayed in a prompt cloud, with prompt sizes and colors indicating their popularity. This encouraged users to browse through others Touch-Talk prompts so that they might leverage and contribute more ideas to the Touch-Talk community. Changing voice prompts was effective immediately with the next phone call.

11.5.2 Preview Feature

Deploying Touch-Talk in a business context showed that people had trouble remembering the keys for specific voice prompts. This problem was likely to be aggravated when users could additionally customize their Touch-Talk prompts. To address this issue, we implemented a preview feature that Touch-Talkers can use during a phone conversation to hear what voice responses will be played to the caller before actually sending them. Pressing '5' immediately followed by a key played the associated voice prompt to the Touch-Talker only. To save time, the prompt was played at double the normal speed.

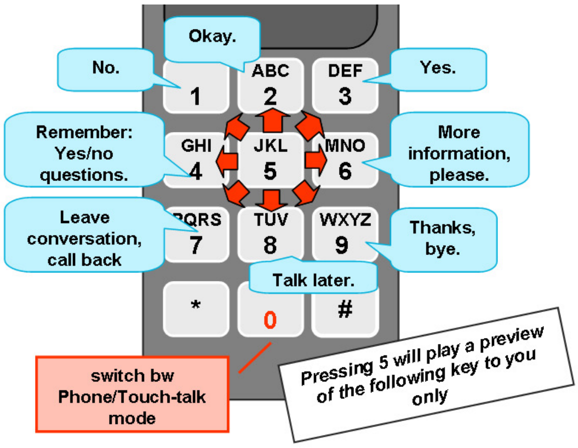


Figure 11.8: The preview feature can be used during a Touch-Talk conversation to hear voice responses before actually sending them.

11.6 Discussion: Possible Future Directions

This Chapter reported on the iterative design and evaluation cycle of Touch-Talk, a mobile phone feature prototype that aims to enable more socially appropriate mobile communication, particularly in public spaces and business contexts. To this end, we have designed Touch-Talk to place negotiation of mobile availability more in the hands of the receiver than previous models of distance communication, which put the burden of guessing the best communication medium on the caller.

Using an open-ended field study where volunteers used Touch-Talk in their actual business conversations, we have learned that short voice prompt responses may be helpful for enabling quite and private mobile communication in meetings and public spaces, and that remembering key-mappings to voice prompts is difficult for users. Directions for future Touch-Talk designs include personalization of Touch-Talk responses and preview options to help remember key-mappings. There are clearly many potential paths for the further design and development of Touch-Talk.

Voice Prompt Themes Integrate different themes of voice prompts could be useful for different callers, such as a boss/manager theme or a family/friends theme, or different contexts, such as a meeting theme or a driving theme.

Business Touch-Talk Using context awareness with smart phones could also add a new dimension to Touch-Talk, particularly in the business context. A Touch-Talk conversation could be enriched with relevant business objects; for example, a caller, who needs confirmations on an order, could make the form appear on the receiver's phone. The receiver could fill out the form, sign it, and return it via Touch-Talk. Assuming that such business objects may be tagged with semantic information, Touch-Talk may automatically supply the relevant business objects (e.g., orders, receipts, meeting agendas, meeting minutes, etc.).

Touch-Talk on Smart Phones Converting Touch-Talk to run on mobile devices (rather than in the network) would enable the use of a graphical Touch-Talk user interface. This will be increasingly important with the inclusion of themes and customized voice prompts. A GUI could also enable displaying and navigating through various Touch-Talk options, customizing them, and perhaps even allow software Touch-Talk buttons. Using integrated phone keypads could as well allow to type in customized answers on the fly.

Touch-Talk and Context-awareness Integrating speech recognition techniques could finally allow caller and receiver to independently decide on input and output modalities. The receiver could decide to communicate via text only, reading what the caller is saying and answering via Touch-Talk. The caller, who decided to call, would speak and hear speech back. The biggest challenge here would be that speech recognition technologies would need to be sophisticated enough to perform reliably on low-quality telephone data.

With the rapid development of new perceptual technologies and context-aware applications in mobile systems, there are new opportunities for Touch-Talk to become more "intelligent" and relieve the user of decision-making burdens. As context-aware technologies improve, they could extend the Touch-Talk system, despite potential privacy concerns. For example, by reading calendars or sensing that the phone is plugged into the car docking station, Touch-Talk could automatically adapt the template to the user's current context.

Chapter 12

Lessons Learned From Real-Life Field Deployments

Conducting user studies in business contexts poses very different challenges compared to studies in university contexts. These busy industry employees do not take part in user studies primarily for monetary incentives, or in order to receive course credits. It is much more important to convince people of the chance to test and influence the development of a new technology that could be useful for them, fun to use, and to the interest of their company.

Some lessons learned from our experiences may be useful to other researchers who are using similar methodologies.

12.1 With regard to system design

- Pilot extensively amongst yourselves before ever releasing a system for a real-life field study (e.g., several months)

- Continue using the system yourselves during the actual study so that you can help troubleshoot problems

- Create an exhaustive list of possible call outcomes and then design around them and/or log them, including special cases (e.g., receivers getting multiple calls at once, callers hanging up before a call is connected, calls going to voicemail, or frustrated users pounding keys in rapid succession)

- Beware of unreliable DIDs; test every number to ensure that users are not receiving calls from inappropriate callers (e.g., people trying to call the previous owner of the number)

- Build a robust prototype that can run 24/7, and set up a back-up system

12.2 With regard to field studies

- Present and demonstrate the system to potential participants
- Build a publicly-available website with FAQs for both callers and callees
- Small compensations, such as Amazon gift certificates, show appreciation for people's time commitment
- Use a per-call payment scheme to encourage use of the system; this provides useful incentives to busy participant volunteers
- Build a training system with computer-generated responder for people to try the system without bothering real people
- Minimize work for volunteers as much as possible (e.g., send email templates to share with potential callers, create Vcards to distribute phone numbers, design questionnaires that can be answered via Blackberry)
- Users will vote with their silence; do not wait for them to report problems
- Personal interviews are critical for gauging feedback with the system, particularly these participants are not with you in the lab at the time they use your system

Part IV

Summary And Conclusions

Chapter 13

Summary And Conclusions

This thesis has discussed the iterative development of the 'Connector Service'. The goal was to build a mobile communication system that could help people better manage the risk of untimely interruptions in an "always-on" world, where we can reach each other anytime, and anyplace.

13.1 Summary of Content

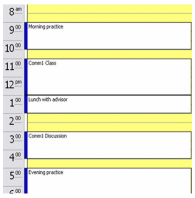
The goal of the Connector Service was to intelligently and non-disruptively integrate mobile technologies into human communication processes. Timelier, more effective and less distracting communication follows from better management of the tension between unwanted interruption and desirable communication.

Designing such a system involves several balancing acts: what sort of mediation is appropriate, who to give control to, and what kind of contextual information is useful. We have explored two fundamentally different strategies when developing and evaluating different version of the Connector Service, one that augments knowledge by using context-awareness, and another that augments modality by using more flexible interfaces. Table 13.1 gives an overview of the different Connector prototype systems.

We assumed that context-sensitive systems, such as the ones described in Part II, would have great potential to reduce call management loads for both sender and receiver, and free the receiver from inappropriate interruptions. Acting similar to a virtual secretary, an intelligent agent should be able to bear much of the burden of negotiating and facilitating calls. We have explored different versions of such context-aware Connector agents, that were based on contextual information from calendars (Chapter 5), smart offices (Chapter 7) and smart phones (Chapter 8).

Systems making use of more flexible or customizable (e.g. multimodal) interfaces, like the ones described in Part III, follow a different approach. Rather then prevent the interruption, they empower the receiver to deal with untimely communication and answer previously unanswerable calls, e.g. by allowing to answer calls without speaking aloud (Chapter 11). Such systems follow a more circumvention-based, rather than mediation-based approach. They provide tools to cleverly deal with interruptions in socially inconvenient situations, and thus put the receiver back into the loop of negotiating with the caller the most appropriate medium and style of communication.

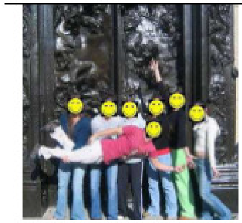
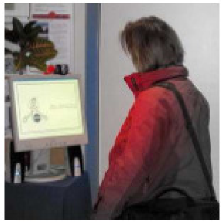
Table 13.1: Overview of Connector prototype systems.

	System	Strategy	Chapter
	*Calendar-based Connector*	augment knowledge	5
	*Connector in Smart Offices*	augment knowledge	7
	*Connector on Smart Phones*	augment knowledge	8
	*One-way Phone*	augment modality	10
	*Touch-Talk*	augment modality	11

Mobile phone usage can be much better observed in a natural setting rather than in a contrived laboratory environment. This is why versions of the Connector Service were studied and evaluated in open-ended as well as experimentally controlled field experiments, with real users engaging in real-life activities. Hundreds of students, as well as dozens of employees from a large IT company used our systems over extended periods of time. Table 13.2 lists all user studies that tested different versions of the Connector Service.

This work generated a number of design insights we believe are broadly applicable within the space of mobile communication, context-awareness, and interruptability. In the following we will review what we believe are most important insights from our investigations.

Table 13.2: Overview of user studies.

	Study name	Participants	Location	Data	Duration	Type	Section
	*Calendar-based Connector*	90 Stanford students	Stanford University	710 calls 15h 38min rec. 131 questionnaires	2 weeks	**field experiment** (2x2 between subjects experimental design, artificial coordination and collaboration tasks)	5.3
	*Availability Study*	92 Stanford students	Stanford University	724 calls 724 diary entries 92 questionnaires	2 weeks	**field study** (exploratory, multiple methods)	6.1
	*Connector in Smart Offices*	an assistant professor + callers/visitors	Universität Karlsruhe (TH)	150 calls & visits 150 questionnaires 1 interview	3 weeks	**field experiment** (study & control cond., real-life interactions)	7.5
	*One-way phone*	10 SAP employees	SAP Palo Alto	203 calls 2h 11min rec. 41 questionnaires 9 interviews	2 weeks	**field experiment** (3 conditions, within-subjects, artificial group scheduling task)	10.3
	*Touch-Talk*	15 SAP employees	SAP Palo Alto	53 calls 13 interviews 35 questionnaires	4 weeks	**field study** (exploratory, real-life interactions)	11.4

13.2 Summary of Contributions

The research presented in this thesis has a significant contribution to make to the discussion of how technology can and should facilitate future mobile phone interactions: both by pushing the state of the art technologically and in socially intelligent design, as well as in terms of collecting principles for what kinds of technological intervention would best balance benefits and burdens of mobile communication in our always-on world.

Towards understanding availability: social factors matter most

In order to design intelligent agents that mediate communications similar to a virtual secretary, we must understand the nature of people's decisions about when to engage in a communication. In a large-scale field study (Section 6.1) we tried to explore what factors might play a role in constructing a person's availability model. 90 college-aged students were randomly called by our server and reported their current availability by pressing 1 to 9 on their mobile phone keypad. Prospective calendars and open-ended post questionnaires were used to assess availability from different perspectives.

Content analysis of post-questionnaires was used to determine different dimensions of mobile context, and subsequent statistical analysis discovered which of the dimensions could predict availability for mobile phone communication. The study revealed that the social context, such as the presence of others, collocated or remotely, or being in a private or public location, was the most powerful and consistently significant predictor of availability. People chose to renounce engaging in mobile phone communication on behalf of those with whom they were already engaged. Moreover, it was the only factor that could as well predict more nuanced measures of general availability, compared to merely predicting whether the person would pick up the phone or not.

The current activity, physical engagement and planned calendar information were also useful predictors of whether people would pick up a phone call.

These findings informed subsequent sensor-based Connector approaches. For example, smart offices were successfully equipped with sensors detecting the in-the-moment social situation inside of the office, such as whether the person was alone, in a face-to-face conversation or on the phone, in order to help people decide whether it was a good time to interrupt (see Section 7.1).

Moreover, these findings and the proposed methodology that combined multiple empirical methods in order to both describe and predict availability may be useful for research methods, theory and design of context-aware systems in general. We will always need to decide which context provides useful information.

Human decision-making is critical: empower the caller with context information

Systems that do actively intervene in order to facilitate human communication must choose to what extend to place decision making responsibilities with the technology, or with the humans. Ideally, an intelligent agent can bear much of the burden of negotiating and facilitating call success. However, accurate algorithmic predictions of availability have shown to be difficult and error-prone (Section 6.3.2).

We assumed that availability is based on a receiver's general situation, as well as on the relationship to the conversational partner, and the nature (importance, urgency, etc.) of a particular instance of communication. And humans are still overwhelmingly better at gauging the urgency and importance of a particular matter. This assumption was confirmed when deploying a virtual secretary in smart offices (see Section 7.5): interruptions were rated as significantly more appropriate, even though the person did not report to generally be more available. Availability depended on the particular instance of interruption.

This should be taken into account when systems try to automatically infer a measure of availability (a sort of "availability-meter"), instead of displaying what the person is actually doing. Thus, our context-aware systems empowered senders to make more informed decisions through the broadcasting of relevant contextual cues. The caller is no longer functioning blind, and thus has less fear of being disruptive, but still retains enough control to force a critical communication through if needed. The receiver also wins because fewer interruptions manage to eventually trickle through the system, as will be discussed next.

Context-awareness can help reduce inappropriate disturbances and improve team performance

We assumed that context-aware active mediators, acting similar to a virtual secretary, would have great potential to free receivers from inappropriate interruptions, and help them work more efficiently. We conducted a series of studies aimed at representing the receiver's state of availability via sensor-based, context-aware systems.

A large-scale field experiment with approximately ninety mobile phone users revealed that providing callers with information from others' calendars at the time of the call had a positive effect on team performance (Section 5.3). In this study eight teams of 11-12 students had access to individual and group Connector features, and solved a collaboration and a coordination task during their normal daily activities. In both cases, Connector-enabled calls were more successful and teams with access to Connector features performed better.

Installing a context-aware virtual secretary in a smart office helped to significantly reduce the number of inappropriate workplace interruptions (Section 7.5). In this study, the virtual secretary mediated all actual phone calls and in-person visits to the office of a senior researcher over a duration of three weeks. Callers and visitors were informed about the current situation inside of the office, before deciding how to contact. Our participant annotated every interruption according to how appropriate it had been. While call success remained the same, there were significantly fewer inappropriate workplace interruptions with the virtual secretary.

Three major limitations of context-aware communication: mobility, privacy and lack of control

However, we found that context-aware agents do have a number of important limitations and design trade-offs. First of all, they are limited to instrumented environments, where perceptual technologies can sense situations and infer availability. It is especially troublesome to carry out this sensing on mobile platforms, given limited processing power, battery life, and sensor obtrusiveness (see Section 8.1). But mobile contexts are where communication is most critical, and potentially most disruptive, since no common ground exists between caller and receiver.

Privacy is also a critical aspect of usability - one that becomes an increasingly broad concern because of ubiquitous sensing and detection. Further, the less reliable sensing becomes in mobile contexts, the more detailed the context information required by the caller to decide whether his call would be appropriate.

In most context-aware systems the receiver lies removed from the locus of control: the decision of the timing and medium of communication lies entirely with the caller or the system. The receiver, though relieved from inappropriate interruptions, does inevitably lose a degree of control over her communications. This can be problematic since retaining control over interruption management may be important to the way many people deal with everyday situations.

This leads us to a more practical consideration. Widespread privacy concerns, as well as the difficulty in developing and deploying robust sensing technologies, makes it difficult to study and evaluate context-sensitive systems with real users in realistic, non-laboratory situations.

Multimodal interfaces are a promising alternative

Systems making use of more flexible or customizable (e.g. multimodal) interfaces, like the ones described in Part III, follow a different approach. They empower the receiver to deal with untimely communication and answer previously unanswerable calls. Such systems follow a more circumvention-based, rather than mediation-based approach. They provide tools to cleverly deal with interruptions in socially inconvenient situations.

Not reliant on contextual information, such passive facilitators do not need a specific monitoring environment, maximizing mobility and minimizing privacy concerns. This opened the way for us to test our systems in field deployments at a large IT company. Busy industry employees used our systems for their real business communications over extended time periods.

Efforts towards introducing one-way voice communication into a business context indicated that synchronous communication is inherently two-way and, if at all possible, should remain this way (see Section 10.3). Touch-Talk communication, on the other hand, was widely accepted and appreciated by busy employees of a large IT company (Section 11.4). Fifteen busy researchers and managers volunteered to use Touch-Talk in their actual business communications, with real colleagues or customers, over a period of four weeks. ‘Yes’ and ‘No’ were the most favorite voice prompts.

Of course, such passive systems have their limitations as well. All the decision-making is placed in the hands of the humans, which could potentially lead to intolerable levels of cognitive interruption on the receiver’s side. Moreover, new interfaces are difficult to learn and get used to. For instance, senders in a Touch-Talk conversation needed to rephrase their whole communication in terms of yes-no questions. Only an iterative development process and evaluations with real users will make such systems easy and natural to use.

Methodologies for large-scale mobile user testing: in the network is better than on-board phone

Mobile use is notoriously difficult to study because of the lack of control inherent when situations vary widely. We have demonstrated a relatively lightweight means for collecting data on mobile use for large sample sizes. Using a customized telephony server connected to DID[1] numbers, we were able to immediately deploy new calling features from any phone capable of sending DTMF digits (see Sections 5.2, 10.2, 11.3). Interactive voice responses (IVR) were used to interact with users. Overall, our setup allowed calls connecting two parties in traditional PSTN or GSM networks to be controlled, monitored, and logged by a server in a VoIP network.

The scalability of this light-weight network-based approach made it possible to test and evaluate prototype systems in large-scale field deployments, and guaranteed the ability to gather sufficiently large data sets for confident generalization. For example, study participants could use their own mobile phones, without needing to share actual phone numbers, which could have raised privacy concerns. All hardware involved was a standard PC running a PBX telephone system. Problems could be solved remotely on the server, rather then on the mobile devices. Call outcomes (e.g. study conditions) could be manipulated on a per-call basis, and were active with the next incoming call. Moreover, IVR scales well to handle large call volumes, and is a familiar concept to use. E.g., callers interacting with the virtual office secretary were completely un-informed (Section 7.5.2), and users found one-way calling easy and natural to send and receive (see Section 10.3.2). Moreover, we were able to "experience sample" ninety users, probing participants for information via their familiar mobile device on a daily basis (see Section 6.1.1).

A major limitation of this network-based approach was the absence of visual browsing via the phones GUI. Visual browsing would for example have helped to remember and increase the number of Touch-Talk voice prompts. Additionally, the network-based approach required usage of third party numbers instead of integrating features with actual phone numbers. This can be prevented when calls are automatically forwarded to the service's dial in number, as we did with calls to office phones in our smart offices (Section 7.2.1 on page 65), but can result in loss of caller ID information.

While locating the agent on-board the phone would certainly be the preferable solution for a future product, the network-based implementation was much better suited for prototyping and large-scale field testing, as it is particularly necessary in early stages of system design.

13.3 Conclusions

This thesis explored ways of better integrating the affordances of mobile technologies into human-human communication, trying to mitigate tensions between desirable communication and inappropriate disturbances. So what can one say about the problem, after our empirical real-world evaluation of different versions of the Connector Service?

To revisit the trade-offs, all systems are inherently limited by how intelligently their rules manage interruption, even if we could in theory design perfect sensors. Thus, the most reliable systems will necessarily

[1]Direct Inward Dialing

involve human decision-making. Empowering the sender requires an intelligent mediation system, but inevitably introduces privacy concerns. Empowering the receiver on the other hand requires circumvention, which can be achieved with flexible interfaces, but involves cognitive effort that can lead to interruption.

If the goal is to minimize interruption and decision-making for the receiver, empower the sender with context information. To optimize for maximum receiver control, empower the receiver with better interfaces. Active mediation systems, that can take decision-making burdens of humans, work best in predictable contexts where monitoring actually provides useful clues. Circumvention, on the other hand, might be the best option in unpredictable, mobile contexts where users are willing to regulate and reshape their conversation behavior and strategies to deal with the situation.

The future ultimate Connector would be an intelligent, context-aware service on a single device. Ideally this device would construct a representation of availability, and then gracefully communicate this information to human conversational partners. The system would be capable of supporting and suggesting the best modality for communication, creating options in real time for caller and receiver for all mobile conversations. Moreover, it might appropriately foreground or background its agency by either actively mediating to lessen the burden of the user, or fading into the background and offering up flexible tools when context did not determine availability but constrained it. Until such a system exists, having a set of proven strategies to build solutions that make use of different strengths of communication technologies seems like a step in the right direction.

If we are to continue living in an always-on world, we might as well equip ourselves with technologies that give us greater say in, and a better basis for connecting and communicating with our fellow citizens.

Chapter 14

Appendix

We would like to introduce in more detail empirical methods of user study design and related statistical methods, as they are typically used in the social sciences. More information on designing evaluations of interactive systems in general, and experimental evaluation in particular, can be found in [22].

Experimental Study Design

A controlled experiment is one of the most powerful methods when evaluating an interactive system, as it provides empirical evidence to support a particular claim. All experiments have the same basic form. The experimenter chooses a hypothesis to test, and manipulates and measures variables in order to produce and compare different experimental conditions.

There are a number of factors which are important to the overall reliability of an experiment: the participants chosen, the variables tested and manipulated, and the hypothesis itself.

Participants Participants should represent the user population as closely as possible. The number of participants (sample size) must be large enough for the experiment design and the statistical method chosen. As a rule of thumb, the minimal number of users for a controlled experiment is 5 to 6, while 10 is commonly recommended.

Variables and experimental design Experiments manipulate and measure variables under controlled conditions, in order to test the hypothesis. Variables identify what you are going to manipulate and what change you expect to see. There are as well the two different types of variables: those that are manipulated in order to obtain the different conditions for comparison (*independent variables*), and those that are measured (*dependent variables*), and depend on changes made to the independent variables.

The dependent variables must be measurable, dependent on the manipulation (i.e., independent variables) and unaffected by other factors. That is why experiments in the social sciences often compare a study condition against a control condition, which is identical to the study condition except for the aspect that is being tested (e.g. access to a new feature, system). Common dependent variables are time to complete a task or quality of performance.

In a *between-subjects design*, a participant is assigned to only one condition. If we have 20 participants, 10 could be randomly assigned to the study condition, and the other 10 to the control condition. Since the user performs only under one of the conditions, there are no learning effects. A disadvantage is that differences between users can affect the results. In a *within-subjects design*, every user performs under each of the conditions. In the previous example, we would only need 10 participants, and each would perform under both conditions. In order to avoid learning effects, the order of conditions can be varied. A popular compromise between the two methods is a mix of both regarding different variables.

Testing hypothesis A hypothesis is the prediction of the outcome of an experiment, stating that there is a difference between the experimental conditions (e.g. that the task completion time is different with and without access to a new system). The aim of the experiment is to show that the prediction is correct. Statistical tests are used to rejecting the null hypothesis, which states that the differences could as well have occurred by chance. We are searching only for differences that make a difference[1].

We have seen that the goal of an experiments is to relate cause and effect, by ruling out as many alternative causes as possible: choosing participants that match the user population, random assignment to conditions, study and control conditions that differ only in the aspect that is being tested, etc..

Statistical Analysis

The choice of the statistical analysis depends on the type of data and the question we want to answer. The data comes from our dependent variables (what we measured), and the questions compare differences between the independent variables which defined the experimental conditions (what we manipulated).

Types of variables

Variables can be either *discrete* (e.g. gender) or *continuous* (e.g. age). The collected data can be measurement or categorical data. *Measurement data* (or quantitative data) is the result of some sort of measurement (e.g. time till task completion, scores on a scale of likeability). *Categorical data* (or qualitative data or frequency data) on the other hand is a categorization, and the data consists of frequencies for each category (e.g. ”22 calls were answered, while 10 were not picked up, and 5 went to voicemail.”).

Choosing a statistical test

We cannot describe all the different statistical techniques here, but will concentrate on a few guidelines and standard tests that were used in this work. Figure 14.1 provides an organizational scheme, which we will discuss next. A more detailed description can be found in statistics books, such as [37].

[1]As I have heard Cliff say so many times :-)

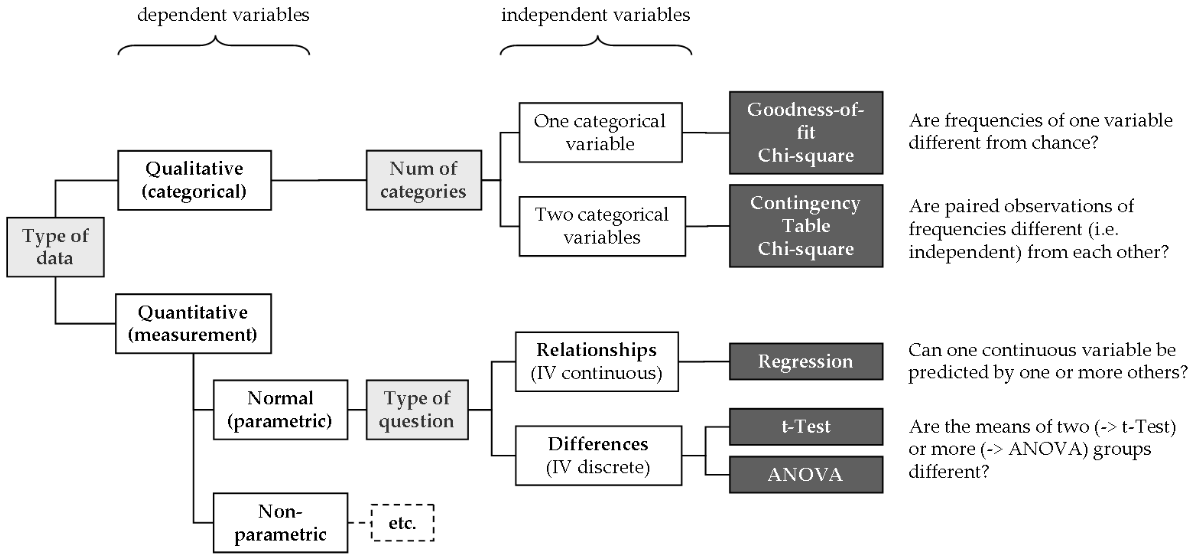


Figure 14.1: Simplified decision tree for chosing a statistical test.

When our **data** (the dependent variables) **is qualitative**, we know frequencies for each category, and are asking the question of whether these frequencies could as well have occurred by pure chance. Such data can be plotted in a contingency table. Chi-square tests can be used to decide whether observed frequencies (one categorical variable) are different from random (-> *Chi-square Goodness-of-fit test*), or whether the frequencies of two categorical variables are different from each other (-> *Chi-square contingency table test*). In experimental analysis the Chi-square contingency table test is often used to determine whether differences in observed frequencies in control and study conditions are statistically significant, and thus influenced by our manipulation.

When our **data is quantitative**, we have to work with means and deviations instead of frequencies. If the random variation of the data takes a special form, in most cases the normal distribution[2], more powerful statistics can be used. Such tests are called *parametric* tests. It is not always easy to decide whether data is *normal*. As a simple and general rule, if data can be seen as the "sum or average of many small *independent* effects they are very likely to be normal" [22]. When data is not normal, we have to work with *non-parametric* tests, which are based purely on the ranking of the data (e.g. Wilcoxon rank-sum test, rank-sum versions of ANOVA, Spearman's rank correlation, etc.), and are thus less likely to detect a difference.

Different statistical methods working with quantitative data depend on the question or hypothesis we want to answer. Quantitative data is commonly used to either compare differences among groups (e.g. we may want to compare whether people in the study and control condition liked a system more, or performed a task faster), or test relationships between variables (e.g. we may want to know whether people's mental engagement was related to their availability).

[2]This means that if we plot the histogram of the random errors, they will follow the bell-shaped form of the normal distribution.

Comparing differences requires a discrete independent variable (e.g. the groups to compare), and is as such the question we ask very often in controlled experiments. We can then use a *t-Test* to compare whether the means of two groups are statistically different from each other. *ANOVA (ANalysis Of VAriance)* tests are used for more complex analysis, e.g. to compare three or more groups.

Testing relationships on the other hand requires comparing a continuous dependent variable (e.g. measure of availability) to one or more continuous independent variables (e.g. measure of mental engagement, physical engagement, etc.), or in other words whether we can predict availability from mental engagement and/or physical engagement. *Regression analysis* is commonly used to compare the relationship between one or more (-> Multiple regression) continuous variables with each other.

The following will review in some more detail statistical tests that were referred to throughout this thesis.

Categorical Data and the Chi-Square Test

Chi-square tests are based on the Chi-square distribution. Chi-square tests are frequently used in the experimental sciences to compare whether paired observations on categorical variables (very often the control and the study condition) are independent from each other.

Chi-square is calculated by finding the difference between each observed and expected frequency (asserting independence of the two variables) for each possible outcome, squaring them, dividing each by the theoretical frequency, and taking the sum of the results:

$$\chi^2_{\text{n-1}} = \sum_{\text{i}=1}^{\text{n}} \frac{(O_i - E_i)^2}{E_i} \text{ , where}$$

O_i = an observed frequency,
E_i = an expected (theoretical) frequency (asserting independence of the two variables)
n = the number of possible outcomes of each event.

Given the degrees of freedom (df) and a given level of confidence (α), one can look up the minimal Chi-square value (χ^2) in a standard table of significance (available as an appendix in the back of most statistics books) and determine whether the calculated χ^2 is large enough to be significant. A probability of 0.05 or less is commonly interpreted as justification that the observed frequencies did not occur by chance, but were influenced by the independent variables (e.g. the different conditions in the study).

Fortunately, statistical computer programs (e.g. SPSS, Minitab, etc.) routinely print the significance test results (χ^2, df, p) and save the trouble of looking them up in a table. Chi-square analysis can as well be done using contingency tables in Excel.

T-Tests

The t-test assesses whether the means of two groups are statistically different from each other, and is often used in controlled experiments to test whether a measurement variable was influenced by the manipulation (e.g. control and study group), or could as well have occurred by chance.

The t-test judges the difference between the group means relative to the variability of their scores, and is calculated as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma^2_{\bar{X}_1}}{n_1} + \frac{\sigma^2_{\bar{X}_2}}{n_2}}}, \text{ where}$$

$\bar{X}_{\mathrm{i}}$ = sample mean of group i,
$\sigma^2_{\bar{X}_{\mathrm{i}}}$ = standard error found in group i,
$\mathrm{n_i}$ = number of scores in group i (e.g. people in group).

To test the significance, one needs to set a risk level or confidence level (α). In most social research, the "rule of thumb" is to set the confidence level α at 0.05.

In the t-test, the degrees of freedom (df) is the number of persons in both groups minus two. Given the confidence level, the df, and the t-value, one can look up the t-value in a standard table of significance (available as an appendix in the back of most statistics books, or of course in the internet) and determine whether the calculated t-value is large enough to be significant. If it is, one can conclude that the means of the two groups are different (even given the variability).

Statistical computer programs routinely print the significance test results, which are the t-value, the degrees of freedom (df), and the significance level (p), and save the trouble of looking them up in a table.

Analysis of Variance (ANOVA)

When we want to compare the means of more then two groups, we could theoretically do pairwise t-tests. However this is not a good idea for a number of reasons. First of all, the number of tests we need to run increased rapidely with the nmber of groups in the study. We would like to do one test to answer our question. As well, the type I error increases with the number of tests we perform.

The **An**alysis **of** **Va**riance (ANOVA) analysis compares differences between two or more sample means. Note that ANOVA for two groups gives the same results then a t-test. Similar to t-tests, ANOVA uses variances to decide whether the means are different.

Understanding the underlying model of the Analysis of variance helps to understand how this rather complex statistical procedure works. X_{ij} represents the score (what we measured) for Person i in Condition j. μ represents the mean of all subjects who could theoretically be run in the experiment, and μ_{j} represents the population mean in Condition j. τ_{j} is the degree to which the mean of Condition j deviates from the grand mean ($\tau_{\mathrm{j}} = \mu_{\mathrm{j}} - \mu$). Finally, ϵ_{ij} is the amount by which Person i in Condition j deviates from the mean of his or her group ($\epsilon_{\mathrm{ij}} = X_{\mathrm{ij}} - \mu_{\mathrm{j}}$). The underlying model for the analysis of variance expresses a score in terms of the mean of the entire population plus an extra component due to the condition plus the unique contribution of each score:

$$X_{\mathrm{ij}} = \mu + \tau_{\mathrm{j}} + \epsilon_{\mathrm{ij}}$$

This basic model underlies all analysis of variance, and can be extended to more complex situations. ANOVA analysis are used to compare data from one variable among two or more groups (-> *1-way ANOVA*), but can as well be used to deal with two or more independent variables simultaneously, e.g. in a 2:2 experiment design (-> *Factorial ANOVA*).

The null hypothesis we want to test assumes that the means of the different conditions are the same:

$$H_0 = \mu_1 = \mu_2 = ... = \mu_\text{k}$$

The Analysis of Variance checks the variation (variance) *within* the groups, then works out how that variation would translate into variation (i.e. differences) *between* the groups, taking into account how many subjects there are in the groups. If the observed differences are a lot bigger than what one would expect by chance, we have statistical significance.

And this is how it is done: We can calculate a population variance as the mean of squared deviations (MS) about the mean, or as the sum of the squared deviations about the mean (SS) divided by the degrees of freedom:

$$\sigma^2 = MS = \frac{SS}{df} = \frac{\sum(X-\bar{X})^2}{N-1}$$

Let's assume we had n persons perform in k different conditions. In the following X_ij will again be the score (what we measured) for Person i in Condition j. $\bar{X}$ represents the mean of all subjects across conditions, and $\bar{X}_\text{j}$ represents the mean in Condition j.

To test the null hypothesis, we calculate two estimates of the population variance. The first estimate (MS_within) is the mean square error within the groups. This estimate is independed from the outcome of our null hypothesis, because we calculate this estimate seperately on each sample. We can obtain the variance within the groups as follows:

$$MS_\text{within} = \frac{SS_\text{within}}{df} = \frac{\sum_{j=1}^{k}\sum_{j=1}^{n_\text{j}}(X_\text{ij}-\bar{X}_\text{j})^2}{k(n-1)}$$

The second estimate assumes that the null hypothesis is true, and thus assumes that the means drawn from this population equals the variance of the population divided by the sample size. This treatment effect (MS_between) can be estimated as follows:

$$MS_\text{between} = \frac{SS_\text{between}}{df} = \frac{\sum_{j=1}^{k} n_\text{j}(\bar{X}_\text{j}-\bar{X})^2}{k-1}$$, where n_j means to do it once for each deviation.

Now we have two estimates of the populations variance. One of them (MS_{within}) is independent of the hull hypothesis, and the other one (MS_between) is only an estimate of the populations variance if the hull hypothesis is true. If the two estimates disagree sufficiently, we reject the hull hypothesis and conclude that the means between the groups are different. The test statistic that we can use is the F ratio, and can be expressed by:

$$\text{F}=\frac{MS_\text{between}}{MS_\text{within}}$$

Again, knowing the F ratio together with the degrees of freedom (df) provides the degree of confidence (p) that the observations could have occurred by chance. A probability p of 0.05 or less is commonly expected in order to reject the null hypothesis (this was that the means have occurred by chance) and conclude that there were significant differences among the means.

Analysis of Variance with more then two groups cannot easily be calculated by hand, but are done using statistics programs, which provide the significance test results, which is the F-value, the total number of degrees of freedom (df), and the significance level (p).

Regression Analysis

Regression can be used to answer questions of relationships, such as asking whether one continuous variable can be predicted or inferred by another. In regression analysis the dependent variable is assumed to be a function of independent variables, plus an error that accounts for all other factors. In *linear regression*, this function is linear:

$y_i = b_0 + b_1 x_i + b_2 x_2 + ... + b_i x_\text{i} + \varepsilon_\text{i}$, where

y_i is the dependent variable (to be predicted),
x_i are the independent variables (the predictors),
b_i are parameters (the regression coefficients),
ε_i is an error term

The goal is to find estimates for the b parameters, which indicate how a change in one of the independent variables affects the dependent variable.

Once a regression model has been constructed, we are interested in the overall fit of the model, or in other words the error of the prediction. R (or R^2) are commonly used to give an estimate of the relationship between the variables. We now want to test whether the model (and thus R) predicts the dependent variable better then chance. Testing the significance of R gives the p-value (and sometimes the F-value of the hypothesis test). If $p > .05$, we conclude that the regression model does not fit the data, and the data thus was independent, and not correlated (e.g. mental engagement could not predict availability, since p $>$ .05).

In **multiple regression** we have more than one predictor variable in the equation. E.g. when studying availability in Section 6.2.3, we tried to predict availability from different measures, such as level of mental engagement, level of physical engagement, and level of social engagement. When we have multiple predictor variables, we want to additionally know which of the predictor variables were statistically significant. This information can be found in the b, or the standardized β, coefficients. The standardized coefficients β (a linear transformation of b so that mean of data is 0 and standard deviation is 1) can be used to compare relative contributions of each predictor variable. Once we have the regression coefficients, standardized or not, we can test for significance, and again get a p-value describing the significance level for each predictor.

Intercoder Reliability (e.g., Krippendorff's Alpha, Cohen's Kappa)

Intercoder reliability, as well called intercoder agreement, means the extend to which independent coders (or judges) reach the same conclusion, and is a critical component of content analysis and interpretation of qualitative data. If no intercoder reliability can be shown, human interpretation of the data can not be considered valid. Intercoder reliability measures the extent to which different coders tend to assign exactly the same rating to each object. High levels of disagreement among coders can suggest weaknesses in research methods, operational definitions, categories, and training of coders.

There are lots of different measures of intercoder reliability. The simplest is to use *percent agreement* (e.g., "our coders agreed in 85% of the time"). However, despite its simplicity and widespread use, this measure does not account for agreement that could be expected to occur by chance (i.g., overestimate true agreement). Note that Chi-square cannot be used to calculate reliability, since it produces high values for both agreement *and* disagreement since they both differ from agreement expected by chance. An extended discussion of different measures can be found in [44].

One of the more widely used measures is *Krippendorff's alpha* score (α). An α score of 0 stands for no identical ratings at all, and 1 being only identical ratings. An alternative is *Cohen's kappa* (κ). The authors of [44] state that coefficients of .90 or greater are nearly always acceptable, .80 or greater is acceptable in most situations, and .70 may be appropriate in some exploratory studies for some indices. Higher criteria should be used for indices known to be liberal (i.e., percent agreement) and lower criteria can be used for indices known to be more conservative (i.e.Krippendorff's alpha). Because Krippendorff's alpha is an extremely conservative measure of intercoder reliability, a coefficient of .65 would actually be considered acceptable by many researchers.

List of Abbreviations

AGI	Asterisk Gateway Interface
AI	Artificial Intelligence
AJAX	Asynchronous JavaScript and XML
ANOVA	Analysis of Variance
ASR	Automatic Speech Recognition
CID	Caller ID, Caller Identification, Calling number Identification
DID	Direct Inward Dialing
DTMF	Dual-Tone Multi-Frequency
FAQ	Frequently Asked Question(s)
GSM	Global System for Mobile communications
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HTTP	Hypertext Transfer Protocol
IM	Instant Messaging
IP	Internet Protocol
IVR	Interactive Voice Response
MMS	Multimedia Messaging Service
PBX	Private Branch eXchange
PDA	Personal Digital Assistant
PSTN	Public Switched Telephone Network
PTT	Push To Talk
SIP	Session Initiation Protocol
SMS	Short Message Service
TCP	Transmission Control Protocol
TT	Touch-Talk
TTS	Text-To-Speech
VoIP	Voice over Internet Protocol
XML	Extensible Markup Language

List of Figures

List of Tables

Bibliography

[1] Piotr D. Adamczyk and Brian P. Bailey. If not now, when?: the effects of interruption at different moments within task execution. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–278, New York, NY, USA, 2004. ACM.

[2] Philip E. Agre. Welcome to the always-on world. *IEEE Spectr.*, 38(1):10–13, 2001.

[3] Jens S. Allwood. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–26, 1992.

[4] F. Ameka. Interjections: The universal yet neglected part of speech. *Journal of Pragmatics*, 18:101–118, 1992.

[5] Daniel Avrahami, Darren Gergle, Scott E. Hudson, and Sara Kiesler. Improving the match between callers and receivers: A study on the effect of contextual information on cell phone interruptions. *Behaviour & Information Technology (BIT)*, 26(3):247–259, 2007.

[6] Keni Bernardin and Rainer Stiefelhagen. Audio-visual multi-person tracking and identification for smart environments. In *Proceedings of the ACM Multimedia 2007*, Augsburg, Germany, 2007.

[7] Barry Brown and Rebecca Randell. Building a context sensitive telephone: Some hopes and pitfalls for context sensitive computing. *Comput. Supported Coop. Work*, 13(3-4):329–345, 2004.

[8] JJ Cadiz, Attila Narin, Gavin Jancke, Anoop Gupta, and Michael Boyle. Exploring pc-telephone convergence with the enhanced telephony prototype. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 215–222, New York, NY, USA, 2004. ACM.

[9] Steven M. Cherry. Communications: Im means business. *IEEE Spectr.*, 39(11):28–32, 2002.

[10] H. H. Clark and M. A. Krych. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1):62–81, 2004.

[11] Herbert H. Clark. *Using Language.* Cambridge University Press, 1997.

[12] Mihaly Csikszentmihalyi. *Flow: The Psychology of Optimal Experience.* Harper Perennial, New York, March 1991.

[13] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. Instant messaging: Effects of relevance and timing, 2000.

[14] Maria Danninger, Gopi Flaherty, Keni Bernardin, Hazim K. Ekenel, Thilo Köhler, Robert Malkin, Rainer Stiefelhagen, and Alex Waibel. The connector: Facilitating context-aware communication. In *ICMI '05: Proceedings of the 7th international conference on Multimodal Interfaces*, pages 69–75, New York, NY, USA, 2005. ACM Press.

[15] Maria Danninger, Gopi Flaherty, Keni Bernardin, Robert Malkin, Rainer Stiefelhagen, and Alex Waibel. The connector - facilitating context-aware communication. In *Proc. of the 2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI)*, 2005.

[16] Maria Danninger, Tobias Kluge, and Rainer Stiefelhagen. Myconnector: Analysis of context cues to predict human availability for communication. In *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, pages 12–19, New York, NY, USA, 2006. ACM Press.

[17] Maria Danninger, Erica Robles, Abhay Sukumaran, and Cliff Nass. *Computers in the Human Interaction Loop*, chapter The Connector Service - Representing Availability For Mobile Communication. Springer, 2008.

[18] Maria Danninger, Erica Robles, Leila Takayama, Qianying Wang, Tobias Kluge, Rainer Stiefelhagen, and Clifford Nass. The connector service - predicting availability in mobile contexts. In Steve Renals, Samy Bengio, and Jonathan G. Fiscus, editors, *MLMI*, volume 4299 of *Lecture Notes in Computer Science*, pages 129–141. Springer, 2006.

[19] Maria Danninger and Rainer Stiefelhagen. A context-aware virtual secretary in a smart office environment. In *Proceedings of the ACM Multimedia 2008*, Vancouver, Canada, 2008.

[20] Maria Danninger, Leila Takayama, Qianying Wang, Courtney Schultz, Joerg Beringer, Paul Hofmann, Frankie James, and Clifford Nass. Can you talk or only touch-talk? a voip-based phone feature for quick, quiet, and private communication. In *ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces*, New York, NY, USA, 2007. ACM Press.

[21] Maria Danninger, Qianying Wang, Michael Weiksner, Dean Eckles, Michael Camacho, Morgen Hentzen, and Clifford Nass. The one-way phone: Who gets to hit the receiver's mute button? In *CSCW '06: International Conference on Computer Supported Collaborative Work*, 2006.

[22] Alan Dix, Janet E. Finlay, Gregory D. Abowd, and Russell Beale. *Human-Computer Interaction (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.

[23] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 107–114, New York, NY, USA, 1992. ACM.

[24] Hazim K. Ekenel and Rainer Stiefelhagen. Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization. In *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, page 34, Washington, DC, USA, 2006. IEEE Computer Society.

[25] Thomas Erickson. Some problems with the notion of context-aware computing. *Commun. ACM*, 45(2):102–104, 2002.

[26] Thomas Erickson and Wendy A. Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.*, 7(1):59–83, 2000.

[27] James Fogarty, Scott E. Hudson, Christopher G. Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny C. Lee, and Jie Yang. Predicting human interruptibility with sensors. *ACM Trans. Comput.-Hum. Interact.*, 12(1):119–146, 2005.

[28] James Fogarty, Scott E. Hudson, and Jennifer Lai. Examining the robustness of sensor-based statistical models of human interruptibility. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 207–214, New York, NY, USA, 2004. ACM Press.

[29] Batya Friedman, Peter H. Kahn, Jennifer Hagman, Rachel L. Severson, and Brian Gill. The watcher and the watched: Social judgments about privacy in a public place. *Human-Computer Interaction*, 21:235–272, 2006.

[30] Erving Goffman. *Interaction Ritual: Essays in Face to Face Behavior*. Aldine Transaction, 1982.

[31] L. Gong, C. Nass, C. Simard, and Y. Takhteyev. When non-human is better than semi-human: Consistency in speech interfaces. In *Smith, M.J., Salvendy, G., Harris, D. and Koubek, R. eds. Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality*, pages 1558–1562, 2001.

[32] Steve Harrison and Paul Dourish. Re-place-ing space: the roles of place and space in collaborative systems. In *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 67–76, New York, NY, USA, 1996. ACM.

[33] Ken Hinckley and Eric Horvitz. Toward more sensitive mobile phones. In *UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 191–192, New York, NY, USA, 2001. ACM.

[34] Joyce Ho and Stephen S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 909–918, New York, NY, USA, 2005. ACM.

[35] Jim Hollan and Scott Stornetta. Beyond being there. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 119–125, New York, NY, USA, 1992. ACM.

[36] Eric Horvitz and Johnson Apacible. Learning and reasoning about interruption. In *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces*, pages 20–27, New York, NY, USA, 2003. ACM Press.

[37] David C. Howell. *Statistical Methods for Psychology*. Duxbury Press, 5 edition, 2002.

[38] James M. Hudson, Jim Christensen, Wendy A. Kellogg, and Thomas Erickson. "i'd be overwhelmed, but it's just one more thing to do": availability and interruption in research management. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 97–104, New York, NY, USA, 2002. ACM Press.

[39] Mizuko Ito, Daisuke Okabe, and Misa Matsuda Eds., editors. *Personal, Portable, Pedestrian: Mobile Phones in Japanese Life*. The MIT Press, October 2006.

[40] Brad Johnson and Saul Greenberg. Judging people's availability for interaction from video snapshots. In *HICSS '99: Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 1*, page 1026, Washington, DC, USA, 1999. IEEE Computer Society.

[41] Robert B. Katz. Automatic versus user-controlled methods of briefly interrupting telephone calls. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(2):321–334, 1995.

[42] Panu Korpipää, Jonna Häkkilä, Juha Kela, Sami Ronkainen, and Ilkka Känsälä. Utilising context ontology in mobile device application personalisation. In *MUM '04: Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia*, pages 133–140, New York, NY, USA, 2004. ACM.

[43] Steinar Kristoffersen and Fredrik Ljungberg. An empirical study of how people establish interaction: implications for cscw session management models. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1–8, New York, NY, USA, 1999. ACM Press.

[44] Matthew Lombard, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28:587–604, 2002.

[45] Steve Love and Mark Perry. Dealing with mobile conversations in public places: some implications for the design of socially intrusive technologies. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 1195–1198, New York, NY, USA, 2004. ACM Press.

[46] Robert Malkin. *Machine listening for context-aware computing*. PhD thesis, Carnegie Mellon University, 2006.

[47] Robert Malkin, Datong Chen, Jie Yang, and Alex Waibel. Multimodal estimation of user interruptibility for smart mobile telephones. In *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, pages 118–125, New York, NY, USA, 2006. ACM Press.

[48] D.C. McFarlane. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17 (1):63–139, 2002.

[49] Allen E. Milewski and Thomas M. Smith. Providing presence cues to telephone users. In *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 89–96, New York, NY, USA, 2000. ACM.

[50] Y. Miyata and D. A. Norman. Psychological issues in support of multiple activities. In D. A. Norman and S. W. Draper, editors, *User Centered System Design: New Perspectives on Human-Computer Interaction*, pages 265–284. Erlbaum, Hillsdale, NJ, 1986.

[51] Andrew Monk, Jenni Carroll, Sarah Parker, and Mark Blythe. Why are mobile phones annoying? *Behav. Inf. Tech.*, 23(1):33–41, 2004.

[52] Kristine S. Nagel, James M. Hudson, and Gregory D. Abowd. Predictors of availability in home life context-mediated communication. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 497–506, New York, NY, USA, 2004. ACM.

[53] Clifford Nass and Scott Brave. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press, 2005.

[54] Donald A. Norman. *The Design of Everyday Things*. Basic Books, September 2002.

[55] Brid O'Conaill and David Frohlich. Timespace in the workplace: dealing with interruptions. In *CHI '95: Conference companion on Human factors in computing systems*, pages 262–263, New York, NY, USA, 1995. ACM.

[56] Antti Oulasvirta, Mika Raento, and Sauli Tiitta. Contextcontacts: re-designing smartphone's contact book to support mobile awareness and collaboration. In *MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, pages 167–174, New York, NY, USA, 2005. ACM.

[57] Leysia Palen, Marilyn Salzman, and Ed Youngs. Going wireless: behavior & practice of new mobile phone users. In *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 201–210, New York, NY, USA, 2000. ACM.

[58] Elin Ronby Pedersen. Calls.calm: enabling caller and callee to collaborate. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 235–236, New York, NY, USA, 2001. ACM.

[59] Mika Raento, Antti Oulasvirta, Renaud Petit, and Hannu Toivonen. Contextphone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing*, 4(2):51–59, 2005.

[60] Emanuel A. Schegloff and Harvey Sacks. Opening up closings. *Semiotica 7*, 4:289–327, 1973.

[61] Bill N. Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *Workshop on Mobile Computing Systems and Applications*, pages 85–90, 1994.

[62] Bill N. Schilit, David M. Hilbert, and Jonathan Trevor. Context-aware communication. *IEEE Wireless Communications*, 9:46–54, 2002.

[63] Daniel Siewiorek, Asim Smailagic, Junichi Furukawa, Andreas Krause, Neema Moraveji, Kathryn Reiger, Jeremy Shaffer, and Fei Lung Wong. Sensay: A context-aware mobile phone. In *ISWC '03: Proceedings of the 7th IEEE International Symposium on Wearable Computers*, page 248, Washington, DC, USA, 2003. IEEE Computer Society.

[64] H. Soltau, F. Metze, C. Fügen, and A. Waibel. A one pass-decoder based on polymorphic linguistic context assignment. In *Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2001.

[65] Chris Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. *cvpr*, 02:2246, 1999.

[66] Lucy Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, Cambridge, 1987.

[67] D. Swinehart. Telephone management in the etherphone. In *Proc. IEEE Global Telecommunication Conference*, pages 392–402, 1987.

[68] Lorant Szasz-Toth. Open-set face recognition, 2007. Studienarbeit, Universität Karlsruhe (TH), Fakultät für Informatik.

[69] John C. Tang, Nicole Yankelovich, James Begole, Max Van Kleek, Francis Li, and Janak Bhalodia. Connexus to awarenex: extending awareness to mobile users. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 221–228, New York, NY, USA, 2001. ACM.

[70] G.H. TerHofte, R.A.A. Otte, H.C.J. Kruse, and M. Snijders. Context-aware communication with live contacts. In *CSCW '04: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, New York, NY, USA, 2004. ACM.

[71] Steve Whittaker, David Frohlich, and Owen Daly-Jones. Informal workplace communication: what is it like and how might we support it? In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 131–137, New York, NY, USA, 1994. ACM Press.

[72] Terry Winograd. Shifting viewpoints: Artificial intelligence and human–computer interaction. *Arti-icial Intelligence*, 170(18):1256–1258, 2006.

[73] Christian Wojek, Kai Nickel, and Rainer Stiefelhagen. Activity recognition and room-level tracking in an office environment. In *IEEE Int. Conference on Multisensor Fusion and Integration for Intelligent Systems - MFI06*, Heidelberg, Germany, September 2006.

[74] P.T. Zellweger, D.B. Terry, and D.C. Swinehart. An overview of the etherphone system and its applications. In *Proceedings of the 2nd IEEE Conference on Computer Workstations*, pages 160–168, 1988.