A Context-Aware Virtual Secretary in a Smart Office Environment

Maria Danninger
Universität Karlsruhe
Interactive Systems Labs
Karlsruhe, Germany
maria@ira.uka.de

Rainer Stiefelhagen
Universität Karlsruhe
Interactive Systems Labs
Karlsruhe, Germany
stiefel@ira.uka.de

ABSTRACT

A lot of the communication at the workplace - via the phone as well as face-to-face - occurs in inappropriate contexts, disturbing meetings and conversations, invading personal and corporate privacy, and more broadly breaking social norms. This is because both, callers and visitors in front of closed office doors, face the same problem: they can only guess the other person's current availability for a conversation.

We present a context-aware Virtual Secretary designed to facilitate more socially appropriate communication at the workplace. This service aims towards understanding a person's activity in smart offices, and passes on important contextual information to callers and visitors in order to facilitate more informed human decisions about how and when to initiate contact.

We have deployed this Virtual Secretary in the office of a senior researcher, mediating all his actual phone calls and in-person meetings for several weeks. With the Virtual Secretary active, the number of inappropriate workplace interruptions could be significantly reduced.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation (e.g., HCI)**]: User Interfaces[User-centered design]

General Terms

Design, Experimentation, Human Factors

Keywords

Computer-mediated communication, Context-awareness, Interruptibility, Availability, Virtual Secretary, Smart space, Workplace, Field experiment

1. INTRODUCTION

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. In particular the mobile phone has fundamentally changed our basic mechanics of social life by making communication possible irrespective of context. 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.

The problem is that callers place phone calls when convenient for them, but they know little to nothing about the receiver's situation. Such phone communications that occur in inappropriate contexts result in potentially disturbing others in close proximity and break the flow of concentration [5]. They can as well lead to invading personal and corporate privacy, and more broadly breaking social norms [20, 22]. It is perhaps for this reason that in 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, previous research found that at most half of all workplace phone calls lead to an immediate conversation, the others fail to reach intended recipients [6, 30].

This paper will discuss a context-aware system that acts 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 ultimately free users from inappropriate interruptions.

Deploying a context-aware Virtual Secretary required an instrumented environment - able to sense situations, recognize people and track activities. Employing the Virtual Secretary 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.

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. However, 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 [19].

Systems that make use of contextual information to actively intervene, such as our Virtual Secretary, must choose to what extent to place decision-making responsibilities with the technology, or with the humans. We will argue that availability for communication is extremely complex and hard to predict. Even if we had perfect sensors, human common sense is still a quality that Erickson has described as "(at best) awfully hard to implement" [9]. Our Virtual Secretary therefore relied on passing on relevant contextual information about the receiver's situation to callers and visitors, to help them take more informed decisions on how and when to initiate contact.

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

2. RELATED WORK

There has been a significant amount of effort placed on adding context to phone communication. Schilit et al have suggested a two-dimensional space for classifying such applications, based on a distinction between the "communication action" and "context acquisition" from user-driven to system-driven [26]: context can be entered manually, or sensed automatically, and the communication act can be achieved manually by the user, or autonomously by an agent. Such contextually appropriate communication technologies are articulated predominately in research on interruptibility and awareness.

2.1 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. Researchers have put extended efforts into automatically predicting people's in-the-moment interruptibility. Systems range from rule-based and user-driven [18] to sensor-based and system-driven [10] solutions. These systems appear across a variety of contexts ranging from the office [16] to the home [23] and to the mobile user [1, 12, 13].

For example, Horvitz et al. have examined models based on calendar information, PC activity and audio and video data [15]. Fogarty and Hudson et al. have studied an office environment, and 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 [10, 11].

Ho and Intille [13] 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. Hinckley and Horvitz [12] 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.

2.2 Awareness

Awareness - "an understanding of the activities of others which provide a context of your own activity" [7] - 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".

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 [14]. Brown and Randell [3], in their essay on context sensitive telephony, discussed the possibility of an automated agent that blocks calls on the behalf of users. They concluded that a better solution would be to provide context information to the caller to help the caller make a more informed decision about whether or not to initiate contact. A number of mobile awareness systems align well with this approach. People can share their context and see others' location and availability status with an interface similar to today's instant messaging buddy lists. Examples are "Context Phone" [25], "Awarenex" [29], "Live Addressbook" [21], "Calls.calm" [24].

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.

3. OVERVIEW & OUTLINE

In this paper, we will present a system that broadly follows the awareness approach: pass on relevant contextual cues to let the caller make a more informed decision about how and when to initiate contact.

However, we believe that user-driven context acquisition, such as manually updating a presence status, requires the kind of management that is so cumbersome as to be ignored completely by most users. Pre-planned information, such as from personal calendars, can be an unreliable source of information. This is why our system used perception technologies to detect the in-the-moment situation in smart offices. While this approach required an instrumented environment, and technology able to sense activity and situations, it is a much more reliable and less intrusive source of information for in-the-moment judgments, such as whether or not to initiate a phone call.

This paper will first introduce perception technologies that were installed in smart offices, and then describe how a context-aware Virtual Secretary mediated both phone and in-person interactions at the workplace.

Next, we will report results from deploying this Virtual Secretary in the office of a senior researcher, mediating all his actual phone calls and in-person office meetings over a period of several weeks. Through a controlled experiment we tried to empirically find out whether this Virtual Secretary could actually help to reduce inappropriate and untimely interruptions, e.g. during meetings or phone conversations.

4. PERCEPTION 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 installed in three smart offices in our research lab.

4.1 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 [31]. For this system, offices were equipped with cameras and omni-directional microphones. Since the system so far requires 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 [27]. 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

Figure 1: Camera snap shots of two of the smart offices with occupants present.

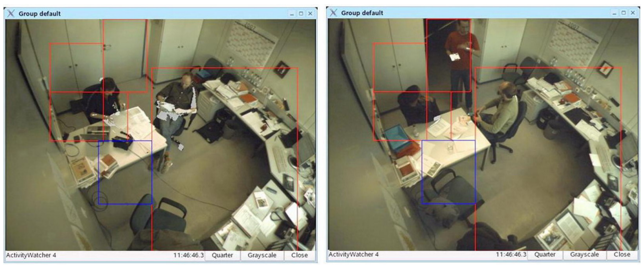

Figure 2: Camera snap shots of a third office in different meeting situations.

office. Figure 1 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 2 shows camera snap shots of meeting 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: whether the office was empty, if the occupant was alone, or in a face-to-face meeting with others (see Section 4.3). 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. We are currently working on technology that can actually track and distinguish single people [2], which is a much harder computer-vision task.

4.2 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 offices. This webcam sat on the screen that greeted visitors in front of our smart offices, and can be seen in the picture in Figure 6. The face recognition system used a local appearance-based algorithm that extracted features using discrete cosine transform, and is described in more detail in [8]. To be able to also distinguish unknown persons, it was necessary to enhance this system to open-set face recognition. We therefore extended it to accurately identify all members of our research lab, approximately 15 persons, and distinguish them from unknown persons, using Support Vector Machines [28].
The system worked very robustly, and could handle varying illuminations and face sizes, as can be seen in Figure 3. 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 [28].

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

Figure 3: An open-set face recognition system could recognize everyone in our lab (approx. 15 persons) and distinguish them from unknown persons.

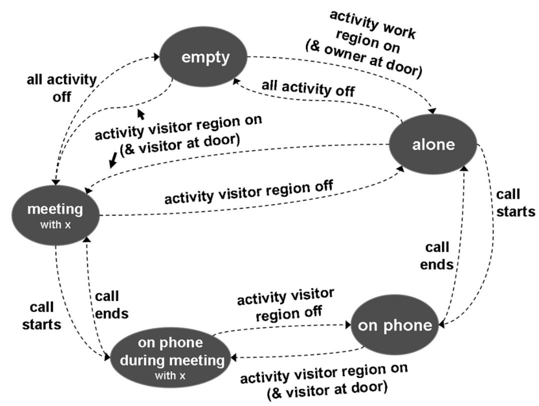

Figure 4: Finite state machine illustrates office activities and possible transitions between them.

4.3 Modeling Office Activity

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 (as will be discussed in Section 5.1), 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 4. 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 changes in lighting conditions, which caused the entire background to adapt. However, when transitions were too difficult, the system could fail to observe a new state, such as quickly leaving the office to get a cup of coffee.

4.4 Evaluation

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 who was very frequently occupied by meetings. An automated script recorded the current office situation as detected by our system. Simultane-

Table 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%

ously, 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%, the occupant was alone in 33.7%, and a face-to-face meeting took place in the remaining 11.4%. Overall results showed that the correct office state could be determined in 90.6%.

Table 1 shows that detection of meeting situations was most challenging. Meetings could be recognized in 88.0%. However, it happened that the system detected a meeting, when there was actually none (22.2% false positive rate). When the person was alone the system correctly detected the situation in 91.1%. Very rarely did the system detect that the person was alone, when he actually wasn't.

Given the relative simplicity of the perception technologies involved to detect activity in offices, these results were very encouraging. Of course, the system could be further improved by detecting more nuanced classes of activity, such as whether the person is working on the computer, or doing paperwork (as in [31]).

In the following, we will describe how a Virtual Secretary used contextual knowledge provided by perception technologies installed in our smart offices in order to mediate contact attempts.

5. THE VIRTUAL SECRETARY

We have developed a Virtual Secretary that mediated phone calls as well as face-to-face interactions in smart offices.

5.1 Mediating Phone Conversations

Currently, callers place phone calls when convenient for them, but they have little knowledge about the receiver's situation. 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.

5.1.1 Adding context to phone communication

The Virtual Secretary mediated all calls made to the office phone, warning senders at the time of the call when the current situation was likely to be inappropriate for talking on the phone, e.g. when a meeting was taking place, or when the person was engaged in another phone conversation. In order to maintain a reasonable level of privacy for the receiver, we programmed the system to moderate the levels of disclosure about the receiver's current situation based on the relationship between the office occupant and the caller (who was identified by caller ID). Here is an example of what a sender could hear when placing a call:

Virtual Secretary [synthesized voice]: *"Hello Jane. Thank you for calling Bob Smith. 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."*

5.1.2 The caller decides

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. If the caller chose to leave a message, the receiver's phone did not ring. Instead, the receiver was sent an audio attachment via email, thus mitigating an unnecessary interruption while still transmitting relevant information.

Email from the Virtual Secretary: *Dear Bob: just wanted to let you know you were left a 0:15 long message (number 12) in your mailbox from <0721608xxxx>, so you might want to check it when you get a chance. Thanks! –Your Digital Secretary.*

(Attachment: msg-2007-11-05-003.wav)

The Virtual Secretary could also assist with re-scheduling the call at a more convenient time. If the sender opted to schedule a call, the Virtual Secretary proactively initiated a call back when the receiver next became available, as determined by the situation in the office:

Virtual Secretary: *"Hello Jane, this is Bob's Virtual Secretary. Bob is now available. Please hold while your call is being connected, or you may hang up the 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, our system initiated the call back 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 a smart office, the scheduled call was initiated only when both parties were simultaneously available.

5.1.3 Network-based implementation

The Virtual Secretary was implemented entirely on the network and could be immediately deployed from any phone capable of sending DTMF signals (aka touch tones). 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 (this experiment will be described in Section 7).

People using the system were assigned a local direct inward dialing number (DID) hosted by a telephone company, and connected to our custom telephone system (Figure 5). This telephone system, a custom private branch exchange (PBX) network, was set up to act as Voice over Internet Protocol (VoIP) router. Asterisk^{TM} was used as open source software implementation of such a telephone PBX, and the Cepstral text-to-speech engine was used to dynamically generate voice responses on the server.

We configured our telephone system to act as Virtual Secretary. Interactive voice responses (IVR) were used to interact with callers. The concept is familiar as many business applications, e.g. call centers, telephone voting systems, etc., employ this technology: phone menus are fed to the caller, allowing choices such as "Please press 1 for English, and 2 for Spanish.". The Asterisk Gateway Interface (AGI) allows adding Java programming functionality, such as database access, to control the dial plan. Phone numbers, relationships between people and other settings were stored in a local database. Thus, changing the behavior of the Virtual Secretary was as simple as a changing an entry in the local database, and was effective immediately with the next incoming call. As calls were presented to the PBX system, the number that the caller dialed was given, so that the system could route the call to the actual office or mobile phone number.

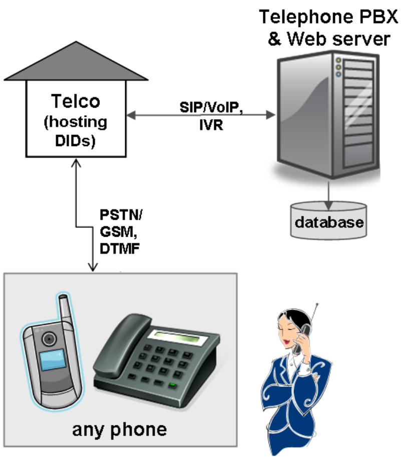

Figure 5: Deploying a private branch exchange system allowed phone calls to be controlled, monitored and logged by a server in a VoIP network.

Figure 6: 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.

Figure 7: The Virtual Secretary informs visitors about the current situation inside of the office. Known visitors get more detailed and personal context information, such as why the person is busy or who is in a meeting.

When testing a previous system with a similar setup, we learned that it can be very difficult to get people to call a new number, especially when a system is tested over a limited period of time [6]. This problem is especially critical if one does not know who the callers will be. Therefore, 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 put through.

All in all, this setup allowed calls connecting two parties in traditional landline or cellular networks to be (1) controlled, (2) monitored, and (3) logged by a server in a VoIP network.

5.2 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. Some 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.

The research team placed a screen, speakers and webcam outside the office door (see Figure 6), enabling the Virtual Secretary to communicate with visitors while capturing identity information via facial recognition (as was described in Section 4.2). 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). The goal was to prevent untimely disruptions from taking place, such as during meetings or phone conversations. As with phone-based interactions, the relationship between the visitor and the occupant determined different levels of information disclosure about the circumstances within the office. Finally, the system maintained a record of all visitors so that if the office occupant had been out for some time, they could see a list of missed connections upon their return.

5.3 Web-Based Office Activity Diary

The Virtual Secretary provided a web-based diary of all office activity. 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], as shown in Figure 8. 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, some examples are shown in Figure 9. E.g., clicking on one of the meeting elements (orange color) opened an info bubble with details about the meeting, including a snap shot of the situation. This snap shot helped remind people what the meeting was like, or who it was with, and could be

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

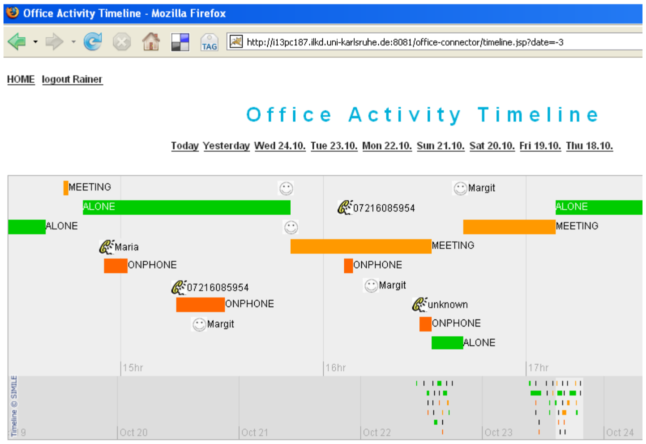

Figure 8: The Office Activity Diary was shown as interactive time line widget. The two bands were synchronized and could be scrolled indefinitely by dragging with the mouse pointer.

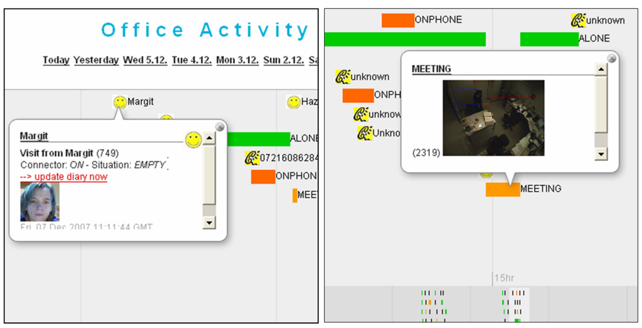

Figure 9: When clicking on any element in the activity time line, an info bubble displayed further details, such as snap shots of the situation, or image of the visitor.

enlarged to full resolution by clicking on it. 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.

6. OVERALL SYSTEM ARCHITECTURE

In our smart offices, multiple perception technologies, actuators, 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. We addressed the problem 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 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, e.g. between perception components and service modules. A central

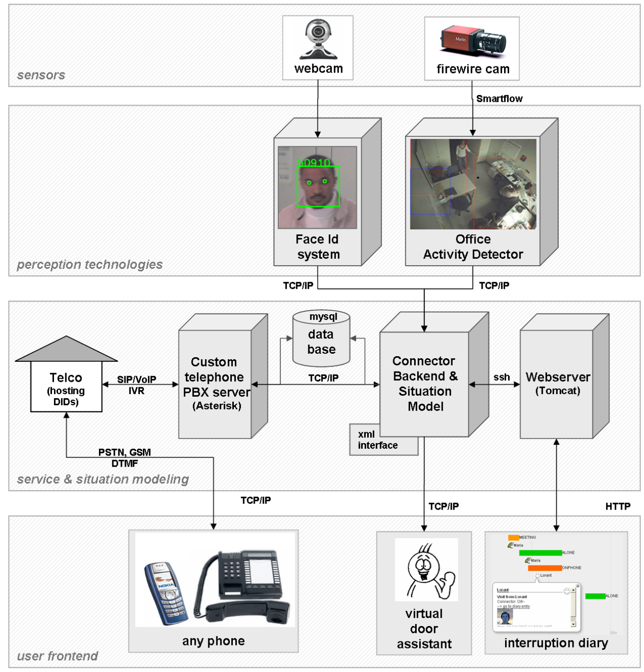

Figure 10: Overall architecture diagram.

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.

The overall system architecture is illustrated in Figure 10. The system could be easily extended. e.g. by other technologies on the sensor layer.

7. EVALUATION

The Virtual 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 or inappropriate interruptions, from the perspective of the receiver. 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 6 shows how a visitor consults the Virtual Secretary at the office door, and is informed that the person she was trying to see was currently not present.

7.1 Experiment Design

In order to explore whether the Virtual Secretary was able to make a measurable difference in reducing inappropriate workplace interruptions, 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.

Figure 11: 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.

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 5.1). The proactive call back feature was disabled for the duration of the study. 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.

7.2 Data and Measurements

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 11). 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.

7.3 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. Moreover, 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 way the system was set up 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 across networks on call forwarding. Our participant mentioned that he would have liked to keep the Digital Secretary permanently after the duration of the study, if only the issue with caller IDs could be resolved.

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. Only a single 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, since we tested the system with real-life interactions. 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, user observations and informal interviews showed that all parties felt that the Virtual Secretary interface was relatively easy to interact with; most people seemed to find the system understandable despite the lack of any prior experience with it. Thus, we could be relatively confident that evaluations of system efficacy were not particularly affected by any usability breakdowns.

7.4 Performance of the Virtual Secretary

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.

7.4.1 *Comparing call success*

Figure 12 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 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.

7.4.2 *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 inter-

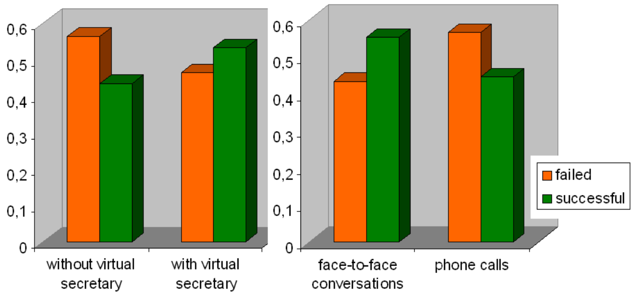

Figure 12: Success of contact attempts across modalities and study conditions.

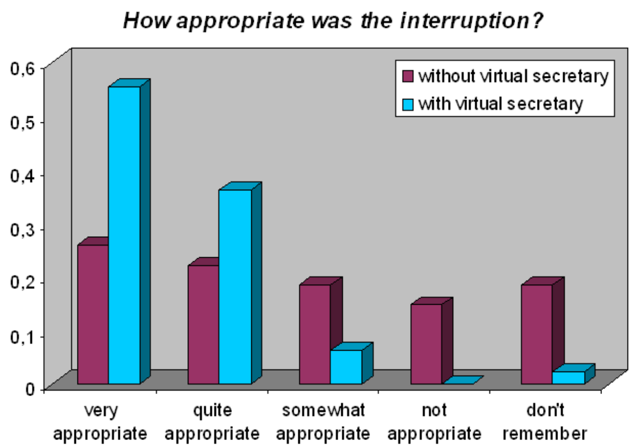

Figure 13: Ratings of how appropriate interruptions were show that there were significantly fewer inappropriate workplace interruptions with the Virtual Secretary.

rupted 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 (Figure 11) were extremely useful, and helped him remember the exact situation when the interruption occurred.
We subsequently analyzed the diary ratings for each social interaction to determine whether or not the Virtual Secretary affected the number of inappropriate interactions experienced by the office occupant. Figure 13 shows distributions of how appropriate interruptions were. Where the Virtual Secretary mediated interactions, there were no interruptions in the "not appropriate" category, and only 6% of interruptions fell into the "somewhat appropriate" category, as compared to 19% "not appropriate" and 15% "somewhat appropriate" interruptions in the control condition.
We used a Chi-square 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 ($X^2 = 19.93$, $df = 4$, $p < 0.002$). Thus, using the Virtual Secretary in smart offices did in fact help to decrease the amount of inappropriate interruptions in our study.

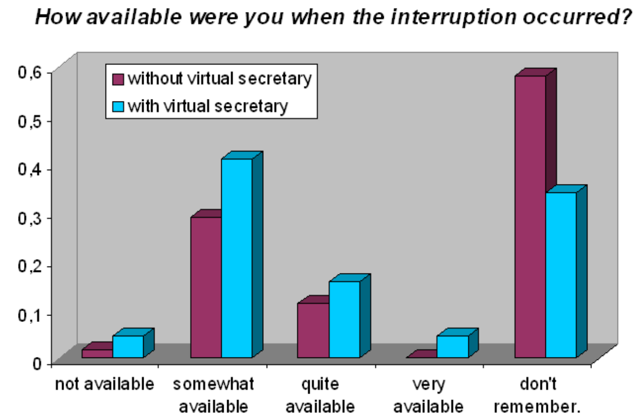

Figure 14: Ratings of general availability showed no differences between conditions.

7.4.3 Comparing general availability

Given the reduction in inappropriateness of interruptions, we might expect that the office occupant would indicate greater availability for communication during those times. We looked at how available the person had reported to be at times when interruptions occurred (Figure 14).
However, analysis indicated that it was quite difficult to self-report availability with much precision. There were no significant differences in how available our participant was when interruptions occurred between the different conditions. Most of the time, the office occupant indicated that he had been "somewhat available" (64%) or "quite available" (25%), rather than being in a state of "very available", or "not at all available". This result may derive from a general expectation that people should be available while in the workplace.
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. There is a general expectation that people should be available while in the workplace. 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. We will further discuss implications of this finding in the following Section.

8. DISCUSSION

Deploying the 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.

8.1 Understanding Availability

Systems that make use of contextual information to actively intervene must choose to what extent to place decision-making responsibilities (e.g. best time and media to communicate) with the technology (system-driven), or with the humans (user-driven).

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. This is perhaps not surprising, since it has been shown that even humans show variations when judging people's general availability in common office situations. 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 [17]. People displayed very different assessments of availability when interpreting particular situations.

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.

8.2 Dealing with Privacy

However, privacy is a major concern with any system that exposes personal information and our Virtual Secretary is no exception. Such systems must balance the trade-offs between making information public and protecting the user's senses of privacy and information ownership.

A mediator such as our Virtual Secretary can to some extend enable privacy decision making to be abstracted away, by transmitting various details of personal information (as we did), or simply collapsing complex ensemble of personal information points to a single indicator: available or not? However, 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.

8.3 Designing for Mobility

Our Virtual Secretary helped to increase common ground [4] for office communications, and decrease the number of untimely interruptions, via the phone as well as in-person. However, this system was tied to a static context, the office, and mediated mostly business conversations. Office phones have the great advantage of being associated to a location - someone's office - which provides a limited set of possible receiver states, and some place-based guidance for both caller and receiver about the social expectations of the office space. E.g. the participant in our study did never feel extremely available or unavailable when in his office, since he expected interruptions.

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. However, it is much more expensive and technically very difficult to develop and deploy sensing technologies in everyday mobile environments, such as on smart phones [12, 13]. It is especially troublesome to carry out sensing on mobile platforms, given limited processing power, battery life, and sensor obtrusiveness. 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 we go, even if this is where we would need them most.

9. CONCLUSIONS

We have presented a context-aware Virtual Secretary that mediated phone and in-person communications in smart offices. Our smart offices were equipped with cameras to detect the in-the-moment situation, and recognize people in front of the office door.

The Virtual Secretary passed on important contextual information about the receiver's current situation, such as whether the person was currently present and if the person was alone, in a meeting situation or on the phone. The goal was that both callers and visitors could take more informed decisions on when and how to initiate contact.

In a controlled experiment we deployed the Virtual Secretary in the office of a senior researcher, where it mediated all his real-life interactions over a period of several weeks. We could empirically show that this person received significantly fewer inappropriate workplace interruptions with the Virtual Secretary active.

We discussed the benefits and burdens of this user-centered decision-making approach, including privacy concerns, and the challenges that we might face when deploying a similar system in a mobile context.

10. ACKNOWLEDGMENTS

We would like to thank our colleagues Hazim K. Ekenel and Lorant Toth who developed the face recognition system, as well as Keni Bernardin who helped set up the computer-vision technologies in the smart offices. Thanks as well to Prof. Cliff Nass and Leila Takayama from Stanford University for sharing their ideas, and support with the user study.

This work has been partly funded by the European Commission under Project CHIL (http://chil.server.de, contract nr. 506909).

11. REFERENCES

[1] D. Avrahami, D. Gergle, S. E. Hudson, and S. 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.

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

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

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

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

[6] M. Danninger, L. Takayama, Q. Wang, C. Schultz, J. Beringer, P. Hofmann, F. James, and C. 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.

[7] P. Dourish and V. 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.

[8] H. K. Ekenel and R. 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.

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

[10] J. Fogarty, S. E. Hudson, C. G. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. C. Lee, and J. Yang. Predicting human interruptibility with sensors. *ACM Trans. Comput.-Hum. Interact.*, 12(1):119–146, 2005.

[11] J. Fogarty, S. E. Hudson, and J. 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.

[12] K. Hinckley and E. 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.

[13] J. Ho and S. 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.

[14] J. Hollan and S. 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.

[15] E. Horvitz and J. 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.

[16] J. M. Hudson, J. Christensen, W. A. Kellogg, and T. 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.

[17] B. Johnson and S. 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.

[18] P. Korpipää, J. Häkkilä, J. Kela, S. Ronkainen, and I. 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.

[19] S. Kristoffersen and F. 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.

[20] S. Love and M. 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.

[21] A. E. Milewski and T. 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.

[22] A. Monk, J. Carroll, S. Parker, and M. Blythe. Why are mobile phones annoying? *Behav. Inf. Tech.*, 23(1):33–41, 2004.

[23] K. S. Nagel, J. M. Hudson, and G. 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.

[24] E. R. 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.

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

[26] B. N. Schilit, D. M. Hilbert, and J. Trevor. Context-aware communication. *IEEE Wireless Communications*, 9:46–54, 2002.

[27] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. *cvpr*, 02:2246, 1999.

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

[29] J. C. Tang, N. Yankelovich, J. Begole, M. V. Kleek, F. Li, and J. 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.

[30] S. Whittaker, D. Frohlich, and O. 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.

[31] C. Wojek, K. Nickel, and R. 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.