TALKING ASSISTANT: A SMART DIGITAL IDENTITY FOR UBIQUITOUS COMPUTING

Erwin Aitenbichler, Jussi Kangasharju, Max Mühlhäuser *

Abstract

A predominant issue in a ubiquitous computing world is that of future personal devices in the post-PC era. In this paper we present the Talking Assistant *(TA), a prototype device we have developed to serve as a personal device. The main function of the Talking Assistant is to serve as the user's digital identity. In addition, the TA features local processing and storage, wireless networking capabilities, and a set of sensors. We present the rationale behind developing the TA and discuss possible use case scenarios from the real world.*

1. Introduction

The vision of the future ubiquitous computing world is very attractive. Computers have become interwoven with our everyday environment and we use them naturally, without even noticing that they are there. However, to use such an infrastructure, we still need a way of interacting with the ubiquitous devices. Interacting with the infrastructure is likely to happen in an unobtrusive manner; this calls for investigating methods and devices for interacting with such an ubiquitous infrastructure. In this paper, we present our vision of such an interaction device, namely, the Talking Assistant headset (TA).

The TA is part of our Mundo [4] project which aims at investigating and prototyping ubiquitous computing infrastructures. It fulfills two critical functions. First, the TA acts as the user's representative in the digital world. By this we mean that the Talking Assistant carries the user's digital identity and is able to perform operations such as authenticating the user to other parties. We believe that interaction with the ubiquitous infrastructure will mainly happen through a personal device, owned by each user. This is because it will represent the user in the digital world and carry out transactions, sometimes with only implicit consent of the user. Hence, it is vital that a user is able to *trust* the device carrying out these actions. The easiest way to instill this trust in users is to have them actually *own* the device instead of relying on a networked service, which may remain an abstract concept for many users.

Second, the TA plays a role as an interface device for interacting with the computers and services of the world. By itself the TA has only minimal processing, storage, and interaction capabilities, but it can boost them by using services that are deployed in the infrastructure. This way, we can keep the device small (both in terms of size and power consumption), offer basic services in all situations, and offer enhanced services when connected to a network. Should network connectivity not be available, the Talking Assistant must still be capable of providing a minimal functionality.

*Telecooperation Group, Department of Computer Science, Darmstadt University of Technology, Germany, {erwin,jussi,max}@tk.informatik.tu-darmstadt.de, http://www.ubicomp.tk/

2. Mundo

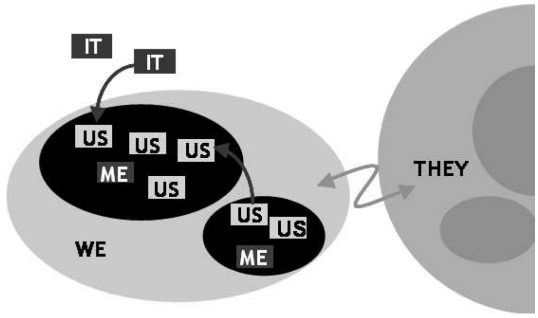

Figure 1. Mundo vertical architecture

In this section, we provide a brief overview of the Mundo project. Figure 1 shows an overview of our Mundo architecture. A more complete description can be found in [4]. Mundo has five different entities: ME, US, IT, WE, and THEY. The names have been chosen to reflect their functionality and role in Mundo, as described below.

ME (Minimal Entity): We consider ME devices as the representation of their users in the digital world. The personal ME is the only entity always involved in the user's activities. We have decided to base our architecture around a single, minimal device, because such a device is easy to carry around in all situations. Compare this to a normal PDA which typically does not comfortably fit in a pocket, and needs to be carried in a bag. Even a mobile phone may be too big to be carried around while jogging, for example. Our goal in developing Mundo around a Minimal Entity is to identify the fundamental requirements of the ubiquitous world. In other words, what functionality must be available to the user at all times? As we present below, any ME can augment its capabilities through association with other entities. However, the user is not required to carry around many devices with her.

In terms of pure functionality, i.e., holding the user's digital identity, such a ME device could potentially be constructed in a very simple manner, for example, based on an RFID chip or similar [3, 12]. We have decided to augment the basic functionality of the ME by adding audio input/output capabilities. Although this makes the device larger, we believe that the demand for audio communications is ubiquitous, as demonstrated by the popularity of mobile phones. Therefore, it is likely that users will (almost) always carry an audio device with them anyway. This also provides us with a very convenient way of interacting with the ME through voice commands and audio output.

US (Ubiquitous aSsociable object): Minimalization pressure will not permit feature-rich MEs in the near future. Hence, they must be able to connect to devices such as memory, processors, displays, and network devices in an ad hoc manner. We call this process *association* and such devices *ubiquitous associable objects* (US). Through association, the ME can personalize the US to suit the user's preferences and needs. For privacy reasons, any personalization of an US becomes unavailable if it is out of range of the user's ME.

IT (smart ITem): There are also numerous smart items that do not support association that would turn them into an US. Vending machines, goods equipped with radio frequency IDs, and landmarks with "what is" functionality are just a few examples. We call such devices *smart items*. An IT is any digital or real entity that has an identity and can communicate with the ME. Communication may be active or passive. Memory and computation capabilities are optional.

WE (Wireless group Environment): We expect ad hoc networking to be restricted to an area near to the user of a ME device, as connections with remote services will involve a non-ad hoc network infrastructure. The functionality of a WE is to bring together two or more personal environments consisting of a ME and arbitrary US entities each. A WE makes connections between the devices possible and also allows for sharing and transferring hardware (e.g., US devices) and software or data between WE users.

THEY (Telecooperative Hierarchical ovErlaY): We regard overlay cells as the backbone infrastructure of Mundo. These *telecooperative hierarchical overlay cells* (THEY) connect users to the (non-local) world, deliver services, and data to the user. THEY support transparent data access, for example, frequently used data may be cached on US devices. They also offer cooperation between different physical networks, transparent to the users.

3. Talking Assistant

From the general properties of a ME device described above and a number of application scenarios, like an interactive tour guide [1] and lecture support [9], we derived the following requirements:

Voice-centric: The device must feature audio in- and stereo output via headset, local speech recognition (limited vocabulary of 10–20 words) and audio stream compression/decompression. The primary use is that of an ears-and-mouth device. We believe that audio communication is an ubiquitous demand and that users would in any case be carrying around a device for audio communication. Displays are considered optional and can be implemented as US devices.

Networked: A network interface permits the ME to communicate with other Mundo entities. The underlying network technology must support spontaneous networking, a reasonable number of clients per area, and sufficient bandwidth for audio streaming. For example, WLAN can fulfill these requirements, but PAN technologies like Bluetooth not.

Context-aware: Incorporating context information promises to enable ”smarter” applications and easier-to-use ubiquitous user interfaces. For our smart office and museum guide scenarios, we require location sensors with an accuracy in the centimeter range and measurement of orientation.

Open architecture: For our research, we require an extensible hardware architecture to be able to add additional sensors. The operating system on the ME must be a modern 32-bit OS with networking and multi-threading in order to run our communication middleware MundoCore. Network and communication interface drivers are required to be open-source, because our WLAN-based location system [7] needs modifications at driver level.

Small: Since our goal is to have each user carry this personal device, it is vital that it is small and easily carried in any circumstances. Building earplug-size MEs in our research work is out of scope. However, we try to build our prototypes as compact as possible. In a few years, advanced industrial manufacturing techniques should permit implementing the ME as small as nextlink’s earplug-size Bluetooth headsets (`http://www.nextlink.to`).

Since no PDA or Smartphone currently on the marked fulfilled our requirements, we started to build own hardware. We consider the TA as a first prototype towards a ME device. We show how the TA is constructed, what capabilities it has, and how it integrates with other systems we have already built.

Figure 2. Talking Assistant headset

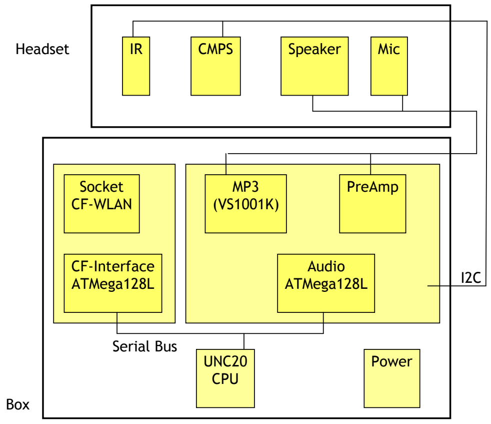

Figure 3. Hardware design

3.1. Talking Assistant Hardware

The TA is shown in Figure 2 and comprises two parts: the actual headset and a small box. CPU, WLAN card and battery are located in the box. The headset consists of a headphone, microphone and a number of sensors. These two parts have been split up, because earlier versions of the TA which had everything built onto the headset turned out to be heavy and impractical to wear. In our current prototype version, the user wears the headset on her head and wears the box on her belt. Because the headset only contains a few sensors, it is no more cumbersome to wear than a standard headset without any sensors. All the sensors are housed in the small box which is attached to the top of the headset and connected to the main TA box through the cable as shown in the figure. Figure 3 shows a high-level overview of the components in the Talking Assistant and their locations (i.e., whether on the headset or in the box).

The main CPU is an ARM7 variant run at 55 MHz. The system has 16 MB RAM and 8 MB flash memory. It runs uClinux with a 2.4.20 kernel. A number of Atmel 8-bit microprocessors are dedicated to special functions. These processors are interconnected by a serial bus. A multiprocessor communication protocol supports packet-oriented push-communication between the processors.

3.2. Audio

The audio hardware consists of a DSP that supports direct playback of PCM and MP3-encoded audio data. For MP3, all sample rates up to 48 kHz may be decoded from bitstreams up to 192 kbit/s. Normally, variable bitrate files with up to 320 kbit/s peaks may be played without glitches. We typically use 128 kbit/s, which is considered almost HiFi-quality. MP3 streaming provides very good audio quality at reasonable network bandwidth requirements. The audio input hardware includes a microphone pre-amplifier with a wide gain range and supports sampling rates up to 22 kHz. We operate it typically at 16kHz because this sampling rate matches that of the speech recognition software we plan to use on the TA.

3.3. Sensors

The sensors are used for determining the position of the mobile user in space and capturing the head orientation in all three rotational axes. The combination of sensors is essential for best operation. An acceleration sensor is used to measure the tilt angles. The heading is measured with an electronic compass which must usually be operated at level to the ground. When there is significant tilt present, the heading value is compensated depending on the pitch and roll values measured by the acceleration sensor. Finally, our infrared location system (described in Section 3.4.) uses the heading to determine the direction in which to emit beacon signals.

The acceleration sensor measures both dynamic acceleration (e.g., vibration) and static acceleration (gravity). We use it as a dual-axis tilt sensor. The force of gravity is used as an input vector to determine the tilt of the device in space. Three magnetometers mounted orthogonal to each other are used to get the compass heading. The measured values are proportional to the magnetic field strengths in the respective directions. With an affine transformation, these values are then rotated into the horizontal plane. The rotation angles are given by the pitch and roll information provided by the accelerometer. The compass has an accuracy of 3–4 degrees at tilt angles up to ± 40 degrees.

3.4. Integration with Local Positioning Systems

We will now present a brief overview of our low-cost infrared local positioning system [2], which we have developed in parallel. The eight IR emitter diodes on the TA, each covering an angle of about 45 degrees, send out modulated signals which permit the receiver to distinguish between multiple users. Each room where we want to use the positioning system is covered by two webcams set on a rail as a stereo camera and equipped with wide-angle lenses and infrared-filters. When a user walks around in the room, the signals from the headset are captured by the stereo camera and we can calculate the user's position through triangulation. The system offers an accuracy of about 8 cm in near range and about 16 cm when covering about 100m^2. A detailed description and evaluation can be found in [2]. The combination of positioning system, compass and tilt sensors enable us to determine the direction of the user's gaze and what object she is looking at. The positions of objects are known from a detailed world model [1]. In case the high-resolution system is not required, we also have a system based on WLAN signal strength, which requires less support by infrastructure [7].

4. Related Work

The CardBIT terminal [6] addresses issues similar to the ones we study with the ME. It is a credit-card-sized, battery-less information terminal that realizes location-and-direction-based interaction. The device only consists of a reflective sheet for tracking and a speaker that is directly connected to a solar cell. Therefore, the system relies on infrastructure for tracking, supports only a single client and audio transmission requires a line of sight to a strong light source modulated with the audio signal.

The Personal Server [12] is a small device which enables the user to access data and applications through interfaces in the user's environment. However, it relies on the existence of an infrastructure, since the device itself has no means of interacting with the user. Furthermore, the TA can act as a trusted terminal that is required for secure transactions.

The main goals of SoapBox [11] are to function as a sensor or an interface device. In particular, it does not include the possibility to store a digital identity and mainly appears to be designed as a small, independent sensor which sends out signals to enable other devices to perform their tasks. In this sense, it is more like an IT in our Mundo architecture.

There are currently many different approaches for carrying data and having it always available. Some examples are MetaPad (`http://www.research.ibm.com`), Pockey (`http://www.`

pockey.co.kr), Apple's iPod (http://www.apple.com/ipod/), and the Q Drive (http://www.agatetech.com/products_q.html). Note that these devices only provide mobile data storage and access; they do not contain any features which would make them a ME device. They also require support from a host computer and must be physically connected to the host computer for accessing any data on them. In contrast, the TA has built-in wireless networking.

Approaches based on thin clients, such as VNC [5], Roma [8], Infotab [10], and Parctab [13], are quite different from our approach with the TA. Thin clients require support from a networked server and are not capable of independent operation. In addition, most such solutions assume a relatively low-latency and high bandwidth connection to the server (or home agent) which might not be feasible in many cases. The TA is not dependent on any particular server, but can interact with any servers in the network. It is even capable of autonomous operation in the absence of a network connection.

5. Conclusion

In this paper, we have presented the Talking Assistant headset. Its main features are (i) that it contains a digital identity for trust establishment, (ii) audio input/output for voice-based interaction and (iii) context awareness through its sensors. Initial results strongly indicate that the TA and its associated positioning system can provide enhancements in several application fields.

References

[1] E. Aitenbichler and M. Mühlhäuser. The Talking Assistant Headset: A Novel Terminal for Ubiquitous Computing. Technical Report TK-02/02, Telecooperation Group, Darmstadt University of Technology, 2002.

[2] E. Aitenbichler and M. Mühlhäuser. An IR Local Positioning System for Smart Items and Devices. In *3rd International Workshop on Smart Appliances and Wearable Computing*, Providence, RI, May 2003.

[3] M. D. Corner and B. D. Noble. Zero-Interaction Authentication. In *Proceedings of MOBICOM*, Atlanta, GA, September 23–28, 2002.

[4] A. Hartl, E. Aitenbichler, G. Austaller, A. Heinemann, T. Limberger, E. Braun, and M. Mühlhäuser. Engineering Multimedia-Aware Personalized Ubiquitous Services. In *Proc. of IEEE MSE*, Newport Beach, CA, December 2002.

[5] S. F. Li, M. Spiteri, J. Bates, and A. Hopper. Capturing and Indexing Computer-based Activities with Virtual Network Computing. In *Proceedings of ACM Symposium on Applied Computing*, Como, Italy, March 19–21, 2000.

[6] Y. Nakamura, T. Nishimura, H. Itoh, and H. Nakashima. A System of Card Type Battery-Less Information Terminal: CardBIT for Situated Interaction. In *First IEEE International Conference on Pervasive Computing and Communications (PerCom'03)*, pages 369–377. IEEE Computer Society, 2003.

[7] Y. Song. In-House Location Tracking. Diplomarbeit. FG Telekooperation, Technische Universität Darmstadt, 2002.

[8] E. Swierk, E. Kiciman, N. Williams, T. Fukushima, H. L. Yoshida, and M. Baker. The Roma Personal Metadata Service. *Mobile Networks and Applications*, 7(5), September/October 2002.

[9] C. Trompler, E. Aitenbichler, and G. Rößling. Intelligent Headsets for Supporting Digital Lecture Halls. Proceedings of the 2003 International Workshop on Multimedia Technologies in ELearning and Collaboration, 2003.

[10] T. Truman, T. Pering, R. Doering, and R. Brodersen. The InfoPad Multimedia Terminal: A Portable Device for Wireless Information Access. *IEEE Transactions on Computers*, 47(10), October 1998.

[11] E. Tuulari and A. Ylisaukko-oja. SoapBox: A Platform for Ubiquitous Computing Research and Applications. In *Proceedings of Pervasive Computing*, Zurich, Switzerland, August 2002.

[12] R. Want, T. Pering, G. Danneels, M. Kumar, M. Sundar, and J. Light. The Personal Server: Changing the Way We Think about Ubiquitous Computing. In *Proceedings of UbiComp 2002*, Gothenburg, Sweden, September 2002.

[13] R. Want, B. Schilit, A. Norman, R. Gold, D. Goldberg, K. Petersen, J. Ellis, and M. Weiser. An Overview of the ParcTab Ubiquitous Computing Environment. *IEEE Personal Communications*, 2(6):28–43, December 1995.