Talking Assistant – Car Repair Shop Demo

Dirk Schnelle Erwin Aitenbichler Jussi Kangasharju Max Mühlhäuser
Telecooperation
Department of Computer Science
Darmstadt University of Technology
Germany
{dirk,erwin,jussi,max}@tk.informatik.tu-darmstadt.de

ABSTRACT

In this video paper we present the Talking Assistant and the STAIRS project, and how the two interact. The Talking Assistant is a device for interacting in ubiquitous computing environments. It features sensors, wireless communications and simple local speech recognition. The STAIRS project concerns browsing of structured hypertext documents in audio. One key feature of STAIRS is, that beside speech commands, navigation can be controlled by changes in context. In this paper we show one example of how context changes can be detected with infrared tags.

1. INTRODUCTION

The video "Talking Assistant – Car Repair Shop Demo" presents the Talking Assistant (TA) headset, a ubiquitous interaction device, and the STAIRS project which investigates navigating information over audio. The TA and STAIRS are both general-purpose solutions to a wide array of situations, and in this video we show how they can help the work in one particular context, a car repair shop.

The Talking Assistant [1] is an audio-based device intended as a general purpose personal device for ubiquitous computing environments. It provides a well-defined minimal functionality which permits very small form factors, allowing the user to carry the device always with her. Audio devices do not need large displays to offer useful interaction possibilities, and through advanced manufacturing techniques, could be reduced to ear-plug size in a few years. We describe the TA in more detail in Section 3.

The STAIRS project[1] develops mechanisms for delivering structured information over an audio interface. Examples of such information are hypertext, user manuals, procedure checklists, etc. The STAIRS project aims at producing a general framework for building and delivering audio-based information which can be tailored to each particular scenario. Section 4 presents the details of STAIRS.

In this video paper we present one example scenario of how the TA and STAIRS can be used to improve the information flow in a real-world context and how this can be used to enhance the work process. Section 5 describes an overview of the video.

[1] STAIRS is an acronym for Structured Audio Information Retrieval System

2. RELATED WORK

W3C's Voice Browser Activity [7], is working to expand access to the Web to allow people to interact via key pads, spoken commands, listening to prerecorded speech, synthetic speech, and music. Their goal is to use any telephone to access appropriately designed web-based services, focusing on people with visual impairments or people needing access while keeping their hands and eyes free. One approach in this direction is HearSay [6] which is based on automatically creating audio browsable content from hypertext Web documents. It uses two technologies, partitioning of web documents and VoiceXML.

In the field of mobile service systems the most popular application that come into mind is the the Boeing aircraft inspection application [5]. Navigator 2 is a multimedia wearable with a focus on graphical output and a joystick as primary input device. the input can be combined with speech. A modern variant is the Xybernaut Mobile Assistant (MVA) [4]. Xybernaut is a industry level wearable where several input and output devices, like a headset, a scanner, a keyboard or a headmounted display can be attached. In addition it features LAN/WLAN capabilities.

A number of different location systems based on infrared tags are documented in literature. Determining location at room granularity was first implemented in the Active Badge system. Our tags are designed to have a limited range and viewing angle in order to determine a user is standing in front of the tag. Compared to Cooltown, where beacons emit static URLs, the STAIRS browser allows to dynamically associate tags with hypertext nodes.

3. TALKING ASSISTANT

The TA was built in our Mundo [3] project which aims at investigating and prototyping ubiquitous computing infrastructures. We decided to base our architecture around a single, minimal device, because such a device is easy to carry around in all situations. A Minimal Entity (ME) can augment its capabilities through association with other entities or by utilizing services hosted in the network.

The ME carries the digital identity of the user, is equipped with a wireless network interface, and audio input/output capabilities. We believe that items identifying users must have a certain intelligence and a user interface, to be able to act as a secure terminal during transactions and to enforce privacy policies.

The TA in its current form implements our requirements for such a ME device. However, because suitable manufacturing techniques are not available today, the device is not as small as we would expect it to be for use in everyday life. Our current applied research focuses on mobile work scenarios, where hands- and eyes-free interaction is desired, like shown in this video.

The device has 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.

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. The audio hardware consists of a DSP that supports direct playback of PCM and MP3-encoded audio data. MP3 streaming provides very good audio quality at reasonable network bandwidth requirements. The TA senses the user's head orientation in all three rotational axes and supports an absolute and a relative positioning system [1]. In the following, only the relative positioning system is discussed.

Tags are small infrared emitters that we can attach to both fixed and mobile devices. A tag has a unique ID which it broadcasts periodically. These broadcasts can only be received by nearby receivers within direct line of sight. Hence, reception of a tag ID conveys the context that the TA is close to and facing a tagged object.

4. STAIRS

The STAIRS project is aimed at developing a simple audio browser for use with the TA. Some application scenarios for the project are delivering information to laboratory workers, mobile inspectors in automotive and aerospace industries, and training at work applications. The worker interacts with the system using a Talking Assistant. In these scenarios the hands and eyes are typically busy performing the tasks, leaving voice as the logical interaction modality.

The information accessed by the worker is stored in an information base in the network of the organization. We use text-to-speech for the text corpus and auditory icons for structuring. It is known from the context of audio web browsers that structured audio is easier to follow because the structure of the information becomes evident to the user and also contains information on how to navigate through the information base. Workers can navigate through it and access the information they need at any given time. Voice interaction, however, requires a new type of interaction device and information access paradigms. We determined a set of 15 basic commands in a survey that was based upon ETSI standard ES202076 [2].

The information base contains information needed by the worker, such as how the repair has to be performed, what things need to be checked, etc. It also contains solutions for common problems, as shown by the following example. Consider a worker inspecting a car who sees fluid leaking from a valve. Using the information base she can get information about possible causes of the leakage and solutions to these.

The Talking Assistant can also act as a source of context, thanks to its positioning system. This allows us to restrict the information delivered to the worker to cover only information relevant to the context of the worker. For example, consider a worker in a garage who needs to perform a list of repairs on a car. The positioning system tells us at which part of the car the worker is, so we can send the worker information only about the repairs to be done for the current part. This avoids overloading the worker with information.

5. CAR REPAIR SHOP DEMO

The video shows an application of the TA and STAIRS in a car repair shop. We show how the TA can improve the information flow and facilitate the repair process, by reducing the possibilities of errors, through a tight integration between the backend system and the front-line workers. All of the features shown in the video are already implemented, except for the actual backend system which is simulated, since implementing such systems is beyond the scope of our work.

One of the key features of STAIRS shown in the video are auditory icons. These are simple sounds, played out through the TA, and give the user feedback about his actions. We have defined about 10 different auditory icons for different situations. In the video, we use auditory icons for the following three purposes: (i) user is near an infrared tag, (ii) speech command recognized, and (iii) structural markers for sections, paragraphs, and list items.

The information which is delivered to the user is passed through a text-to-speech engine. It is also possible to use pre-recorded audio messages, but we have chosen to use TTS because it allows for rapid prototyping and development. We are currently using AT&T Natural Voices [8] as our TTS engine.

In the video, a customer brings his car for a check-up. The work order is first created using the backend system, in combination with the TA. Then, the supervisor places infrared tags on the car, to mark places where work is needed. When the worker starts working on the car, he automatically gets a list of tasks to be done, and can get more information about them, as well as add notes about the repairs he has done. Completed tasks and voice notes are automatically recorded in the backend and can be used when the supervisor is preparing the bill.

6. REFERENCES

[1] E. Aitenbichler, J. Kangasharju, and M. Mühlhäuser. Talking Assistant Headset: A Smart Digital Identity for Ubiquitous Computing. In *Advances in Pervasive Computing*, 2004.

[2] ETSI. Human factors (HF); user interfaces; generic spoken command vocabulary for ict devices and services. Technical report, ETSI, Nov 2000.

[3] A. Hartl et al. Engineering Multimedia-Aware Personalized Ubiquitous Services. In *IEEE Fourth International Symposium on Multimedia Software Engineering*, pages 344–351, 2002.

[4] Mobile assistant v. Technical report, abcus.

[5] navigator 2. `http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/vuman/www/navigator.html%`.

[6] I. V. Ramakrishnan, A. Stent, and G. Yang. Hearsay: enabling audio browsing on hypertext content. In *Proceedings of 13th WWW Conference*, 2004.

[7] Voice browser activity. `http://www.w3.org/Voice/Activity.html`.

[8] AT&T Natural Voices. website. http://www.naturalvoices.com.