Enhancing the Flexibility of a Multimodal Smart-Home Environment

Thimios Dimopulos[1], Sahin Albayrak[2], Klaus Engelbrecht[1],
Grzegorz Lehmann[2], Sebastian Möller[1]

[1] Deutsche Telekom Laboratories; [2] DAI-Labor, TU Berlin
Ernst-Reuter-Platz 7
D-10587 Berlin
{thimios.dimopulos|sebastian.moeller|klaus-peter.engelbrecht}@telekom.de
{grzegorz.lehmann|sahin.albayrak}@dai-labor.de

Abstract: The INSPIRE smart-home environment is an open platform for multimodal home appliance control. We describe the latest enhancements of the platform in order to increase its modularity and flexibility. These include connecting the system to a VoiceXML-based voice platform, and an abstract home device controller that interfaces the system to Universal Plug and Play devices.

1 Introduction

The smart home offers a new opportunity to augment people's lives with ubiquitous computing technology that provides increased communications, awareness, and functionality. Moreover, the technological complexity of electronic appliances in the home context is often causing problems to users [Th05]. The need for interfaces that resemble the intuitive human-to-human interaction becomes apparent.

Natural language dialogue is easy to learn, hands free, and accessible from any place in the household or outdoors. However, strictly spoken interaction is less effective than the multimodal face-to-face interaction that humans are used to. During multimodal communication we use speech, gestures, eye gaze and body posture in a powerfully expressive coordination [OC00]. Nevertheless, some features of a device may be convenient to use with traditional controls (remote control, nods etc.) which can also be considered as additional modalities. We therefore realize the need for intelligent interfaces, e. g. in the form of a "home assistant", which should hide the complexity of the appliances through an intuitive interface.

Such a home assistant was developed in the EU-funded IST project INSPIRE [Mö04]. During the INSPIRE project, system development was focused mainly on natural language control of domestic devices, e.g. TV, video recorder, answering machine, lamps, blinds etc. It has since then evolved into a framework for multimodal application design, by the work of Cenek, Melichar and Rajman [CMR05].

In a collaboration between the SRL Usability of Deutsche Telekom Laboratories and the DAI-Labor, Berlin University of Technology, we have augmented the modularity and flexibility of the system, by exploiting recent advances in consumer electronics, wireless communication and user interface technologies. We decided to increase the flexibility at two points: 1. *The user interface*: By coupling it to a powerful VoiceXML platform which provides remote or local voice interaction in a unified, standards-based fashion. 2. *The device and environment connection*: By leveraging the system's modularity we have extended it with an abstract, protocol-independent device controlling mechanism. In Section 2, an overview of the current system architecture is presented. The main parts of the system along with our recent enhancements are discussed in Section 3. In Section 4, conclusions and plans for future work are sketched.

2 Overview of the System Architecture

The current implementation of the INSPIRE system is based on a distributed component architecture. This enables a flexible integration and the reuse of heterogeneous software components. Such an architecture is considered particularly suitable for multimodal dialogue systems, due to their high complexity [HK03], [Mc02].

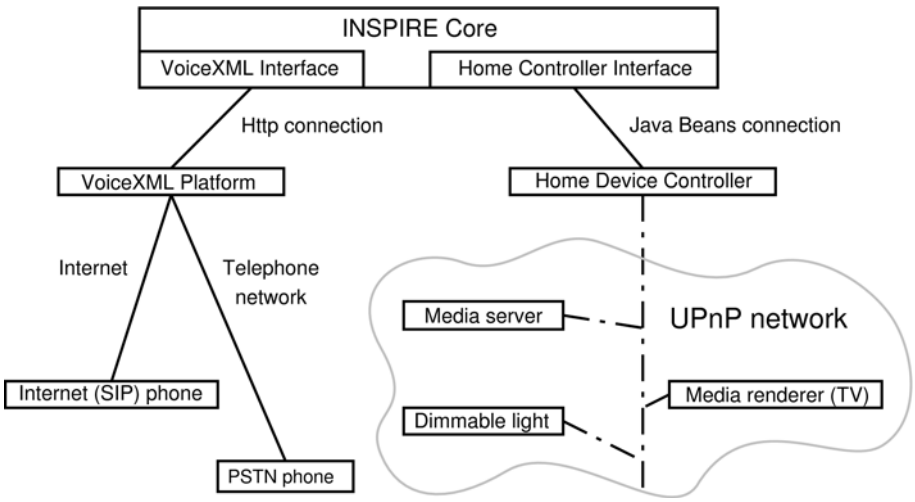

Figure 1: System architecture overview

An overview of the current system setup is presented in Figure 1. The INPIRE core is the main part of the system where the dialogue logic resides. Its components bridge the user multimodal interaction and the appliance actions with the dialogue logic. The Voice Platform is where the speech recognition (ASR) and the text-to-speech (TTS) systems reside. It enables calling the INSPIRE system either through a normal telephone, or an Internet phone. The Home Device Controller provides an abstraction of the home appliances. It keeps track of the appliances that get connected to the system, discovers their capabilities and hides the complexity of their operation.

3 Recent System Enhancements

The system architecture is the result of a recent effort to enhance the flexibility of the INSPIRE platform. In the following sections, we describe these efforts in more detail, addressing the INSPIRE core, the Voice Platform and the Home Device Controller.

3.1 Dialogue and Interaction Manager – INPIRE Core

The INSPIRE core – presented in Figure 1 as a single entity – is not a monolithic component. Its implementation follows the Enterprise Java Beans (EJB) component architecture paradigm, which renders it a distributed system [EJB06]. The central component that controls the interaction flow with the user is the dialogue (or interaction) manager. Its input is a formal representation of the user's expression, e.g. a spoken utterance, a gesture, or a combination of them. The dialogue manager output is a formal message from which an output component can generate a response to the user. In addition, it communicates action-taking commands to the action-performing components. A detailed description of the dialogue manager has been published by Bui et al. [BRM04].

Further core components bridge the dialogue manager with the human interaction modalities, and with the devices of the smart home. Input components are concerned with gathering the input from the possible modalities, semantically interpreting it and feeding it to the dialogue manager. Output components drive the output of the dialogue manager towards the user, and action components realize the actions that the dialogue manager decides to take. Two such components have been added recently:

The newly implemented VoiceXML interface is both an output and an input component. For each turn of the dialogue, it produces a VoiceXML document, which describes the next spoken interaction with the user. This description is then forwarded to the Voice Platform which handles the actual user interaction. Once the user response has been recognized, the Voice Platform forwards it to the VoiceXML interface component.

The second INSPIRE core enhancement is the Home Controller Interface. It is an action component that maps the actions emitted by the dialogue manager to the abstract device description provided by the Home Device Controller, see Section 3.3. Special effort is put in the flexibility of this mapping. For example, the dialogue manager might decide to display a video recording to the user. The home controller interface will instruct the controller to find the video-rendering-capable screen which is closest to the user's current position in the smart home, and render the video on that one.

3.2 User Interface Enhancement – Voice Platform

During the INSPIRE project, the system was thoroughly tested with locally-connected speech recognizers and TTS systems [Mö04]. With the advent of sophisticated voice platforms based on the VoiceXML specification [MB04], developing a VoiceXML interface for the INPIRE system emerged as an appealing advancement. VoiceXML is currently a widely accepted standard for building speech interfaces. By enhancing the core

of the INSPIRE system with a VoiceXML interface component, we gained a vendor-independent connection to any voice platform.

In the current setup, a commercial Voice Platform is used. It incorporates a speech recognizer which we have trained using a natural language modeling approach based on bi-grams. For speech output, one option is to use the TTS system of the Voice Platform. However, despite their flexibility, current TTS systems lack some naturalness. To cope with that, the INSPIRE core includes a component that generates prompts by concatenating prerecorded speech. The prompts are then served to the Voice Platform which can be accessed either by an IP-phone through the internet, or by mobile and land line phones.

Despite its strong points, VoiceXML was mainly designed for call center applications. In our system, it is used in the simplest way, namely as a straight interface to the Voice Platform. The purpose of this is to keep all crucial functionality like the dialogue logic or the semantic interpretation of user utterances in the Java components that have already been developed. This way, the modularity and flexibility of the system is preserved, while a new way to interact with it is incorporated.

3.3 Environment Enhancement – Home Device Controller

In order to evaluate the INSPIRE system in a real-world smart-home environment, we have integrated it with the Home Device Controller (HDC) architecture. The HDC architecture has been developed during the works on the “Seamless Home Services” project at DAI-Labor, and has been designed to serve as the back-end for smart-home systems [FB06]. It provides the means to easily access and use devices and services in the home environment. The HDC defines its own abstract, ontology-based description of devices, which is independent of any standard or vendor. This approach is necessary to cope with the heterogeneity of today’s home appliances. Moreover, it also makes the HDC architecture easy to use for developers. All device types are encapsulated and their specific control protocols become transparent, since the HDC handles them.

Internally, each device discovered in the environment by the HDC is managed by a controller, which is responsible for the standard-specific communication with it. If a new standard needs to be supported, a new controller must be implemented. This can be done in an easy way, because a lot of effort has been put to assure good extensibility of the HDC. Often this is not necessary though, because multiple controllers have already been created during the works on the HDC. The development focused on the UPnP technology, which is currently the most established standard for device discovery and control in the home environment.

By using the HDC architecture, the integration of the INSPIRE system within the home environment has been greatly simplified. As already described, the INSPIRE core has been enhanced with an action component that maps the actions emitted by the dialogue manager to the abstract device description provided by the HDC. This way our system can fully benefit from the flexibility of the HDC and its already implemented controllers. Furthermore, new device types can be integrated using the mechanisms provided by the HDC without any changes in the rest of the system.

4 Conclusions and Open Problems

We have leveraged the flexibility and extensibility of the INSPIRE system, by extending the core with a VoiceXML interface component that enables coupling it to any VoiceXML enabled platform. We are currently using a commercial Voice Platform so that the system can be accessed with any kind of speech telecommunication device. We have extended the smart-home environment with a controller framework that enables control of home appliances in a unified way, regardless of protocol or vendor. We already use UPnP appliances including a media server and various media renderers, along with non-UPnP based blinds and lights.

An open problem is the auto-generation of the currently manually-produced task and dialogue models. Ideally, connecting a new appliance in the smart home should automatically extend the dialogue to include the interaction with the new appliance. Our interests also go beyond the system itself. We would like to establish a framework where the overall quality of multimodal systems could be formally evaluated. In the MeMo project [ME06], we aim in evaluating and predicting the usability of spoken dialogue systems in an automated way.

References

[BRM04] Bui, T.; Rajman, M.; Melichar, M.: Rapid Dialogue Prototyping Methodology. Lecture Notes in Computer Science, Volume 3206, Jan 2004.

[CMR05] Cenek, P.; Melichar, M.; Rajman, M.: A Framework for Rapid Multimodal Application Design. Lecture Notes in Computer Science, Volume 3658, Aug 2005.

[EJB06] Sun Microsystems Corp.: The Enterprise Java Beans Homepage. http://java.sun.com/products/ejb/, accessed June 2006.

[FB06] Feuerstack, S.; Blumendorf, M.; Lehmann, G.; Albayrak, S.: Seamless Home Services. AmI.d Ambient Intelligence Developments Conference, September 20-22, 2006.

[HK03] Herzog, G.; Kirchmann, H.; Merten, S.; Ndiaye, A.; Poller, P.: MULTIPLATFORM Testbed: An Integration Platform for Multimodal Dialog Systems. In: Proc. HLT-NAACL, 2003.

[MB04] McGlashan, S.; Burnett, D.; Carter, J.; Danielsen, P.; Ferrans, J.; Hunt, A.; Lucas, B.; Porter, B.; Rehor, K.; Tryphonas, S.: VoiceXML Version 2.0. 2004.

[Mc02] McTear M.: Spoken Dialogue Technology: Enabling the Conversational User Interface. ACM Computing Surveys, 2002.

[ME06] Möller, S.; Engelbrecht, K.; Englert, R.; Hafner, V.; Jameson, A.; Oulasvirta, A; Raake, A.; Reithinger, N.: MeMo: Towards Automatic Usability Evaluation of Spoken Dialogue Services by User Error Simulations. Accepted for: Interspeech 2006 – ICSLP, Pittsburgh, 2006.

[Mö04] Möller, S.; Krebber, J.; Raake, A.; Smeele, P.; Rajman, M.; Melichar, M.; Pallotta, V.; Tsakou, G.; Kladis, B.; Vovos, A.; Hoonhout, J.; Schuchardt, D.; Fakotakis, N.; Ganchev, T.; Potamitis, I.: INSPIRE: Evaluation of a Smart-Home System for Infotainment Management and Device Control. Proc. LREC 2004, Lisbon, 1603-1606.

[OC00] Oviatt, S.; Cohen, P.: Perceptual User Interfaces: Multimodal Interfaces that Process What Comes Naturally. Commun. ACM 43, 3 (Mar. 2000), 45-53.

[Th05] Thompson, C.W.: Smart Devices and Soft Controllers. Internet Computing IEEE Volume 9, Issue 1, Jan-Feb 2005.