Agent-oriented, multimedia, interactive services in home automation

Luigi Ceccaroni and Xavier Verdaguer

:tmtfactory>
Edificio Forum de la Tecnología, c. Marie Curie s/n, 08042 Barcelona
luigi@tmtfactory.com, xavier@tmtfactory.com

Abstract. Several academic and industrial research groups are working to deploy home-automation systems and to improve their user interfaces. Based on user studies and prototype implementation, this paper presents the development of an innovating appliance that incorporates interactive services of leisure and information, offered through a high-quality user interface on the surface of a mirror. This appliance is thought to be both integrated in an agent-based network of a home and used as an independent element. The prototype will offer basic services, like interactive television, the presentation of personalized weather data, news and selections of musical tunes, and will incorporate some advanced functionalities, like an intelligent service of personal motivation and a reminder service.

1 Introduction

Modern man is basically sleepwalking if he is not multitasking. His lifestyle has evolved in such a way that optimizing time is the most important thing. The newest innovations show that we can be online while watching TV, be email-ready while driving the car, be taking calls while ascending Mount Everest, be watching television while shaving in front of the mirror or access email using a phone while traveling. Most evenings most first world's kids do their homework on the computer while instant messaging friends and talking on the phone. The stereo is typically playing and they are probably downloading MP3s and DivX movies. Occasionally their foot extends to gently connect with the dog. There may be something wrong with this picture, but the kids are definitely not sleepwalking. The pressure to be wired - everywhere, always, fast- is hard to ignore [8].

Software developments are more and more enabling systems to *think for themselves*. Artificial intelligence is now reaching the stage where systems can be programmed to predict situations and outcomes based on previous experience, and then take action accordingly. By using *fuzzy logic*, systems can continue to operate efficiently where accurate data is either not available or not strictly necessary, in much the same *rough and ready* way we ourselves often work. These are software programs which *know* what the user wants to do and can act autonomously on his or her behalf. The user will put a problem to the agent, and the agent will then monitor events and perform tasks which meet the user's goal. It will search for information the user needs regularly. It will monitor changes, and even represent the user where necessary. It will learn from experience, automating regular patterns. At any one time we may have a number of these agents working for us: they will be able to expand their knowledge-base by learning about us from each other.

Before the increasing interest of the industry and the market for home automation, the concept of intelligent human-habitat managed by machines has left the science-fiction domain to become, in the last years, a necessary investment-area for the companies that try to lead the sector in the mid term. Intelligent appliances are starting to appear on the market, appliances we can communicate with on both a functional and emotional level, appliances which use voice recognition and synthesis, software agents, fuzzy logic and virtual reality. But, is home automation really a necessity for consumers? It's too easy, of course, to be a skeptic, to regard each new invention with a sense of wise detachment and, while praising its state-of-the-art capabilities, wonder aloud if we really need it. After all, nobody wants to risk missing real progress.

2 Related work

Several projects and public initiatives worldwide are dealing with ambient intelligence (e.g., [1] [6] [10] [12] [13] [15] [16]) and a few of them are developing interactive mirror-displays on which, based on personal preferences, different applications can be run. In particular, the HomeLab [17] at Philips Research [18] is working on a project which is much related to the subject of this paper.

In the summer of 2004, a promising residential observational research facility in Cambridge Massachusetts opened. The PlaceLab has been explicitly designed, in the frame of House_n and Changing Places (see below), to be a highly flexible and multi-disciplinary facility for the scientific study of people and their interaction patterns with new technologies and home environments. House_n is a Department of Architecture research consortium at the Massachusetts Institute of Technology that explores how new technologies, materials and strategies for design can make possible dynamic, evolving places that respond to the complexities of life. Changing Places is a joint Architecture and Media Laboratory consortium that includes House_n and emphasizes links between the home and places of healing, work, learning, and community.

Several other groups dealt with the development of dialogue systems in diverse domains, not only home automation, with projects such as: SmartKom [19] on the design of new interfaces, which aims to help reducing the hesitations people presently feel upon using information technology and thus to make a contribution to the user friendliness and user centeredness of technology in today's information society; TRINDI [20] on task oriented instructional dialogue, which built a computational model of dialogue moves; Jupiter [21], a conversational system that provides up-to-date weather information over the phone, and other projects at MIT's Spoken Language Systems; the Conversational Interfaces project [10], whose aim is to build a general purpose dialogue system which supports multi-modal (i.e., speech and gesture) activity-oriented dialogues with devices, applications and services, and other projects at Stanford University's Center for the Study of Language and Information; and August [6], a Swedish spoken dialogue system.

3 Overview

Often, really useful technologies are not considered a need in a first phase (e.g., telephones, computers, PDAs) and part of the challenge in this project, called Magical Mirror, is demonstrating a technology to help creating new needs. In the software industry, much emphasis is placed on computational modules embedded in appliances, hand-held devices and information systems, which often have to interchange significant amounts of multimedia content.

To be sure, current technologies that transmit images and voice fall far short of anything remotely realistic [8]: at present, home automation generally involves jerky video and canned sound. In Magical Mirror it is necessary to deal with these limitations in order to develop an interactive appliance able to provide the following realistic services, controlled by voice, in Spanish (English and Catalan, in the future), and presented with the support of a hand-designed, high-quality, advanced visual interface (see Fig. 1):

- S1. customized weather data
- S2. customized information on traffic and public transport
- S3. customized news (stock market, sports)
- S4. customized music management
- S5. interactive television (integration with the IntegraTV service[1])
- S6. Internet access and search
- S7. personal agenda
- S8. home automation (control of other devices: lights, heating, oven)
- S9. communication among users (voice-message recording, email)

4 Architecture

Magical Mirror is a meta-service, with the layered structure shown in Fig. 2. It is important that this service's environment appears, before the user, as one interface. It would not be acceptable that he has to use an interface in order to handle home devices (lights, heating, oven), another one for communications devices (telephone, email) and a different one for audio and video devices (TV, music). In order to obtain this unity in the interface, a high level of interaction, among the heterogeneous infrastructures of communications that will support it, is needed. Some of the layers of the architecture are described in the following sections

[1] See [http://integra.tv/].

Fig. 1. Magical Mirror's high-quality interface (BCNMedia and **:tmtfactory>** are trademarks of the same company.)

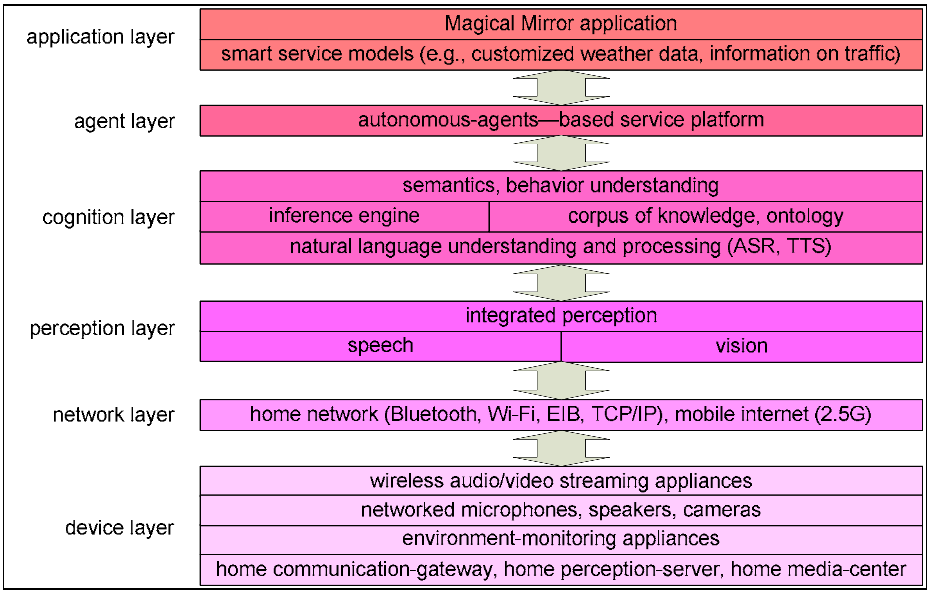

Fig. 2. Layered structure of the Magical Mirror service

4.1 Perception and cognition layers

Voice and image feeds will be collected in the perception layer and processed in the cognition layer. The main challenges during perception are:

- automatic speaker-identification;
- voice-silence detection;
- robustness with respect to the environmental noise (basically, water noise, echo and other voices).

Research and commercial tools are used for automatic speech recognition (ASR) and text to speech (TTS) processing. VoiceXML (VXML)[2] is used in service models to represent grammars, speech acts and dialogues. Here is a simple example of VXML encoding and a sample human-computer interaction (translated from Spanish to English) in the case of customized music management:

```
<form>
  <field name="artist">
     <prompt>Whom of these artists World you like to lis-
              ten to?: Anticappella? Aphex Twin? Eiffel
              65? Elio e le Storie Tese? o Leftfield?
     </prompt>
     <grammar        src="artist.grxml"
                     type="application/srgs+xml"/>
  </field>
  <block>
     <submit next="http://www.tmtfactory.com/art2.asp"/>
  </block>
</form>
```

A field is an *input* field. The user must provide a value for the field before proceeding to the next element in the form.

C (computer): Whom of these artists World you like to listen to?: Anticappella? Aphex Twin? Eiffel 65? Elio e le Storie Tese? or Leftfield?

H (human): Mogwai.

C: I didn’t understand what you said. (default message.)

C: Whom of these artists World you like to listen to?: Anticappella? Aphex Twin? Eiffel 65? Elio e le Storie Tese? o Leftfield?

H: Aphex Twin

C: (continues in document art2.asp)

4.2 Agent layer

The highest layers of the architecture will be based on independent agents which interact among them (on multiple devices) and with the user, and which accomplish the intelligent control of diverse devices, sensors as well as actuators (in phase of study for *modality **c***). Agent-based architectures present, potentially, a series of fundamental advantages, such as modularity, reusability, scalability and dynamic reconfiguration. Agents will be physically deployed on two JADE platforms: one in the computer of the hall and the other one in the Magical Mirror. Each agent has a basic functionality and in addition is equipped with certain *decision and adaptation* capabilities, in such

[2] See [http://www.voicexml.org].

a way that it can improve its efficiency on the basis of information that receives directly from other agents or indirectly through three kinds of support knowledge-components:

1. a *base of facts and knowledge* that is common to all the agents;
2. a **domain ontology** (*DomOnt*);
3. a number of **service ontologies** (*ServOnt*).

An ontology is a model of the world in which the different agents of the system work. The ontology is made up of a hierarchy of entities, which represent buildings, rooms, elements that compose the rooms, users, tasks that these users are accomplishing, and any other information of interest for the community of agents which belong to the home-automation environment. Agents use the *DomOnt* as a mechanism for obtaining data they need about the environment and, in turn, deposit in the ontology their results so that other agents can take advantage of them.

While the *DomOnt* contains a representation of the environment and (in its instances) of its state, the *ServOnt* models the information flow that is exchanged among the different agents part of the environment, influencing the language and the communication protocols between them. The agents channel the information according to the flow model contained in the *ServOnt*. For example, if an agent deduces that the user requires that a certain flow of video (the output of a camera) is transferred from the hall to the mirror of the bath, what it has to do is to change some parameters of the model of information flow contained in the *ServOnt*. The agents controlling the camera in the hall and the computer in the bath respond to this change in the ontology one emitting the video flow towards the computer in the bath and the other beginning to emit this flow in the Magical Mirror.

An important element of the Magical Mirror project is to allow the addition to the home-automation environment of any sort of new agents and devices developed by third parties using the same standards. With this aim, OWL [5] and OWL-S [12] (both based on XML) are used as the languages of definition of the *DomOnt* and the *ServOnt*, respectively. This assures that the requirements that a new agent will have to fulfill to join the home-automation environment will be minimal and standardized. With respect to the agent systems' architecture, the general requirements are:

1. compliance with the FIPA standard of communication and interaction;
2. ability to use OWL ontologies.

4.2.1 Complexities of application-development in multiple-devices environments

Writing agent-systems for multiple devices is difficult. Devices vary greatly in their input and output characteristics. A PDA or a mirror could have a rich color display, a PC would likely have a large display, while a camera could have no display at all. It may be more appropriate to use a stylus for text input on a PDA and voice input on a phone. A user always carries a phone or a PDA, making them prime candidates as environments for agents supporting time-sensitive interaction. For example, an agent that monitors TV programming for a user could send him a message reminding him to watch a particular program. Most approaches to device heterogeneity have focused on the presentation or user interface aspects only. But in many cases the data or the model that are being used for a device are different. Many approaches have suggested using the model-view-controller (MVC) architecture (see Fig. 3) to address this issue.

MVC is a commonly used and powerful architecture for GUIs. The MVC paradigm is a way of breaking an application, or even just a piece of an application's interface, into three parts: the model, the view and the controller. MVC was originally developed to map the traditional input (C), processing (M) and output (V) roles into the GUI realm [2].

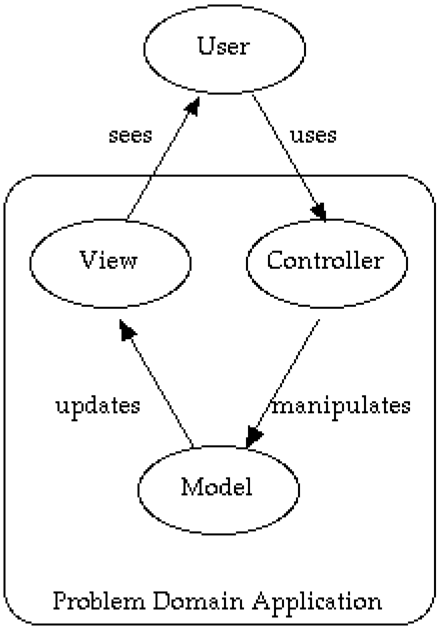

Fig. 3. Model-view-controller (MVC) architecture

In general this framework is valid as long as the model does not change. But what if it changes, as in the case of multiple devices? New approaches suggest this framework is not appropriate for multi-device agent systems [9]. Designing the next versions of the Magical Mirror agent-system will require device- and connectivity-awareness, and a key element in the solution of this problem will be a new class of multidevice-computing--aware design-patterns[3] and perhaps, over time, a pattern language for multi-device computing.

[3] Design patterns capture the static and dynamic structures of solutions that occur repeatedly when producing applications in a particular context. Design patterns are an important technique for improving the quality of software since they address fundamental challenges in software system development. Key challenges addressed by design patterns include communication of architectural knowledge among developers, accommodating a new design paradigm or architectural style, and avoiding development traps and pitfalls that are usually learned only by experience [4].

5 Personalization

In Magical Mirror, due to the presence of a number of different services, personalization is a long and challenging process to be carried out. The personalization methodology that we use is defined by the following steps:

1. Creation of domain models with formal semantics, via ontologies:
   a. user domain;
   b. television-programs domain;
   c. …
2. Creation of a recommendation algorithm, which associates instances of the user-domain ontology with instances of the service-domain ontologies (e.g., in the case of television, in the form of personal programming guides). This step includes a certain degree of learning and constitutes the core of the personalization process.
3. Creation of an exploration algorithm (a search engine) for the different services, which takes into account the user profile.

Domains are formalized using ontologies: the static part of each domain is represented by a subset of *DomOnt*, while the dynamic part of service domains is represented by a *ServOnt*. User profiles, i.e. instances of the user ontology, are created through a Component for Creation of User Models (CCUM), which reuses a module developed for the TVFinder system [3]. In Fig. 4, an example of the interface for the capture of user preferences is shown. In this particular case the user is asked to explicitly express his preference with respect to the *ingredients* (such as *love*, *action* and *suspense*) of his favorite TV series. Explicit preferences are just one kind of preferences used by the CCUM; other kinds include stereotypical and implicit preferences.

A code is provided to each user in order to access his personalized interface with the Magical Mirror (e.g., music libraries and TV channels). The identification of the user will be important also for improving the voice, image and gesture recognition.

6 Modalities

Magical Mirror can work in three modalities:
[a] independent;
[b] connected to the Internet (in particular, to an online data-base);
[c] connected to the Internet and to a home-automation network.

The data-base includes audio-visual contents and preferences of the users. After a process of initialization and content upload, the way to update the system depends on the modality of the mirror:
[a] learning;
[b] updates from a content portal + learning;
[c] updates from a content portal and the home-automation network + learning.

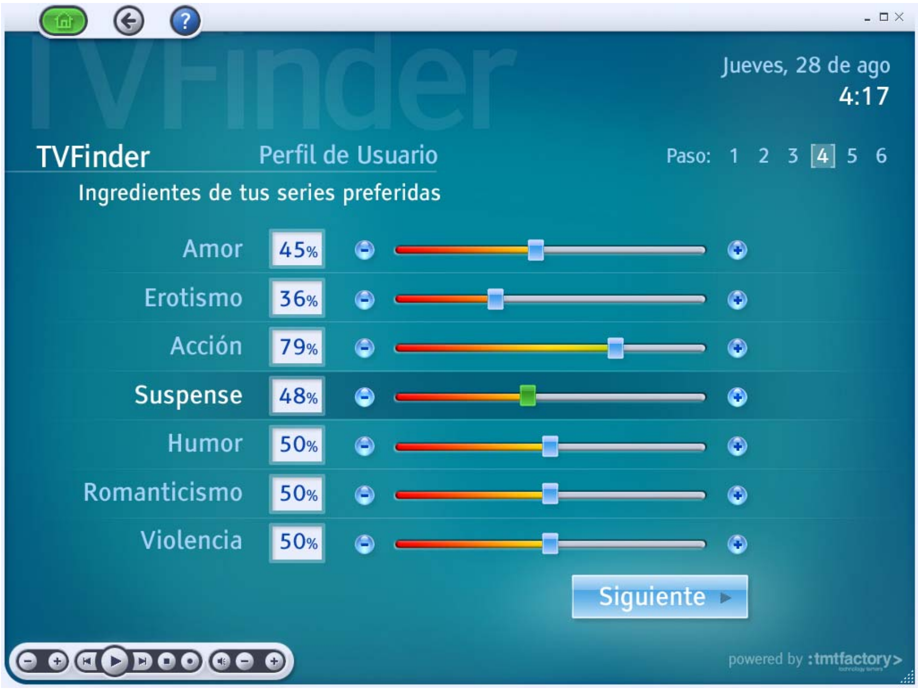

Fig. 4. Interface representing a step in the creation of the user profile (Copyright © 2003, :tmtfactory>)

7 Natural-language dialogue management

Among the necessary technologies to be integrated in the Magical Mirror, so that it is possible to provide the type of personalized services described above, we are building a preliminary speech interface. To this aim, the main artificial-intelligence technique used is an agent for dialogue management, i.e. an agent for natural-language conversation, aware of the profile and preferences of the user.

One of the objectives of the project is that the developed system is able to maintain robust dialogues in natural language. By *robust dialogue* we mean one in which the user of the system can use general expressions (i.e., not restricted to a predetermined language of commands) to access the functions and services of the mirror. Naturally, this objective is excessively ambitious if we pretend that the scope of these dialogues is an ample domain without restrictions. Nevertheless, if the domain is sufficiently restricted, the complexity of the problem is tractable with the present technology of natural language processing. Therefore, a set of dialogue domains has been identified that corresponds with the services offered and is sufficiently restricted to allow such processing.

The architecture of the system is organized around the *VoiceXML framework* and the documents that comprise the *dialogue manager*. These documents are generated with dynamic content coming from the *data bases* of the system: one for the data of the services and one for the user profile. The choice of VoiceXML turns the dialogue manager independent of both the signal-processing technology and the logic used in the application.

The value of the technology being developed does not primarily consist of an advance in each one of the specific techniques of interaction (intelligent processing of the language, agents), but rather in their correct integration to form a coherent whole. Therefore, the process of design and development of a correct architecture of integration of all the modules constitutes an essential part of the project.

8 Use case

Fig. 5. Magical Mirror at a public exposition

A demonstration version of Magical Mirror was shown at various expositions around Catalunya (Spain), in 2003, to check the reaction of the public (see Fig. 5), and received a positive feedback. We hand-designed graphical interfaces on Magical Mirror for interactive television, a multimedia player, weather data and news services. Since we could not yet control the actual applications by voice commands, we simulated control with a laptop (connected to the interface of the mirror) that responded to commands and signals from the users by playing multimedia content that closely matched the response expected of the actual applications. Most people attending the expositions, including kids and seniors, interacted with the mirror speaking in complete comfort.

The average time spent interacting with it was about 5 minutes per person; longer in the case of kids. The attribute that attracted the greatest deal of attention was not the function as information-services provider, but rather the capability of differentiating among interlocutors. In other words, the high degree of personalization and humanity was the most appreciated feature identified by the public. Visitors could listen to their favorite songs and see images corresponding to their specific requests. The importance of the use of stereotypes was absolutely clear at the moment of the mirror's deployment. About 80% of the visitors requested the same artists and type of images; and these artists corresponded to the ones at the top of the charts of the season. Sport, and specifically soccer, was often the object of search requests.

In general, the public showed a positive disposition at adopting a similar mirror at home, given that the price were not much higher then the one for a traditional appliance, demonstrating that people did not consider the apparatus as something completely new and different, and that the additional technology was installed in a satisfactorily transparent way. Additional user studies are in progress to confirm this.

9 Conclusions

A prototype of a device that incorporates interactive services of leisure and information, offered to the user through a natural interface in form of language and the visual superposition of images on the surface of a mirror, is being developed. This appliance is thought to be both integrated in an agent-based network of a home and used as an independent element. The main services, controlled by voice, in Spanish, and presented with the support of an advanced visual interface are: customized weather data, customized news, customized music management, interactive television, Internet access and search. All services are in various phases of research and prototypical development. Nonetheless, a comprehensive prototype, called Magical Mirror, was deployed at various public expositions, where people could interact lively with it.

Acknowledgements

This work was conducted as a part of the Magical Mirror project. The interactive television interface is being developed as part of the TVFinder [3] (phase 2) and IntegraTV-4all projects. Various people at **:tmtfactory>** did a portion of the work on the prototype interface. This work was funded in part by grants FIT-070000-2003-626 and FIT-330300-2004-52 from the Spanish Ministry for Science and Technology, through the PROFIT program. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the Spanish Ministry for science and technology, and the Spanish Ministry of industry, tourism and commerce.

References

[1] Aarts, E. H. L., Roovers, R.: IC Design Challenges for Ambient Intelligence. Proceedings of Design Automation & Test in Europe (DATE 2003), Munich, Germany (2003)

[2] Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P.: Pattern Oriented Software Architecture: A System of Patterns. 1st ed. John Wiley & Sons ISBN: 0471958697 (1996)

[3] Ceccaroni, L., Verdaguer, X.: TV finder: una aproximación semántica a la televisión interactiva. Proceedings of the workshop on Ubiquitous computation and ambient intelligence of the Conference of the Spanish Association for Artificial Intelligence (CAEPIA 2003), Donostia, Spain (2003)

[4] Coplien, J.O., Schmidt, D.C. (eds.): Pattern Languages of Program Design. Addison-Wesley ISBN: 0-201-60734-4 (1995)

[5] Dean, M., Schreiber, G., Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L.A.: OWL Web Ontology Language Reference. W3C Proposed Recommendation (15 December 2003)

[6] Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J., Burgelman, J-C.: Scenarios for Ambient Intelligence in 2010. ISTAG final report for the European commission (2001) Available at [http://www.cordis.lu/ist/istag-reports.htm] (921kb), accessed 23 Apr 2004

[7] Gustafson, J., Lindberg, N., Lundeberg, M.: The August spoken dialogue system. Proceedings of Eurospeech'99, Budapest, Hungary (1999)

[8] Ignatius, A.: We have contact. Time, June 4 (2001)

[9] Islam, N., Fayad, M.: Towards Ubiquitous Acceptance of Ubiquitous Computing. Communications of the ACM, 46(2) (2003) 89-92

[10] IST Advisory Group: Ambient Intelligence: from vision to reality. ISTAG draft consolidated report for the European commission (2003) Available at [http://www.cordis.lu/ist/istag-reports.htm] (225kb), accessed 23 Apr 2004

[11] Lemon, O., Gruenstein, A., Battle, A., Peters, S.: Multi-tasking and Collaborative Activities in Dialogue Systems. Proceedings of 3rd SIGdial Workshop on Discourse and Dialogue, Philadelphia (2002) 113-124

[12] Lindwer, M., Marculescu, D., Basten, T., Zimmermann, R., Marculescu, R., Jung, S., Cantatore E.: Ambient Intelligence Visions and Achievements: Linking Abstract Ideas to Real-World Concepts. Proceedings of Design Automation & Test in Europe (DATE 2003), Munich, Germany (2003)

[13] Marzano, S., Aarts, E.: The New Everyday: View on Ambient Intelligence. Uitgeverij 010 Publishers. ISBN: 9064505020 (2003)

[14] McIlraith, S., Martin, D. Bringing Semantics to Web Services. IEEE Intelligent Systems, 18(1) (2003) 90-93 OWL-S’s latest version available at [http://www.daml.org/services/owl-s/], accessed 23 Apr 2004

[15] Myers, B. A.: Graphical User Interface Programming. In: Tucker, A. B. (ed.): Computer Science Handbook. 2nd ed. CRC Press, Inc., Boca Raton, FL, USA. To appear on 25 Jun 2004, ISBN: 158488360X (2004)

[16] Nichols, J., Myers, B., Harris, T. K., Rosenfeld, R., Shriver, S., Higgins, M., Hughes, J.: Requirements for Automatically Generating Multi-Modal Interfaces for Complex Appliances. Proceedings of the Fourth IEEE Inter-national Conference on Multimodal Interfaces, Pittsburgh, PA, USA (2002) 377-382

[17] Philips HomeLab [http://www.research.philips.com/InformationCenter/Global/FArticleDetail.asp?lArticleId=2753&lNodeId=931&channel=931&channelId=N931A2753], accessed 23 Apr 2004, is part of Philips Research Laboratories

[18] Philips Research [http://www.research.philips.com], accessed 23 Apr 2004. Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

[19] SmartKom project: Dialog-based Human-Technology Interaction by Coordinated Analysis and Generation of Multiple Modalities [http://www.smartkom.org], accessed 23 Apr 2004

[20] TRINDI project: Task Oriented Instructional Dialogue [http://www.ling.gu.se/projekt/trindi/], accessed 23 Apr 2004

[21] Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T.J., Hetherington, L.: Jupiter: A Telephone-Based Conversational Interface for Weather Information. IEEE Trans. on Speech and Audio Processing, 8 (1) (2000) 85-96