

# A Multimodal Shopping Assistant for Home E-Commerce

Mehrdad Jalali-Sohi  
and  
Feza Baskaya  
Fraunhofer Institute for Computer  
Graphics  
Rundeturmstr. 6, 64283 Darmstadt  
jalalijfbaskaya@igd.fhg.de

## Abstract

Electronic Commerce has rapidly grown with the expansion of the Internet. E-commerce has also become a promising field for applying agent and Artificial Intelligence technologies. Software agents help to automate a variety of tasks including those involved in buying and selling products over the Internet. In this paper, we describe a multimodal intelligent Shopping Assistant developed in the EMBASSI project [1]. EMBASSI is a project involving more than twenty big German companies and sponsored by BMBF [3]. Its goal is not to focus on the unlimited possibilities of this technology, but rather on the individual prerequisites of the human in contact with it. Therefore, the user interfaces of a big class of appliances and systems, including shopping and e-commerce, need to be easily and efficiently accessible for everyone, taking into account psychological and ergonomic aspects and using innovative interaction techniques by realization of intelligent anthropomorphous assistants.

## Keywords

Intelligent agents, assistant systems, e-commerce, multimodality, home shopping, mobile agents

## 1. INTRODUCTION

An important agent metaphor that can be applied to computer-aided software engineering is the personal assistant metaphor. This metaphor is discussed by Kirste [4] and Maes [5] among others. The personal assistant is like a personal secretary and hides the complexity of difficult tasks by helping the user through the problem. These personal assistants are usually used in applications such as email, meeting scheduling, news filtering and book recommendations. In this paper, we describe a multimodal personal assistant that has been developed for home e-commerce.

As e-commerce continues to expand, the potential utility of natural language communication with e-commerce applications becomes increasingly apparent. While clickable web pages will continue to support the great majority of e-commerce interactions, an increasing number of on-line customers want to interact with a person. A natural language interface, for example, could successfully reduce some of this burden on customer service. Representations for processing human communication have been mainly concerned with single modalities. Further advances, however, may require acknowledgement of the fact that a great part of human communication takes place in more than one

modality at the same time. In the project EMBASSI, we have developed an intelligent Shopping Assistant, which uses the possibilities of the EMBASSI infrastructure and provides a comfortable human-based interface to the online world. Our Shopping Assistant approach covers four research areas: Multi-Agent Systems, Human-Computer Interaction, Electronic Commerce and Security.

## 2. EMBASSI Global Architecture

In an average household today, there are at least a dozen stationary electrical appliances for the kitchen and the laundry room, as well as for entertainment and telecommunication. In addition to these systems, there are those of heating and cooling, ventilation, lights, house communication and alarm systems, as well as the multitude in the home office. Normally, these appliances and systems are based on different control philosophies. Modern appliances offer numerous functions to be controlled on the basis of existing technological possibilities. We have to equip all appliances with more intelligence in order to simplify the service and to individualize their use. In EMBASSI, we address the following concepts:

- Human-based adaptive and user interface design;
- Use of innovative, multimodal and anthropomorphous interaction technology;
- Logical separation of appliance and operating unit;
- Creation of a comfortable control mechanism through the network;
- Compatibility of all EMBASSI capable systems and appliances.

Corresponding to the conventional methods, the appliance uses an EMBASSI conform user interface with corresponding interaction technologies and assistant functionality. The entire EMBASSI system can, therefore, be structured in the following layers:

- System management
- Semantic EMBASSI protocol
- Anthropomorphous interaction
- Multimodal EMBASSI interface
- Dialogue management
- Media management
- Device management
- Unimodal I/O channels

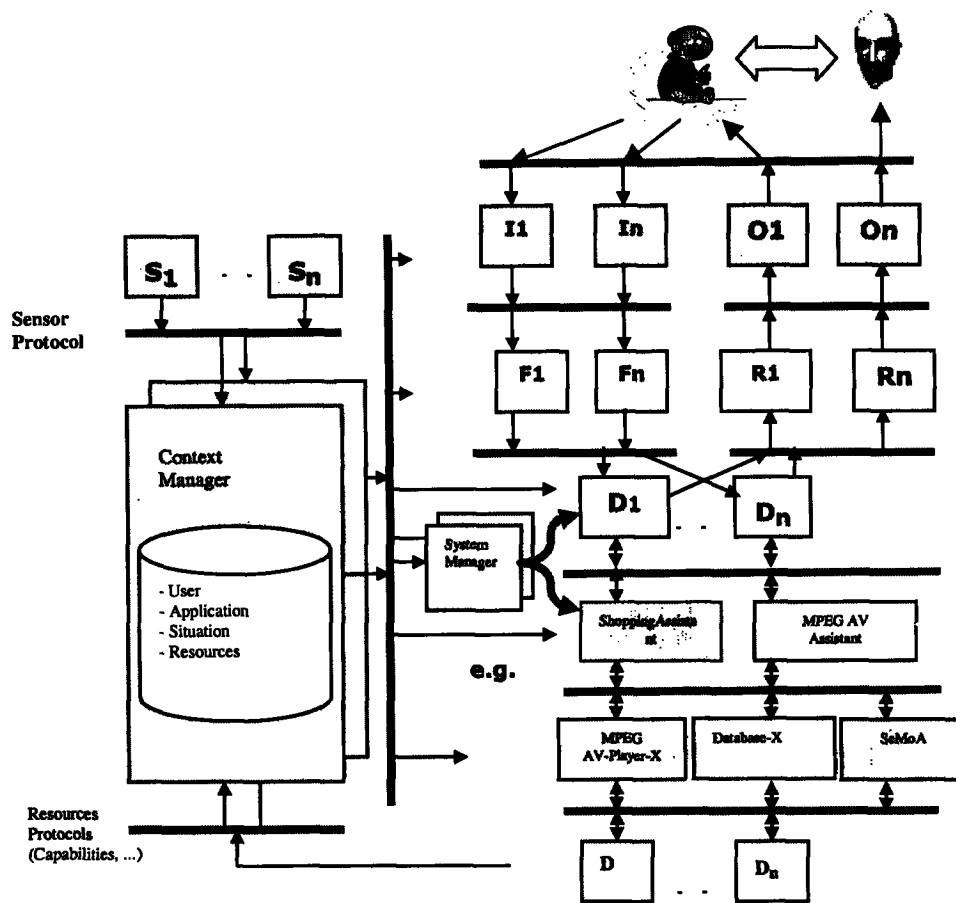


Figure 1 :EMBASSI global architecture

Different layers, from top to bottom, form a generic framework, in which the desired user interface can be configured by selection from a "modular kit" of innovative interfaces (see Figure 1).

The following fundamental aspects belong to the generic framework:

- Network techniques
- Distributed applications and operating systems
- Security aspects
- Online and remote configurability of appliances
- System management

The system management is the highest level of the EMBASSI system. The functionality of the system management is distributed and scalable. The system management is responsible for:

- the recording, storage and administration of user and appliance profiles;
- the communication between EMBASSI compatible appliances;
- the location of data and services.

### 3. Semantic EMBASSI Protocol

The communication between the multimodal interface or assistant and the appliance is managed over the semantic EMBASSI protocol. This is a mixed protocol composed of KQML wrapper

and XML content. All assistants in EMBASSI communicate over this protocol.

#### 3.1 KQML

The Knowledge Query and Manipulation Language (KQML) [2] is a high-level language intended to support interoperability among intelligent agents in distributed applications. It is both a message format and a message-handling protocol to support runtime knowledge sharing among agents. KQML is an interlingua, a language that allows an application program to interact with an intelligent system. It can also be used for sharing knowledge among multiple intelligent systems engaged in cooperative problem solving. This language, originally developed as part of a DARPA Knowledge Sharing initiative, is becoming a de facto standard for inter agent communication languages. A KQML message consists of a performative, the content of the message, and a set of optional arguments. The performative specifies an assertion or a query used for examining or changing a Virtual Knowledge Base (VKB) in the remote agent.

#### 3.2 XML(Extensible Markup Language)

XML, the Extensible Markup Language, is a new format designed to bring structured information to the Web. It is a language for electronic data interchange in the Web. XML is an open technology standard of the World Wide Web Consortium (W3C), the standards group responsible for maintaining and advancing

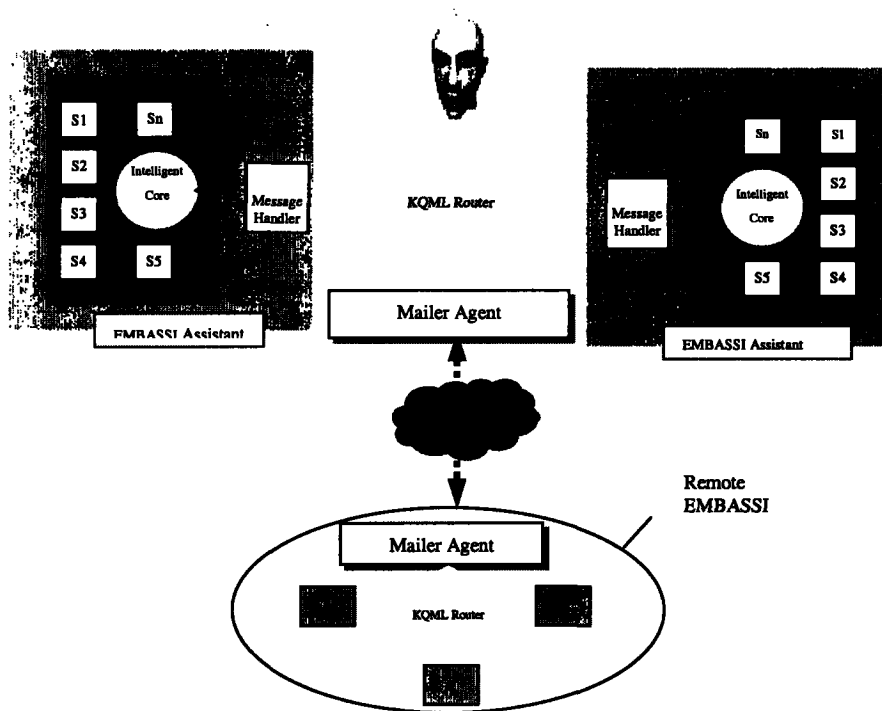


Figure2: All EMBASSI Assistants communicate over KQML

HTML and other Web-related standards. In EMBASSI, we use the XML in the content of the KQML message exchanged between the agents. Below, we give an example of such a message. In this message, the Dialogue Manager orders a western video over the shopping assistant:

Dialogue Manager → Shopping Assistant (searchForContent (ContentInfo ci, Conditions co))

(stream\_all

:sender Dialogue Manager

:receiver SHOPPING A

:reply-with id2

:ontology agent\_ontology

:language XML

:content (

<searchForContent>

<ContentInfo>

<ContentType> Video </ContentType>

<Genre> Western </Genre>

</ContentInfo>

</SearchForContent>))

All EMBASSI assistants exchange messages over the EMBASSI protocol using a KQML router (see Figure 2). In the EMBASSI context, we have developed a secure mailer agent, which helps EMBASSI assistants to communicate over the Internet and KQML router. This architecture overcomes the problem of traditional KQML routers being unable to communicate over the Internet (see Figure 2).

## 4. Multimodal EMBASSI Interface

A multimodal interface is capable of using several unimodal I/O channels simultaneously to get higher information intensity transferred. A classic example on the input side is the combination of voice input and gesture recognition: the user points to an object on the computer screen and says "close it". The combination of graphics and language is an example for the output side. An example of this is user identification through a combination of language analysis and face recognition. The multimodal interface in EMBASSI consists of three hierarchical management layers that steer the process of the interaction between human and appliance on the technology side and I/O channels (see Figure 1).

### 4.1 Dialogue Management

The dialogue management is the highest management layer of the multimodal interface. The dialogue management mediates between users and application. Therefore, it must have all necessary context information concerning them. The user profile, the appliance or application profile, I/O channel for the interaction and the situation are all part of the context information. On the basis of this data made available by the system management, a dialogue is shaped between users and application by the intercession of the EMBASSI interface and handled over existing I/O channels.

### 4.2 Media Management

A piece of information is set aside in such a form that it can be portrayed by a media manager and dialogue manager, situation-dependent on different media or devices. Thus, it is possible that a certain dialogue is established on the TV screen over an anthropomorphous conversation with an Avatar or displayed over a text message on the computer screen. It is crucial that a universal mental model of the type of conversation exists in order

to avoid a break between different interaction types during communication.

### 4.3 Device Management

The device management is the abstract layer between the media management and the different I/O channels. The device management communicates directly using a specific device driver over a network layer with the physical I/O devices. The device management allows a flexible binding of different I/O devices and offers layers for logical access to them, independent of the physical properties of individual devices.

### 5. Unimodal Input Technologies

The term 'unimodal input technologies' either includes a number of different technologies from the direct input of information through the user or from the extraction of information about the user, about the condition or the situation of the user, as well as of other context-relevant data. Classic input technologies like keyboard or mouse are also taken into account in the overall concept, since they are still important for certain applications and user groups. The following technologies have been developed and adapted to the EMBASSI framework and the typical EMBASSI applications:

- Video-based input forms
- Gesture recognition
- Facial expression and emotion recognition
- Lip reading
- Eye tracking
- Pointer device
- Sensors for the recognition of the position

### 6. Unimodal Output Technologies

Among others, EMBASSI supports the user with the following output technologies:

- Audio output unit
- Avatar

#### 6.1 Audio Output Unit

The audio output unit is an important component of the anthropomorphous interaction and consequently an object of the performances in EMBASSI. Another important feature is the intercession of the urgency of a piece of information by means of particular language emphasis.

#### 6.2 "Avatar" and "User Interface Agent"

In the context of User Interface Agent, EMBASSI aims at the

embodiment of an artificial, autonomous and anthropomorphous software agent. This necessitates a corresponding dialogue formation as component of the user-interface designs. For this entity, we use - in the EMBASSI context - the term "Avatar" [11]. This is the anthropomorphous representation of an autonomous and proactive entity that adopts the dialogue with the user. "Avatars" are mostly presented in a synthetic three-dimensional face (see Figure1). This results in a communication in the form of a natural, human-based dialogue.

### 6.3 Shopping Assistant

In the context of the EMBASSI project, we are working on the development of a home shopping scenario. The technology developed in EMBASSI is available for every component plugged to the system in a very simple manner. The shopping assistant is responsible for providing all users at home with all kinds of information and content (video, audio, etc.) from the Internet. Below, we describe a shopping scenario in which the user tries to order a video. Entities involved in the scenario are:

- **Shopping Assistant:** This is the central entity responsible for all kinds of information retrieval tasks on the Internet. This assistant manages all e-commerce-relevant tasks such as:
  - Payment
  - Certification request
  - Authentication against commercial servers
  - Authorization of different users for special orders on the Internet (child protection mechanism)
- **User Assistant:** The User Assistant is responsible for recording and managing profiles for all registered users.
- **Context Manager:** The Context Manager is aware of all situation-dependent questions. This includes the situation of:
  - Users at home
  - Devices
  - Assistants
  - Sensors
- **Generic Dialogue Manager (GDM):** controls all parallel modalities in a dialogue with the user.
- **Database Assistant:** is responsible for the management of all content storage and retrieval tasks in EMBASSI at home. This assistant is equipped with access control mechanisms.

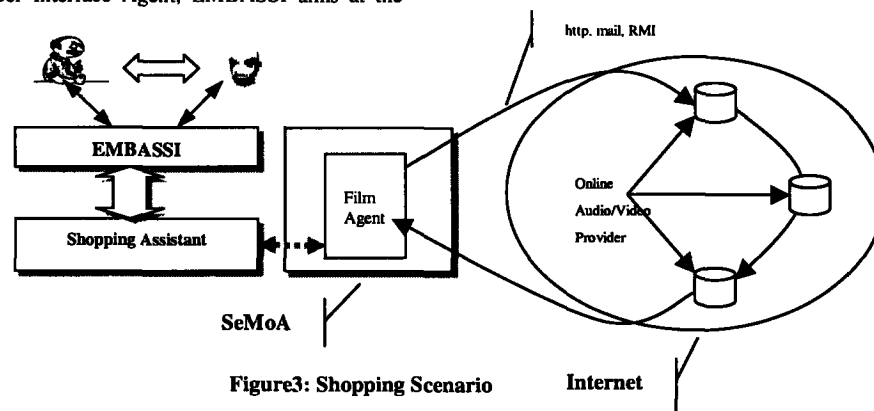


Figure3: Shopping Scenario

- **Mobile agent server SeMoA:** The Shopping Assistant may decide to send a mobile agent to perform the information retrieval tasks in order to minimize the communication costs of the transaction. In this case, the Shopping Assistant configures a SeMoA [9,10] agent and dispatches it to the agent server of the providers. SeMoA is a secure agent server developed by Fraunhofer-IGD focusing on the security requirements of mobile agent technology. The architecture of SeMoA is published mainly in [10]. After performing the tasks, the agent will return to EMBASSI and give the results to the Shopping Assistant. The second application with a mobile agent is the remote configuration of the EMBASSI devices from outside. We will publish this scenario in a separate paper in the near future.

## 7. Shopping Scenario

The User expresses the general wish: "I want to watch a video". The installed microphone captures this wish and produces an audio signal. A device manager specific for the microphone unit produces a KQML message from the input signal and sends it to the generic dialogue manager. Receiving this KQML message, the EMBASSI dialogue manager searches inside his knowledge base and finds out which function of registered assistants could match this wish. In this case, the GDM (Generic Dialogue Manager) finds the function "searchForContent" as appropriate and sends the search task "searchForContent" inside a KQML message to the Shopping Assistant. The Shopping Assistant asks the context manager which user is logged in. The context manager checks the actual situation using installed sensors and gives the UserID of the active user to the Shopping Assistant. For our first implementation, we use a finger print sensor, which is installed inside the remote Control Unit of the television. The Shopping Assistant now asks the User Assistant which genre this user may prefer. The user assistant gives a list of genres for this user to the shopping assistant as a response. The Shopping Assistant locates the desired contents on the Internet and notifies the GDM about them as an answer to the user's inquiry. The GDM presents the results over different modalities based on the user profile and the situation and asks the user to choose a content. The video can now be ordered by the online provider using a digital certificate of the user. The GDM orders the video by calling the "orderContent" function of the shopping assistant over a KQML message.

## 8. Related Work

The idea of intelligent agents evolved in the 1970's. A number of research organizations and companies are working on agents with a focus on different problematics of this topic. As e-commerce continued to expand, numerous institutes began to work on intelligent assistants for e-commerce applications. Jango [6], developed by Netbot, Inc., founded in May 1996, is an application for Windows95 or WindowsNT that works in browsers (Microsoft Internet Explorer or Netscape Navigator). A user enters the name of a product he is looking for and Jango automatically determines which stores and information sites are relevant. Jango communicates with online stores using standard Web protocols and works with existing interfaces of online stores. Tete-a-Tete (T@T) [7] is a project within MIT Media Lab's Agent-mediated

Electronic Commerce (AmEC) Initiative. T@T's approach engages consumer-owned shopping agents and merchant-owned sales agents in integrative negotiations. Mindmaker's Intelligent Personal Assistant "Prody Parrot" [8] uses interactive multimedia to interact with the user. Prody talks and interacts with the user as an Intelligent Assistant for the PC, in the form of an animated avatar on the desktop.

## 9. Conclusions and Future Work

Technology developed in EMBASSI will be available in a few years to consumers all over the world. A very important issue that we are working on is that of Integrity and the confidentiality of the user's personal data, even at home. An integrated security architecture is required to secure heterogeneous agent systems for the sensitive applications described above. This architecture should protect against the threats of open networks and inside attacks while keeping maximum operational freedom for agents, which is fundamental for the actual market exploitation of this technology.

## 10. ACKNOWLEDGMENTS

The work described is being done during the EMBASSI project. EMBASSI is sponsored by BMBF [3], German Ministry for Education and Research. We would like to thank our head of department, Dr. Christoph Busch, for his support during our work to develop the application outlined in this paper.

## 11. REFERENCES

- [1] EMBASSI (Elektronische Multimediale Bedien- und Service-Assistenz), <http://www.embassi.de>
- [2] Tim Finin, Yannis Labrou and James Mayfield, KQML as an agent communication language, in Jeff Bradshaw (Ed.), "Software Agents", MIT Press, Cambridge 1997
- [3] German Ministry for Education and Research <http://www.bmbf.de/>
- [4] Esteban Chávez, Rüdiger Ide and Thomas Kirste, Interactive applications of personal situation-aware assistants, *Computers and Graphics* 23(6) (1999) pp. 903-915
- [5] Pattie Maes, Agents that reduce work and information overload, *Communications of the ACM*, 37(7), July 1994
- [6] Jango Homepage: <http://www.jango.com>
- [7] tete-a-tete, <http://ecommerce.media.mit.edu/tete-a-tete/>
- [8] Mindmaker, <http://www.mindmaker.co>
- [9] Secure Mobile Agents Project, Fraunhofer Institute for Computer Graphics, Germany [http://www.igd.fhg.de/igda8/projects/semoa/semoa\\_de.html](http://www.igd.fhg.de/igda8/projects/semoa/semoa_de.html)
- [10] Volker Roth and Mehrdad Jalali, Concepts and Architecture of a Security-centric Mobile Agent Server, Fifth International Symposium on Autonomous Decentralized Systems, March 26-28, 2001 Dallas, Texas, U.S.A
- [11] Alfredo Pina, Eva Cerezo and Francisco J. Serón, Computer animation: from avatars to unrestricted autonomous actors (A survey on replication and modelling mechanisms), *Computers and Graphics* 24(2) (2000) pp. 297-311