Chapter 11
Multimodal and Agent-Based Human–Computer Interaction in Cultural Heritage Applications: an Overview

Antonio Gentile and Salvatore Vitabile

Summary One of the most recent and interesting applications of human–computer interaction technologies is the provision of advanced information services within public places, such as cultural heritage sites or schools and university campuses. In such contexts, concurrent technologies used in smart mobile devices can be used to satisfy the mobility need of users allowing them to access relevant resources in a context-dependent manner. Of course, most of the constraints to be taken into account when designing a pervasive information providing system are given by the actual domain where they are deployed.

This chapter presents an overview of such techniques, focused on two different approaches to the development of human–computer interaction aimed at providing solutions for engaging fruition of cultural heritage sites and exhibits. The chapter will first present multimodality as a key enabler for a more natural interaction with the virtual guide and its surrounding environment. The second approach will be presented next, offering an overview of the evolution of agent-based human–computer interaction systems for the same domain.

11.1 Introduction

The exponential diffusion of small and mobile devices, third-generation wireless communication devices, and location technologies has led to a growing interest toward the development of pervasive and context-aware services. Such technologies,

A. Gentile
Dipartimento di Ingegneria Informatica, University of Palermo, Viale delle Scienze, Ed. 6, 90128 Palermo, Italy
e-mail: gentile@unipa.it

S. Vitabile (✉)
Dipartimento di Biopatologia e Biotecnologie Mediche e Forensi, University of Palermo, Via del Vespro, 90127 Palermo, Italy
e-mail: vitabile@unipa.it

F. Xhafa et al. (eds.), *Complex Intelligent Systems and Their Applications*,
Springer Optimization and Its Applications 41,
DOI 10.1007/978-1-4419-1636-5_11,

both hardware and software, made the Mark Weiser's vision of Ubiquitous Computing real and more available for users in their in everyday life (Weiser 1991). The Ubiquitous Computing paradigm relies on a framework of smart devices that are thoroughly integrated into common objects and activities. Such a framework implements what is otherwise called a pervasive system, the main goal of which is to provide people with useful services for everyday activities.

As a consequence, the environment in which a pervasive system is operating becomes more complex, that is, enriched by the possibility to access additional information and/or resources on a per-needed basis. Augmented environments can be seen as the composition of two parts: a visible part populated by active users (visitors, operators) and/or objects (inanimate but controlled by some type of artificial intelligence) interacting through digital devices in a real landscape, and an invisible part made of software objects performing specific tasks within an underlying framework. People would perceive the system as a whole entity in which personal mobile devices are used as human–environment adaptable interfaces.

There are many domains where pervasive systems are suitably applied. One of the most recent and interesting applications of pervasive technology is the provision of advanced information services within public places, such as cultural heritage sites or schools and university campuses. In such contexts, concurrent technologies used in smart mobile devices can be used to satisfy the mobility need of users allowing them to access relevant resources in a context-dependent manner. Of course, most of the constraints to be taken into account when designing a pervasive information providing system are given by the actual domain where they are deployed.

In this chapter we will focus on two different ways to approach the development of human–computer interaction aimed at providing solutions for engaging fruition of Cultural Heritage (CH, in short) sites and exhibits. We will next focus on multimodality as a key enabler for a more natural interaction with the virtual guide and its surrounding environment. In this first part, we also evaluate the approaches examined in the chapter in terms of multimodality, presence of pervasive access to contents, and intelligent handling of interaction with the user. We will then offer an overview of the evolution of agent-based human–computer interaction, reviewing some of the most relevant papers in the past years.

11.2 Multimodal Human–Computer Interaction in Cultural Heritage Applications

Cultural Heritage applications pose tremendous challenges to designers under different aspects. Firstly, because of the large variety of visitors they have to deal with, each with specific needs and expectations about the visit. Secondly, no two sites are the same, and pretty much you need a framework that can easily produce a new installation given the site characteristics (indoor versus outdoor, distributed versus centralized, individual centered versus group centered, etc.). Lastly, the technologies involved must be robust to failures, redundant, and, above all, easy and intuitive to use. These are the reasons why there are several research groups that are focusing their attention on this applicative domain. It is a good test bed to validate almost all models and design choices.

Multimodality is usually described of such software applications or computing systems that combine multiple modalities of input and output. Being free to choose among multiple modes to interact with such augmented environments and systems is crucial to their wider acceptance by a vast variety of users. This is often the case when such systems are intended for mass fruition at museums or cultural heritage sites. Multimodality is the capability of a system to allow for multiple modes of interaction, from traditional point-and-click to voice navigation or gesture activation of controls. A pervasive system targeted to cultural heritage fruition, therefore, can no longer be designed without first addressing how it will handle interaction with users. Additionally, multimodality relies on redundant information, resulting on more dependable systems that can adapt to the needs of large and diverse groups of users, under many usage contexts.

11.3 A Timeline of Cultural Heritage Fruition Applications

Several workgroups focused their research in the past five years on the definition of models aimed at providing users of a cultural heritage site with useful services of different kinds. In this section, we will introduce some of the projects that we consider most relevant.

Specifically, we will focus on technologies used to access devices, positioning/location, human–environment interfaces, and ambient intelligence. In order to support our review, we will discuss a selection of that we deem best representing the evolution of systems for fruition of cultural heritage sites. Figure 11.1 depicts the timeline we will follow, along with the projects discussed in this chapter.

The projects presented during 2005–2006 witness the need to combine more than one technology at a time, concurrently used to implement multichannel interaction. As a matter of fact, during this timeframe, we assist to the spreading of several technologies (Wi-Fi, Bluetooth, RFID, GPS, voice recognition and synthesis, image recognition, conversational agents) and smart access devices (such as PDA and smartphones) that more and more are integrated into enhanced systems.

Farella et al. (2005) presented a work that exploits widely available personal mobile devices (PDAs and cellular phones) and software environments (usually based

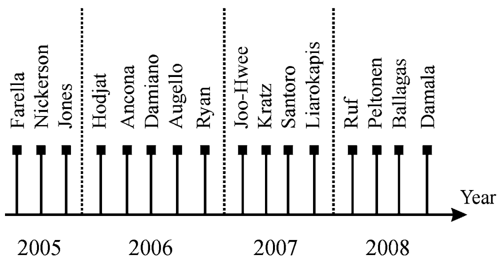

Fig. 11.1 Timeline

on Java) to create highly interactive Virtual Heritage applications based on popular, low-cost, wireless terminals with highly optimized interfaces. Ancona et al. (2006) proposed a system called AGAMEMNON that uses mobile phones equipped with embedded cameras to enhance the effectiveness of tours of both archaeological sites and museums.

The evolution of speech synthesis techniques allow for a new multimedia content delivery way, based on text-to-speech technology. This allows providing the user with new services accessible also by telephone. This solution tries to overcome the well-known limits of static prerecorded audio guides, by vocally presenting a text that can be dynamically composed. Using this technology, Nickerson (2005) proposes the system named History Calls. He developed an experimental system using VoiceXML technology that is capable to deliver an automated audio museum tour directly to cell phones.

The main target of some research group is to make the visit within a cultural heritage site as more natural as possible for the largest part of users, avoiding the use of ad hoc or complex devices. This goal can be achieved by integrating different well-known technologies in a different way.

Augello et al. (2006) proposed a multimodal approach for virtual guides in cultural heritage sites that enables more natural interaction modes using off-the-shelf devices. The system, termed MAGA, integrates intelligent conversational agents (based on chatbots) and speech recognition/synthesis in an RFID-based location framework with Wi-Fi-based data exchange. Moving along the line of a more natural interaction, Jones et al. (2005) proposed a system in which multiple virtual guides interact with visitors, each characterized by its own personality. In their system, it is then possible to interact with two intelligent agents, with divergent personalities, in an augmented reality environment. Another example of system that is capable to adapt itself to the user behavior is CRUSE (Context Reactive User Experience) proposed by Hodjat et al. (2006). It is a user interface framework that enables the delivery of applications and services to mobile users.

In Dramatour (Damiano et al. 2006), Carletto, a virtual guide for the Savoy apartments of Palazzo Chiablese (in piazza Castello—Torino) provides users with information on portable devices presenting them in a dramatized form (being himself a small spider, long-time inhabitant of Palazzo Chiablese).

The past two years (2007–2008) see the advent of ambient integrated guides that are based on mixed reality, where real physical elements interact with virtual ones. Liarokapis and Newman (2007) focus their work on the design issues of high-level user-centered Mixed Reality (MR) interfaces. They propose a framework of a tangible MR interface that contains Augmented Reality, Virtual Reality, and Cyber Reality rendering modes. This framework can be used to design effective MR environments where participants can dynamically switch among the available rendering modes to achieve the best possible visualization. Santoro et al. (2007) propose a multimodal museum guide that provides user with gesture interaction mode: using a PDA equipped with a 2D accelerometer, they control interface and content navigation by means of hand tilt gestures.

Damala et al. (2008) present a prototype of an augmented reality mobile multimedia museum guide, also addressing the full development cycle of the guide, from conception to implementation, testing, and assessment of the final system. They use last-generation ultra-mobile PCs, which allow designers to exploit some of the most powerful software technologies, such as OpenCV for video acquisition, ARToolkitPlus for tracking the paintings, OGRE3D for the insertion of the virtual objects, Open AL for the audio output, and XERCES for XML document parsing.

Overall, in the past five years the attention of researchers in the field of pervasive service provision moved to a higher level of abstraction, mainly a middleware level. Software technologies have been exploited to give systems more adaptability, usability, and, above all, intelligence. These features are the more promising ones in order to achieve the naturalness needed to get the largest acceptance of virtual guide systems among common people. As a consequence, the current research interests in the field of service provision in cultural heritage sites are mainly aimed at reaching these features. A common line researchers are following is the concurrent use of multiple hardware and software technologies at a time. The goal is to give users the possibility to interact in multiple modes, according to their habits, skills, and capabilities. This will have the desired side effect to expand the number of prospective users, allowing also disabled ones to exploit such guide systems with their residual capabilities.

11.4 Multimodal Mobile Access to Services and Contents in Cultural Heritage Sites

The cultural heritage target domain is extremely challenging for the development of systems that assist users during their visit according to their mobility requirements. Moreover these systems should attract and involve users by showing an easy and friendly access. Most of these requirements are fulfilled by means of personal mobile devices, suitably applied as adaptive human–system interfaces.

We start our discussion by summarizing the main features of those of the projects previously discussed that make use of hand-held devices (Table 11.1). Systems are listed by year of presentation, along with their features listed under four main categories: compass/position detection, context awareness, intelligent interaction, and output and input modes.

The *compass/position detection* column shows technologies used to detect the user position and orientation within the cultural heritage site. All positioning-based systems exploit information about the proximity of the user to a particular item or point of interest. With no further information, such systems are not able to detect or estimate what the user is looking at, thus making this solution unsuitable for sites with items that are close one to each other. To disambiguate such situations, information about the spatial orientation of the user within the environment is needed in addition to his position. To this end, some systems use different combinations of technologies and techniques, such as infrared combined with electronic compass or with accelerometers embedded into the hand-held devices, thus providing users with fine-tuned contents. Simpler, cheaper, and less intrusive location frameworks

Table 11.1 System features

Year	System	Compass/ position detection	Context awareness	Intelligent interaction	Input modes	Output modes
2005	Farella	–	–	–	keyboard, touch pen or touch-screen,	text, prerecorded audio, video clip, images, virtual reality
2005	Nickerson	–	–	–	vocal interface	prerecorded audio, synthesized narrations
2005	Jones	GPS	location-based, state-based, profile-based	natural language, proactivity, character	keyboard, touch pen or touch-screen, user position and/or compass	text, prerecorded audio, images
2006	Hodjat (CRUSE)	GPS	location-based, state-based, profile-based	natural language, proactivity	keyboard, touch pen or touch-screen, user position and/or compass	text
2006	Ancona		state-based, profile-based	proactivity	keyboard, touch pen or touch-screen, vocal interface, image recognition	text, prerecorded audio, synthesized narrations, video clip, images
2006	Damiano (Drama-Tour)	–	state-based	proactivity, character	keyboard, touch pen or touch-screen, user position and/or compass	video clip

are made available by means of the RFID technology. In fact, the short tag detection distance can be used to estimate the user interest for a specific item with a good accuracy.

Table 11.1 (Continued)

Year	System	Compass/ position detection	Context awareness	Intelligent interaction	Input modes	Output modes
2006	Augello (MAGA)	RFID	location-based, state-based	natural language, inference, proactivity, character	keyboard, touch pen or touch-screen, vocal interface, user position and/or compass	text, synthesized narrations, images
2007	Santoro	RFID	location-based, state-based	proactivity	keyboard, touch pen or touch-screen, gesture caption, user position and/or compass	text, pre-recorded audio, video clip, images
2007	Joo-Hwee	–	–	–	image Recognition	text, audio, images
2008	Ruf	–	location-based	–	image Recognition	text, audio, images, virtual reality

The *Context-awareness* column reports the different techniques used to build context-related contents. Systems that miss this feature provide users with information that is neither position- nor profile-based. Three different methods have been identified for contents composition to make the human–system interaction more natural and interesting:

- *location-based*: users are provided with information related to items that are located near their current position;
- *profile-based*: delivered contents are generated according to the user's profile, such as preferences, skills, age, etc.;
- *state-based*: information are generated taking into account different context factors, such as the user position, the user's profile, the interaction flow, the followed path, etc. Systems may use a subset of these elements to detect the current state (e.g., the history of inputs).

Research results in the field of artificial intelligence suggest new tools to make the interaction more natural (Jones et al. 2005; Ibanez et al. 2003; Almeida and

Yokoi 2003). To take this trend into account, in the *Intelligent Interaction* column we then report the capability of a system to implement methodologies that are typical in the field of artificial intelligence. Specifically, the *proactivity* feature shows that the system can spontaneously initiate the interaction with the user without his explicit request, according to the detected context. Systems that support *natural language* interaction accept user input in natural language (e.g.: "give me information about marble statues dating from the 5th century BC") and/or provide user with spoken information. Systems with the *character* feature embed a life-like tour virtual assistant with a specific personality. *Inference* means that a system has the capability to make inferences on domain ontology to update its knowledge base and to generate ad hoc contents.

The *Input Mode* column lists the input modes available for each system, whereas the *Output Modes* column lists the content delivery modes a system is enabled to use.

In the following we will discuss how issues and problems of multiple concurrent interaction modes have been faced in the systems and approaches examined so far. In particular we now focus our attention on concurrent input modes, as most of researchers we cited did. Actually, the processing of multiple inputs simultaneously coming from different channels present several constraints, for instance, in terms of synchronization and concurrency, whereas the management of contemporary output modes is less compelling. To this end, in Fig. 11.2 we depict the temporal evolution of such systems, comparing them under three dimensions: degree of multimodality, pervasive access to contents, and intelligence in interaction management.

Any given system is therefore classified as either *none, intelligent*, *pervasive*, or *both* according to the performed interaction, as illustrated in the figure with the bar

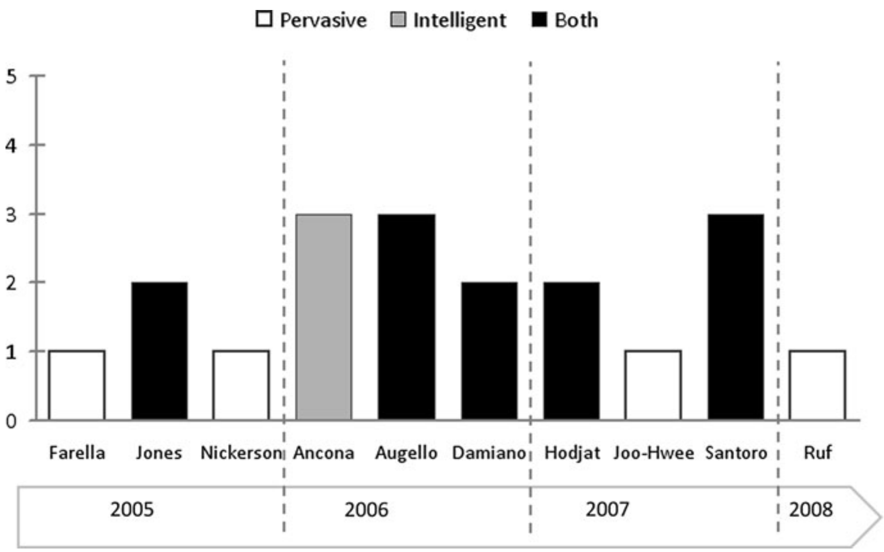

Fig. 11.2 Evolution of multimodal input and performed interaction

filling pattern. Systems classified as *none* do support neither intelligent interaction nor pervasive access. A system offers *pervasive* access to cultural heritage contents if it relies on a framework of smart devices where data are stored and processed. The retrieved information is then suitably formatted to fit the specific access device and transmitted over a wireless connection. We classify as *intelligent* in the interaction a system in which contents generation and/or delivery are managed with techniques that are inherited from the artificial intelligence research field. In particular, from the contents generation point of view, the system should use one of the following methods: state-based, location-based, profile-based, and inference. From the contents delivery point of view, the system should either interact using natural language, or exhibit proactive behavior, or show a life-like character, or any combination of the three.

The degree of *multimodality* of each system is presented on the vertical axis, and it is evaluated by counting the number of input modes by which users gain access to information.

Figure 11.2 shows that in the last decade the pervasive approach in service provision within cultural heritage sites has been largely adopted. This is mainly due to the distributed nature of data and to the need of accessing those data anywhere at any time. The figure also shows that the use of algorithms and techniques of the Artificial Intelligence field is considered a key feature for the success of a system, particularly when the main goal is the naturalness of the interaction. Pervasive systems accessed by mobile devices substituted the prerecorded audio guides, providing data mining algorithms with higher computational power and with more data sources.

In order to better highlight the evolution of interaction, in Table 11.2 we present the detailed list of input modes implemented by each of the discussed systems. This allows readers to realize that, despite new interaction modes are studied and made available, most of existing systems keep showing the traditional point-and-click interface, although being implemented by up-to-date devices.

Table 11.2 Input modes

Year	System	Keyboard, touch pen or touch-screen	Vocal interface	User position and/or compass	Gesture recognition	Image recognition
2005	Farella	X				
2005	Nickerson		X			
2005	Jones	X		X		
2006	Hodjat	X		X		
2006	Ancona	X	X			X
2006	Damiano	X		X		
2006	Augello	X	X	X		
2007	Santoro	X		X	X	
2007	Joo-Hwee					X
2008	Ruf					X

Another commonly used input mode is the user position. In fact, this allows designers to give their systems the context-awareness by taking into account an important context element. Furthermore, the needed framework can be relatively cheap and often may partly rely on some existing implementation (e.g., the GPS).

Different input modes, such as voice, gesture, and image recognition, are commonly considered as useful in order to improve the naturalness of interaction, but they have been not so largely adopted mainly due to their complexity. Advances of both hardware and software technologies in these fields are giving an important stimulus to their adoption, and several research groups are currently working in the field of integration of these technologies.

11.5 Agent-Based Human–Computer Interaction in Cultural Heritage Applications

There is a very large number of research projects coping with the problem of providing information about sites of cultural heritage value (see Table 11.3 for a list of the most relevant research projects). Complex distributed applications in very challenging domains, such as e-Business, e-Government, Cultural Heritage, etc. need to use new models to overcome classical client–server and web-based model limits. In a cultural heritage site, different mobile or fixed devices interact with a great number of services. As example, the number of devices in the museum cannot be easily predicted as well as the number of the available and active services. Environment is highly dynamic: users enter and leave the site, and availability of services cannot be guaranteed (service providers may be busy or simply off-line).

Accordingly to Ducatel et al. (2001) and Penserini et al. (2003), paradigms in which software applications are constructed based on independent component services with interfaces can be seen as a feasible solution (Service-Oriented Computing and Service-Oriented Architectures) for CH applications. Intelligent Agents are a natural choice to implement and develop the above paradigm. In fact, each service can be seen as one or more interacting autonomous agents. At the individual-agent level, each agent requires a representation of the individual behavior through elements such as beliefs, response, intentions, desires, goals, and ego. In addition, each agent-based service should provide intelligent functionalities, such as reactivity, proactivity, and social ability. On the other hand, those applications need intelligent mechanism to exhibit, compose, and adapt its behaviors to external changes and/or triggers.

Accordingly to Lopez-Jaquero et al. (2009), Agent-Based Human–Computer Interaction (AB-HCI) supports the execution of human-centerd applications by means of one or more interface agents aimed at providing an advanced user interaction experience by processing the incoming information from the environment and applying techniques coming from different disciplines (human–computer interaction, software engineering, artificial intelligence, psychology, ...). In the cultural heritage domain, the required software architecture can really benefit from using intelligent

Table 11.3 System features

Project	System	Compass/ position detection	Context awareness	Intelligent interaction	Input modes	Output modes
Minerva	Amigoni and Schiaffonati (2003, 2009)	–	–	natural language	keyboard, natural language	HTML pages, VRML objects
Teschet	Pilato et al. (2004a)	–	–	inference	keyboard, touch pen or touch-screen	Text
DramaTour	Damiano et al. (2006)	–	state-based	proactivity, character	keyboard, touch pen or touch-screen, user position and/or compass	video clip
Row 7 MAGA	Augello et al. (2006)	RFID	location-based, state-based	natural language, inference, proactivity, character	keyboard, touch pen or touch-screen, vocal interface, user position and/or compass	Text, synthesized narrations, Images
PEACH	Stock et al. (2007)	IR, IRDA, RFID	location-based, profile-based	proactivity	keyboard, touch pen or touch-screen	video clip
MOSAICA	Shah et al. (2007)	–	–	–	keyboard, mouse, touch-screen	images, textual and vocal description
Cuspis	Costantini et al. (2008)	satellite signals	location-based, profile-based	inference, proactivity	user position	Text

agents, since they support a much more natural method to design decision-making mechanisms by providing constructs closer to the ones used in human reasoning theories.

At the same time, agent-based systems offer different key benefits. First of all, MAS is the natural paradigm allowing for the distribution of computation, making

Table 11.3 (Continued)

Project	System	Compass/ position detection	Context awareness	Intelligent interaction	Input modes	Output modes
–	Lopez-Jaquero et al. (2009)	–	environment-based, profile-based, device-based	proactivity, adaptivity	keyboard, touch pen or touch-screen	text, images, vocal
AB-MCUI	Huang et al. (2009)	camera, sensor, accelerator	cultural-based, state-based	speech recognition, non-verbal behavior, character	vocal interface, user position and/or compass, user gesture and movements, cultural information	recorded audio, animations

easier the integration with current trends in software design such as service-oriented application development. The main idea is the efficient utilization of the resources available to a mobile device in its vicinity, whether it is in the wired network or whether the mobile device is in an ad hoc network. Accordingly to Lind (2001), a MAS consists of multiple independent entities, which coordinate with themselves in order to achieve their individual and their joint goals. In addition, agent-based design methodologies can be extended to the programming of massively parallel computing platforms, since an agent can be seen as the instance of an agent class that is the software implementation of an autonomous entity capable of pursuing an objective through its autonomous decisions, actions, and social relationships (Gentile et al. 2002). In designing a mobile-based application, certain restrictions are applied, such as limited memory and processing resources. Each agent will be represented to solve a scheduled problem. In this view, the mobile device service application handles the mobile user interface that relates to the interaction between the application and the user, covering user's data-entry and portable device output functions (Abdel-Naby et al. 2007).

On the other hand, a user-centered approach adapted to the design of collaborative technologies is more accurately referred to as a group-centered approach (McNeese et al. 1995). In order to support a user-centered approach to the design of collaborative technologies, it becomes necessary to acquire knowledge about the group, the group work domain, and the group design requirements directly from the group. In a MAS, separate software entities are forming Virtual Communities

(VC) (Rakotonirainy et al. 2000). Software entities of each VC collaborate to obtain certain results, and they usually share the same interests and goals. The most usual features of CSCW (Computer Supported Cooperative Work) applications are commitment (goals), cooperation, coordination, and competition (Kling 1991). So intelligent agents supporting group behaviors and goals can be seen as a powerful tool to develop GDSS (Group Decision Support Systems) for group-centered applications.

In literature, two main approaches have been followed to design and implement AB-HCIs. The first approach exploits Multi-Agent Systems (MASs) facilities and capabilities for designing and developing cultural heritage solutions and applications. A generic MAS gives the freedom to develop specialized agents for intelligent reasoning tasks and for the interaction between human and computer. Agent communication and interaction are implemented exploiting the low-level functionalities supplied by existing frameworks and platforms for agent development. The second approach deals with the development of relational and conversational agents designed to create and maintain long-term social and emotional relationships with users. In the following, the most meaningful solutions for the two approaches are briefly described.

11.5.1 Multi-Agent System-Based Solutions for AB-HCI

In Amigoni and Schiaffonati (2003) a multiagent system, called Minerva, supporting the creative work (the preparation and the allocation of a virtual museum or a virtual art exhibition) of museum organizers is presented. Minerva is related to archaeological findings and provides, through a graphical interface, different level (role-based) user interaction. From the architectural point of view, Minerva is composed of components and agents. The user interface component generates HTML pages with VRML objects, as response to a user request. The user, through a graphical interface with drop down menus and free text fields, chooses some parameters and sends the request to the Minerva system. A natural language processing component implements basic natural language understanding techniques for a subset of Italian language. Minerva architecture is completed by two main intelligent agents interacting with the above components: the Preparator Agent determines, on the basis of the user selected criteria, the rules to display the works of art of a collection. The Allocator Agent is able to find, within a selected environment, the right position for the works of art in the museum rooms taking into account some constraints (the number of works, room order and position, etc.). In Amigoni and Schiaffonati (2009) an evolution of the previous architecture is presented. The new architecture is a MAS-based architecture in which multiple clients interact with the Minerva server and, consequently, with the Preparator Agent and Allocator Agents (one for each client). VRML builder agents build the requested views. The new architecture has been developed using the JADE platform.

In Pilato et al. (2004a, 2004b) a Multi-Agent System (MAS) for automatic and concurrent document retrieval in the cultural heritage domain was presented. The system was composed by four agents: a Trainer Agent, a mobile Neural Classifier Agent, a Librarian Agent, and finally an Interface Agent. The system was based on both the mobile agent paradigm and neural network architecture for a sub-symbolic knowledge representation. The main feature of the system was its versatility in managing documents whose topic class is unknown and not a priori fixed: the system autonomously adapts its document classification capability, exploiting the web directories available in the most common search engines. The entire system was developed using the JADE (Java Agent DEvelopment Framework) platform (Bellifemine et al. 1999). Concerning user interaction, the Interface Agent (IA) implemented a simple Graphic User Interface with drop down menus and provides, on Personal Digital Assistant devices, a front end application to the end user, for checking user inputs and displaying results. The work was developed within the TESCHET (A Technology System for Cultural Heritage in Tourism) research project, having the goal to create a multichannel platform based on pervasive, intelligent, and agent technologies with ontological classification of information for the Italian tourism and cultural heritage domains.

In Stock et al. (2007) the experience and the results of the PEACH (Personal Experience with Active Cultural Heritage) project are reported and analyzed. The PEACH project, located in the Castle of Buonconsiglio in Trento (Italy), was aimed to create an interactive and personalized guide for enhancing cultural heritage enjoyment through individual's background, needs, and interests profiling. From our perspective, the system is composed of two main (interdependent) components: a three-tier "classical" application containing a presentation layer (User Assistant—UA) and a MAS to provide the required services and interactions. Agents communicate in order to provide relevant and personalized presentations to the museum visitor based on his/her location and interest. The UA runs on the user's PDA and provides the system interface, while a Presentation Composer receives explicit and implicit user requests (user interest propagation) and replies with appropriate presentations (small Flash presentations). PEACH focus is essentially on presentations personalization, incorporating the information supplied from a positioning system and the concept of situation-aware content. Peach also focus on the concept of Active Museum, an intelligent and pervasive environment.

In Shah et al. (2007) an agent-based approach to develop the Semantically Enhanced, Multifaceted, Collaborative Access to Cultural Heritage (MOSAICA) pedagogical framework is proposed. MOSAICA is organized as an advanced web portal and has the purpose to design a toolbox for intelligent presentation, knowledge-based discovery, and interactive and creative interactions covering a broad variety of cultural heritage resources. The initial focus of MOSAICA was Jewish cultural heritage. The MAS-based approach (the system is composed of two interacting agents) is used to develop virtual expeditions as specific educational instruments based on conceptual modeling and designed for learning through exploration of virtual worlds. MOSAICA's framework is composed of several navigational interfaces, integrating documents, images, and GIS data for virtual explorations.

In Costantini et al. (2008) the DaliCa multiagent system, exploiting intelligent agents, was proposed. DaliCa was developed as central component of the European

Cultural Heritage Space Identification System (Cuspis) project, and it addresses the dissemination of information about cultural assets. DaliCa was successfully tested at the University of L'Aquila and at the Villa Adriana (Rome, Italy) area. The designed MAS application consists of three application agents and three application environment components. When a new user starts on a visit, the generator agent produces an initial user profile agent that is able to monitoring visitors' interests and behavior. An output agent performs information exchange between DaliCa and external infrastructures. The application environment has three components as well. The ontology interface gives agents information about the cultural heritage site context. The visitor interface sends to the DaliCa systems visitors' positions and data (the Galileo satellite signal is used for the visiting Point Of Interest localization). When a new visit starts, the user profile agent elaborates either the data coming from the initial profile or the new data derived from the user behavior, deducing visitor interests. Successively, the agent suggests the most appropriate sites to the visitors through their PDAs. DaliCa was implemented in DALI, a logical-agent-oriented language (Costantini and Tocchio 2004).

In Lopez-Jaquero et al. (2009) an adaptive interface based on a set of collaborative agents to assist the user in handling some tasks in a museum is proposed. The architecture is able to detect the context of use (AgentDetectContextOfUse) through the received information about the characteristic of the adopted platform and devices (AgentContextPlatform), about the user's goals (AgentContextUser), and about the characteristics and the changes of the environment (AgentContextEnvironment). The information about the perceived context of use is forwarded to the Agent Adaptation Process for the selection of the interaction rules fitting the supplied information. Moreover, agents update the context-of-use models (user, platform, and environment) to keep an up-to-date view of the context where the interaction takes place. As example, in the museum, the orientation of the screen in the PDA could be changed by the user. This change is automatically detected by means of software sensors (AgentDetectContextOfUse) and forwarded to the Agent Adaptation Process for the activation of the related adaptivity rules.

11.5.2 Conversational Agent Based Solutions for AB-HCI

Embodied Conversational Agents (ECA) are computer-generated human-like characters that improve the naturalness of the interaction between humans and computers. To achieve that feature, agents assemble several intelligent features, such as natural language understanding and generation, sensor data processing, gesture recognition and generation, personality modeling, facial expression recognition and generation, and so on. In the previous sections, two examples of conversational agents have been presented and described.

In Augello et al. (2006) a conversational agent for achieving natural interactions and enabling site fruition also by inexpert and/or disabled users was proposed. The conversational agent is enriched by a speech recognition/synthesis module and an

RFID-based auto-localization module. The agent integrates also reasoning capabilities, since an ontology for the specific domain was firstly created and then combined with the agent dialogue module.

In Damiano et al. (2006) a character, representing a teenage spider with an anthropomorphic aspect (Carletto), has been designed to create an interactive guided tour. The application was tested in the historical site of Palazzo Chiablese in Turin (Italy): Carletto's family has inhabited the palace from ages, so that the spider knows either the history of the palace or a lot of funny anecdotes. Carletto has been designed and developed through a new methodology, called DramaTour, for creating information presentations based on a dramatization. The methodology has a modular structure integrating the handling of user interactions, the content organization, and the final delivery of audiovisual contents.

In Huang et al. (2009) a general purpose framework to build an ECA-based customer application is proposed. The framework was developed as part of the "An Agent Based Multicultural User Interface in a Customer Service Application" (AB-MCUI, in short) project. A characterized tour guide of Dubrovnik, Croatia, answers queries coming from human visitors with verbal and nonverbal interactions. The tour guide is able to recognize and interact with Japanese or Croatian visitors, adapting itself to the Japanese mode (speaking and behaving following Japanese rules) or to the Croatians mode (speaking and behaving following European rules). User interaction is implemented through natural language speaking and nonverbal behaviors such as pointing to an object on the background image. Advanced modules for head orientation tracking, hand shape recognition, and head nodding/shaking recognition have been also implemented. Tour guide design and development have been performed using the GECA (Generic ECA) framework. GECA is composed of a low-level communication platform, a set of communication API libraries, and a high-level protocol (XML-based messages).

11.5.3 Discussions and Comparisons

In Table 11.3, there are summarized the main features of the previously discussed projects that make use of hand-held devices, PDAs, or laptop. Systems are listed by year of presentation, along with their features listed under the same four main categories previously listed: compass/position detection, context awareness, intelligent interaction, and output and input modes.

As stated before, the *compass/position detection* column shows technologies used to detect the user position and orientation within the cultural heritage site. The most recent solutions use location systems to detect the proximity of the user to a particular item or point of interest. Some systems use also advanced techniques (IR, RFID) to obtain information about the spatial orientation of the user within the environment. In Huang et al. (2009) an interesting approach based on data coming from IR and CCD cameras, motion, tracking, and accelerator sensors is proposed.

As previously defined, the *Context-awareness* column reports the different techniques used to build context-related contents. In our analysis three new methods have been identified for contents composition in HCI (see Sect. 11.4 for *location-based*, *profile-based*, and *state-based* definitions):

- *environment-based*: the system is provided with information about the physical environment (for example, light condition);
- *device-based*: the system is provided with technical information about the used mobile device (display features and resolution, device orientation, etc.);
- *cultural-based*: the system is provided with information about visitor origin and culture in order to address human–computer interaction with typical movements, speeches, and expressions.

As stated before, the *Intelligent Interaction* column reports the capability of a system to implement methodologies that are typical in the field of Artificial Intelligence or Computational Intelligence. For this purpose, we have also added *speech recognition* and *nonverbal behavior* (gesture, motion, and expression recognition) among the items for intelligent interaction.

The *Input Modes* column and the *Output Modes* column report the input modes available for each system and the content delivery modes a system is enabled to use, respectively.

In Table 11.4, the most meaningful system implementation features are summarized. In the most of cases, MAS-based HCI has been developed using an agent development methodology, an agent platform, or both. Agent development methodologies are software tools for designing and developing multiagent societies. They often integrate design models and concepts from both Software Engineering and Artificial Intelligence approaches. Among the agent development methodologies used in the cultural heritage domain, there are the PASSI (a Process for Agent Societies Specification and Implementation) methodology (Cossentino and Potts 2002), the Tropos methodology (Giunchiglia et al. 2002), and the Prometheus methodology (Padgham and Winikoff 2005). JADE (Java Agent DEvelopment Framework) platform is a software framework that facilitates the implementation of multiagent systems through a middleware and a set of graphical tools for the debugging and deployment phases (Bellifemine et al. 1999). Analogous facilities offer the GECA framework used in Huang et al. (2009). DaliCa multiagent system (Costantini et al. 2008) was developed and implemented using the logical-agent-oriented DALI language (Costantini and Tocchio 2004). So in cultural heritage domain, agent design and development tools and frameworks to simplify application development are widely used for MAS-based HCI . However, they do not implement a character-based interaction, even if it is the most emotional and natural interface for HCI.

In contrast with MAS-based HCI, single-agent-based human–computer interaction is implemented embedding a life-like virtual character with a specific personality. Generally, no methodology or platform are used to develop conversational agents, even if Dramatour (Damiano et al. 2006) has been designed and used for the anthropomorphic spider design.

Table 11.4 System implementation details

System	MAS based	Character based interaction	Agent development methodology	Agent platform/ language
Pilato et al. (2004a)	YES	NO	PASSI	JADE
Damiano et al. (2006)	NO	YES	Dramatour	–
Augello et al. (2006)	NO	YES	–	–
Stock et al. (2007)	YES	NO	TROPOS	–
Shah et al. (2007)	YES	NO	–	–
Costantini et al. (2008)	YES	NO	–	DALI
Lopez-Jaquero et al. (2009)	YES	NO	Prometheus	–
Huang et al. (2009)	NO	YES	–	GECA
Amigoni and Schiaffonati (2009)	YES	NO	–	JADE

11.6 Conclusions

Cultural heritage fruition and communication are an exciting new arena to exercise many enabling technologies and study novel, more natural interaction schema, all aimed at engaging visitors with multisensorial, memorable visit experiences. We have thus explored systems for cultural heritage fruition looking at their capability to engage visitors with pervasive, multimodal, intelligent access to information contents. We have particularly focused on multimodality as it is a key enabler for natural, unobtrusive interaction with exhibits and site virtual environment. In addition, we have also looked at how pervasive access to information contents is deployed and to what degree agent-based designs may provide some intelligence to resemble human tracts in system responses to visitors' queries and preferences.

It appears that, in recent years, applications for cultural heritage fruition have focused on locating visitor position inside the environment, as this piece of information is a key to contextualize dynamic contents and offer a natural interaction, specifically geared for the exhibit/artifact at hand. Voice-based techniques are used to a lesser extent, due to the difficult tuning that vocal interface often require. The development of robust speech and speaker recognition systems that operate reliably in crowded, noisy environments is still an open research. An interaction mode that is receiving a growing interest is movement and gesture recognition. In the literature, many exemplary projects are available that propose immersive fruition of virtual world and augmented reality, in which gesture-based commands are intuitive for the visitor.

On the other hand, agent-based HCI provides intelligent functionalities, such as reactivity, proactivity, and social ability. MAS paradigm gives the freedom to develop specialized intelligent agents for high-level reasoning tasks, while ECAs allow one to develop natural and emotional interaction between humans and humanized computers. Several frameworks for MAS development have been developed, so that MAS-based HCI can exploit the low-level functionalities supplied by existing frameworks and platforms for agent development, communication, and interaction.

Every year, eighty percent of the visitors of cultural heritage sites and museums keep a cellphone in their pockets. It is estimated that in excess of fifty percent of those devices are Java enabled and capable to access Bluetooth or WiFi networks. Whichever their language, they might be tapping into a local server to access their site guide, maybe previously customized on the web before heading off their visit. Those visitors may enjoy a different visit experience, where multimodal interaction with the site, mediated by their own personal device, will result in lasting memories to treasure and relive, once back at home, with family and friends. Enabling this vision could be the mission for Human–Computer Interaction studies on next generation cultural heritage applications.

References

Abdel-Naby, S.; Giorgini, P.; Weiss, M. Design patterns for multiagent systems to elevate pocket device applications. *Agents World (ESAW'07)*. NCSR "Demokritos", Athens, Greece, 22–24 October 2007.

Almeida, P.; Yokoi, S. (2003). Interactive character as a virtual tour guide to an online museum exhibition. *Proceedings of Museum and the Web*.

Amigoni, F.; Schiaffonati, V. (2003). The Minerva multiagent system for supporting creativity in museums organization. *Proceedings of the IJCAI2003 (Eighteenth International Joint Conference on Artificial Intelligence), Workshop on Creative Systems: Approaches to Creativity in AI and Cognitive Science (IJCAI)*, Acapulco, Mexico, 9–10 August 2003, pp. 65–74.

Amigoni, F.; Schiaffonati, V. (2009). The Minerva system: a step toward automatically created virtual museums. *Applied Artificial Intelligence*, 23(3), pp. 204–232.

Ancona, M., Cappello, M., Casamassima, M., Cazzola, W., Conte, D., Pittore, M., Quercini, G., Scagliola, N., Villa, M. (2006). Mobile vision and cultural heritage: the AGAMEMNON project. *Proceedings of the 1st International Workshop on Mobile Vision*, Austria, May 2006.

Augello, A., Santangelo, A., Sorce, S., Pilato, G., Gentile, A., Genco, A., Gaglio, S. (2006). MAGA: a mobile archaeological guide at Agrigento. *Proceedings of the Workshop "Giornata Nazionale su Guide Mobili Virtuali 2006" (ACM-SIGCHI)*, Italy, Torino, 18 ottobre 2006, online proceedings http://hcilab.uniud.it/sigchi/doc/Virtuality06/index.html.

Bellifemine, F., Poggi, A., Rimassa, G. (1999). JADE: a FIPA-compliant agent frame work. *CSELT internal technical report. Part of this report has been also published in Proceedings of PAAM'99*, London, April 1999, pp. 97–108.

Cossentino M., Potts C. (2002). A CASE tool supported methodology for the design of multi-agent systems. *Proceedings of the 2002 International Conference on Software Engineering Research and Practice (SERP'02)*, Las Vegas, NV, USA, 24–27 June 2002.

Costantini, S., Tocchio A., (2004). The DALI logic programming agent-oriented language. *Logics in Artificial Intelligence, Proceedings of the 9th European Conference*. *LNAI*, vol. 3229, Springer, Berlin, pp. 685–688.

Costantini S., Mostarda, L., Tocchio, A., Tsintza P. (2008). DALICA: agent-based ambient intelligence for cultural-heritage acenarios. *IEEE Intelligent Systems*, 23(2), pp. 34–41.

Damala, A., Cubaud, P., Bationo, A., Houlier, P., Marchal, I. (2008). Bridging the gap between the digital and the physical: design and evaluation of a mobile augmented reality guide for the museum visit. *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA '08)*, Athens, Greece, pp. 120–127.

Damiano, R., Galia, C., Lombardo, V. (2006). Virtual tours across different media in DramaTour project. *Workshop Intelligent Technologies for Cultural Heritage Exploitation at the 17th European Conference on Artificial Intelligence (ECAI 2006)*, Riva del Garda, pp. 21–25.

Ducatel K., Bogdanowicz M., Scapolo F., Leijten J., Burgelman J.-C. (2001). Scenarios for ambient intelligence in 2010. *Technical report, Information Society Technologies Programme of the European Union Commission (IST)*, Feb. 2001. http://www.cordis.lu/ist/.

Farella, E., Brunelli, D., Benini, L., Ricco, B., Bonfigli, M.E. (2005). Computing for interactive virtual heritage. *IEEE Multimedia*, 12(3), pp. 46–58.

Gentile A., Cossentino, M., Vitabile, S., Chella, A., Sorbello F. (2002). Intelligent agent mapping on a massively parallel MIMD computing platform. *Proceedings of 2002 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2002)*, vol. II, CSREA Press, Las Vegas, pp. 850–855.

Giunchiglia F., Mylopoulos J., Perini A. (2002). The Tropos software development methodology: processes, models and diagrams. *AAMAS Conference 2002*.

Hodjat, B., Hodjat, S., Treadgold, N., Jonsson, I. (2006). CRUSE: a context reactive natural language mobile interface. *Proceedings of the 2nd Annual International Workshop on Wireless Internet (WICON '06)*, Boston, Massachusetts, 2–5 August 2006, ACM, New York, p. 20.

Huang H., Cerekovic, A., Pandzic, I. S., Nakano, Y., Nishida, T. (2009). Toward a multi-culture adaptive virtual tour guide agent with a modular approach. *AI & Society*, 24(3), pp. 225–235.

Ibanez, J., Aylett, R., Ruiz-Rodarte R. (2003). Storytelling in virtual environments form a virtual guide perspective. *Virtual Reality*, 7(1).

Jones, C.M., Lim, M.Y., Aylett, R. (2005). Empathic interaction with a virtual guide. *AISB Symposium on Empathic Interaction*, University of Hertfordshire, 12–15 April 2005.

Kling, R. 1991. Cooperation, coordination and control in computer-supported work. *Communications of the ACM*, 34(12), pp. 83–88.

Liarokapis F., Newman, R. M. (2007). Design experiences of multimodal mixed reality interfaces. *Proceedings of the 25th Annual ACM International Conference on Design of Communication*, El Paso, TX, USA, pp. 34–41.

Lind, J. (2001). Iterative software engineering for multiagent systems. *The MASSIVE Method. Lecture Notes in Computer Science*, vol. 1994, Springer, Berlin.

Lopez-Jaquero, V., Montero, F., Gonzalez, P. (2009). AB-HCI: an interface multi-agent system to support human-centred computing. *Software, IET*, 3(1), pp. 14–25.

McNeese, M.D.; Zaff, B.S.; Citera, M.; Brown, C.E.; Whitaker, R. (1995). AKADAM: eliciting user knowledge to support participatory ergonomics. *The International Journal of Industrial Ergonomics*, 15(5), pp. 345–363.

Nickerson, M. (2005). All the world is a museum: access to cultural heritage information anytime, anywhere. *Proceedings of International Cultural Heritage Informatics Meeting, ICHIM05*, Paris, September 2005, pp. 2–15.

Padgham, L., Winikoff, M. (2005). Prometheus: a practical agent-oriented methodology. *Agent-Oriented Methodologies*, edited by B. Henderson-Sellers and P. Giorgini, Idea Group.

Penserini, L., Liu, L., Mylopoulos, J., Panti, M., Spalazzi, L. (2003) Cooperation strategies for agent-based P2P systems. *WIAS: Web Intelligence and Agent Systems: An International Journal*, 1(1), pp. 3–21.

Pilato, G., Vitabile, S., Conti, V., Vassallo, G., Sorbello, F. (2004a). Web directories as a knowledge base to build a multi-agent system for information sharing. *Web Intelligence and Agent Systems, An International Journal*, 2(4), pp. 265–277.

Pilato, G., Vitabile, S., Conti, V., Vassallo, G., Sorbello, F. (2004b). A mobile agent based system for documents classification and retrieval. *Intelligenza Artificiale*, 1(3), pp. 34–40.

Rakotonirainy, A., Loke, S. W., Zaslavsky A. Multi-agent support for open mobile virtual communities. *Proceedings of the International Conference on Articial Intelligence (IC-AI 2000)* (Vol. I), Las Vegas, NV, USA, pp. 127–133, 2000.

Santoro, C., Paternò, F., Ricci G., Leporini B. (2007). A multimodal mobile museum guide for all. *Proceedings of the Mobile Interaction with the Real World (MIRW 2007) in Conjunction with the 5th Workshop on "HCI in Mobile Guides"*, Singapore, September 09.

Shah, N., Siddiqi, J. I. A., Akhgar, B. (2007). An intelligent agent based approach to MOSAICA's pedagogical framework. *Proceedings of the Ninth International Conference on Enterprise Information Systems (ICEIS)*, pp. 219–226.

Stock, O., Zancanaro, M., Busetta, P., Callaway, C., Krüger, A., Kruppa, M., Kuflik, V., Not, E., Rocchi, C. (2007). Adaptive, intelligent presentation of information for the museum visitor in PEACH. User Modeling and User-Adapted Interaction, 17(3), pp. 257–304.

Weiser M. (1991). The computer for the twenty-first century. *Scientific American*, pp. 94–10.