Ambient Intelligence
G. Riva, F. Vatalaro, F. Davide, M. Alcañiz (Eds.)
IOS Press, 2005, http://www.ambientintelligence.org

6 A Flexible Architecture for Ambient Intelligence Systems Supporting Adaptive Multimodal Interaction with Users

Stefano PIVA, Carlo BONAMICO, Carlo REGAZZONI,
Fabio LAVAGETTO

Abstract. Ambient Intelligence (AmI) systems aim at augmenting real environments to create Smart Spaces where users are provided with pervasive virtual services. In order to allow users to seamlessly complete their tasks across a multitude of smart devices, and across different physical locations, the AmI infrastructure must be complemented with ubiquitous Intelligent User Interfaces able to adapt the interaction to their characteristics and needs.

In this chapter we address three main issues related to supporting an adaptive interaction between the user and the Smart Space: how can we approximate in an AmI system the deep understanding of the context typical of human beings in the user interface of a Smart Space? How can we replicate the richness and flexibility of human-to-human communication? How do we design a flexible architecture which can effectively link the context sensing and multimodal communication functionalities?

The proposed approach is twofold: on one hand, we use a neuro-biological model of awareness as a blueprint for the design of a flexible architecture for AmI systems; on the other hand, we exploit a Virtual Character based interface to allow the system to address the user by means of a range of natural communication modalities: voice, eye gaze, facial expressions and gestures.

Finally, the chapter discusses significant results and opens issues observed during the validation of the proposed architecture and multimodal interaction paradigm in an experimental test-bed, where an AmI infrastructure is integrated in a limited area of our university.

Contents

6.1 Introduction

Thanks to advances in a range of heterogeneous fields, from video analysis and understanding to software agents, from multi-sensor context assessment to pervasive communications, research on Ambient Intelligence systems is making significant progress towards the implementation of Smart Spaces (SS) where users are provided with an ensemble of ubiquitous virtual services [19][22][25][35]. ISTAG gives a more formal definition of Ambient Intelligence (AmI) that points out how it should provide technologies to surround users with intelligent sensors and interfaces and to support human interactions [1].

It is remarkable that, with the increase in AmI systems capabilities, the paradigm of interaction with the user is progressively becoming more complex than the one of well-defined “applications” which is commonplace on desktop systems. Smart Spaces have the ambitious goal of augmenting user capabilities by supporting them in their tasks, which involve interaction not only across a multitude of smart devices (including PCs, networks, PDAs, home automation systems, 3rd generation mobile phones), but also across different physical locations.

While the specific services being provided can vary significantly according to the application domain, a common requirement for many AmI systems is that of supporting their users with guidance on available services, safety warnings and navigation aids.

Providing effective assistance to users in such a complex scenario requires giving to the system a high degree of awareness of the user’s general preferences and current activities [22]. Moreover, this requires allowing the system to suitably communicate with the user at any time. While the detection and analysis of user activities is a well-known topic in Ambient Intelligence research, the communication between the user and the environment has been less explored [25].

Ideally, the user should be able to interact with the virtual services provided by the Smart Space as directly as he/she interacts with physical objects and tools, and to communicate with the system as intuitively and robustly as with another human [34].

According to this, Brooks states that an Intelligent Environment has to make computation “ready-at-hand”, putting computers out into the real world of people more than people into the virtual world [2].

We believe that a key issue in the development of pervasive Intelligent User Interfaces is the need for adapting the interaction to the user, taking also into account that while hardware and software capabilities of devices constantly increase, the user’s level of attention remains constant [19]. Not only different users need to receive different feedbacks from the system according to their expertise, habits and preferences, but also the same user will vary his interaction according to his current tasks and priorities. As an example, consider the different interaction styles between a novice user exploring the system and an expert user trying to complete a well-known task in a hurry.

In this work we report the results being achieved within two major national research project: *VICOM* [3] and *PER2* [4]. The projects address complementary aspects of the creation of augmented environments: the reference scenario is that of an intelligent university campus, where researchers and students, beside physical facilities available in the real space, also seamlessly take advantage of virtual services provided by the Smart Space infrastructure.

Specifically, in this chapter we address three main issues related to the creation of adaptive interaction and communication between the user and the Smart Space:

- How can we approximate in an AmI infrastructure the deep understanding of the context typical of human beings?
- How can we replicate the richness and flexibility of human-to-human communication in the user interface of a Smart Space?
- How do we design a flexible architecture which can effectively link the context sensing and multimodal communication functionalities?

Context perception and understanding involve all the issues related to intelligent sensors and the management of heterogeneous information sources. As a human being perceives through his senses a huge amount of information and exploits them to decide for his actions, an artificial intelligent system based on contextual sensing must be provided with the appropriate sensors' set and with the not easily achievable ability to extract interesting data at a higher abstraction level [36]. In these systems the central concept of *context awareness* represents the possibility for the system of biasing itself and its reactions to the environment [3][37].

Developing effective interfaces to the virtual services provided by the Smart Space is a challenging task, not only because it is necessary to hide the complexity of the underlying infrastructure to the user, but because it must be taken into account that he will need to access those services pervasively and that his attention will also be partly absorbed by other tasks (such as walking or interacting with other people)[19]. In such a complex scenario, GUI-based interaction paradigms typical of desktop operating systems, show their limitations [34]. For instance, textual and pointer based interaction are often not practical when the user is moving [21][26].

A promising approach, which has been explored in this research, is that of exploiting the advances in multimodal interfaces [15][29] and particularly *Virtual Character based interfaces* [13][17][25], which allow the system to address the user by means of a range of natural communication modalities: voice, eye gaze, facial expressions and gestures [8][32].

Additionally, an high degree of flexibility can be achieved if a Virtual Character (sometimes less precisely referred to as an *Avatar*) can adapt not only message content but also the related emotional information by means of non-verbal modalities.

It is easy to understand that the development of such a system is challenged by a large number of open problems, not last the design of an architecture assembling the fundamental functions of a Smart Space: *Perception*, *Analysis*, *Decision*, and *Action* (including *Communication* with the user). The aim is then the one to define a general structure able to take into account all the possible issues related to the management of so many different heterogeneous technologies.

In this work we try and define a logical architecture we think has the generality and the completeness required to allow for the analysis of many of the possible problems raised by a Smart Space. Adapting the content and presentation modalities of informative messages to the user contributes to avoiding that pervasive communications from the Smart Space become another source of information overload for the user.

The proposed architecture and Virtual Character based interaction paradigm have been validated in an experimental test-bed, where an AmI infrastructure is integrated in a limited area of our university which includes a laboratory open to students. The system aims at providing students with many functionalities, such as help in finding people and resources they do not know how to reach, as well as access to all the personalized services the system decides to tender them thanks to the knowledge of their personal profiles.

This chapter is organized as follows. The first section discusses several fundamental issues which influence the delivery of adaptive multimodal feedback to users of a Smart Space. Section 6.2 proposes the use of a biologically-inspired awareness model as a blueprint for the design of a flexible architecture for Ambient Intelligence systems, which is in turn presented in section 6.3. Section 6.4 discusses the core module of this architecture, the Decision Manager which is responsible for personalizing the interaction with the user, and controlling the Multimodal Communication module described in section 6.6. Section 6.7 describes the test-bed laboratory in which the current implementation of the proposed architecture is being tested, and discusses significant results and open issues. Finally, section 6.8 draws some conclusions and outlines directions for future work.

6.2 Main Issues

The design of a comprehensive Ambient Intelligence architecture needs to take into account several heterogeneous issues. Essentially, the system needs to be able to perceive all the useful information to extract a correct description of the controlled environment.

This knowledge is the fundamental starting point to condition the behavior of the Smart Space and to adapt it to user needs. The aim of the system is to deliver the correct and personalized feedback to the user interacting with the intelligent environment. To effectively achieve this result the system has to collect a variety of data and obtain an instantaneous description of the context. This knowledge becomes the basis for adaptation; so one of the key point for the design of such a structure resides in the communication interface. Adaptive interaction appears to be a very important topic to make this kind of systems useful, effective, and also unobtrusive, so to make users accept them as something they are comfortable with [25].

In order to better outline the motivations and criteria which have influenced the design of the proposed architecture, it is useful to analyze more deeply these three aspects:

- Context awareness
- Multimodal communication
- User-centered adaptive interaction.

6.2.1 Context Awareness

The central idea of the user centered adaptation paradigm is based on the context information exploitation. But what do we mean with *context*?

To give a more formal definition, we can say contextual information can be defined as *an ordered multilevel set of declarative information* concerning events occurring both within the sensing domain of a Smart Space (SS) and within the communication and action domains of the SS itself. Relations among events are also included (explicitly or implicitly) within contextual information [39].

An *event* can be defined as the occurrence of some fact that can be perceived by or be communicated to the SS; an event is characterized by attributes that basically answer questions about *where* (position) and *when* (time) the event occurred. Other attributes involve *what* (core) consists the event of, *who* (identity) is involved in the event, and possibly *why* (reason) the event occurred [38].

Events can be used to represent any information that can characterize the situation of an interacting user as well as of a Smart Space component, i.e. an entity. An entity can be a person, a place, or an object that is relevant to the interaction between a user and an application. The user and the SS parts themselves are entities. The multilevel nature of contextual information is related to the possibility of detecting and representing events at multiple abstraction levels.

Context-awareness refers to the property of a SS to internally represent in terms of events the state of its users, of their surroundings and of SS parts, as well as to be provided with rules that make it possible to adapt its behaviour accordingly. Therefore, in a context aware SS, contextual information can be either used by itself, i.e. it can have its own value for the interacting user, or it can be used to select which services can be provided to the user and which interaction modalities are more suited to let him/her access such services.

In brief, an estimate of the user context can be used to optimize the adaptation and personalization process in such a way to maximize the service value.

Pascoe introduced a set of four context-aware capabilities that applications can support [6]:

- *Contextual sensing*: a system detects the context and simply presents it to the user, augmenting the user's sensory system
- *Contextual adaptation*: a system uses the context to adapt its behaviour instead of providing a uniform interface in all situations
- *Contextual resource discovery*: a system can locate and use resources which share part or all of its context
- *Contextual augmentation*: a system augments the environment with additional information, associating digital data with the current context.

This distinction results particularly useful in the design of a complex distributed systems characterized by a high-level analysis and interaction capabilities as the one described within the scope of this work. In the design of the proposed architecture we focus on *contextual sensing* and *contextual adaptation* capabilities, while the other two listed properties are not explicitly explored in this implementation.

6.2.2 Multimodal Communication

In a Smart Space, where users must be able to access services not only when they are sitting in front of a PC screen, but as they move freely within the environment and at the same time attend to other tasks, traditional user interaction paradigms commonly employed on desktop machines, based on Windows, Icons, Mouse and Pointer (*WIMP*), exploit only a small fraction of the user's communication and interaction capabilities [30].

With the goal of making the interface between the user and the Smart Space as transparent as possible (the *ideal interface is no interface* [34]), a promising approach is that of widening the range of communicative modalities that the system can use to address the user [25]. Our starting point is that *human-to-human* communication is an extremely rich and varied exchange of information, which exploits a multitude of channels and modalities [18].

A *channel* can be seen as a combination of effectory and sensory means that enable the transmission and reception of information: for instance, the human voice and hearing.

A *modality* can be defined as a specific way to use a channel, which determines the type of information transferred and the association of semantic meaning to the individual signals [15]. For instance, using voice and hearing to transfer natural language, or using facial muscles and sight to transfer expressions representing emotional states.

Communication modalities can convey either linguistic or para-linguistic information: for example, speech can contain words and phrases, but can also convey emotions. On the other hand, gestures typically represent non-linguistic information but can also be used for linguistic exchange through sign language.

Multi-modal communication can then be defined as *a simultaneous exchange of information over multiple channels at different levels of abstraction* [15].

Human-to-human communication is intrinsically multimodal: limiting man-machine interaction to keyboard/mouse input and text/graphical output means employing only a small fraction of the overall communication bandwidth that it is available to us [30]. The use of multi-modal interfaces aims at allowing the artificial system to engage in a similar dialog with the user, with the objective of exploiting the richness, robustness, and flexibility of face-to-face conversation.

Mehrabian outlined how in face-to-face conversation only a small fraction of the information is transferred in linguistic form [23]. The *richness* of multimodal dialog allows for conveying subtle nuances in meaning and meta-information about what is expressed in a linguistic form [16]. Multi-modal interfaces aim at detecting and reproducing these subtleties to more effectively present information to the user. As an example, the effectiveness of gestures and gaze for providing spatial information in a concise form is exploited by several multimodal systems [17][31].

The *robustness* of multimodal communication comes from the redundancy and correlation between modalities, which allow for disambiguating the meaning of received stimuli, such as in presence of noise, or when the listener is paying only partial attention to the speaker [29][15][28].

Robustness is also related to the intrinsic *flexibility* of multimodal communication: the possibility of dynamically switching modalities to best suit a given situation [32]. As an example, using gestures instead of speech in a noisy environment. This flexibility makes multimodal interfaces particularly suited to mobile or outdoor use [26].

While significant advances have been achieved in this field [28], the development of multimodal systems still faces significant challenges, particularly on the input/analysis side [29]. Consequently, in this work we will not consider the direct acquisition of multimodal input from the user, but focus on the use of multiple modalities in the system-to-user side of interaction.

Generally, multimodal interfaces do not necessarily need to be personified; on the other hand, to allow the machine to exploit most non-speech output modalities, at least some degree of embodiment is needed [25]. In our research, we consider a particular type of multimodal interfaces which uses computer-animated characters to enable a direct interaction with the user. We will use the term *Virtual Character* to refer to an animated character which is able to interact and communicate with the user by means of verbal and non-verbal modalities, but is not necessarily human-like, or realistic.

In a Smart Space, the character can be employed to act as a *Virtual Assistant* providing suggestions and feedback to users.

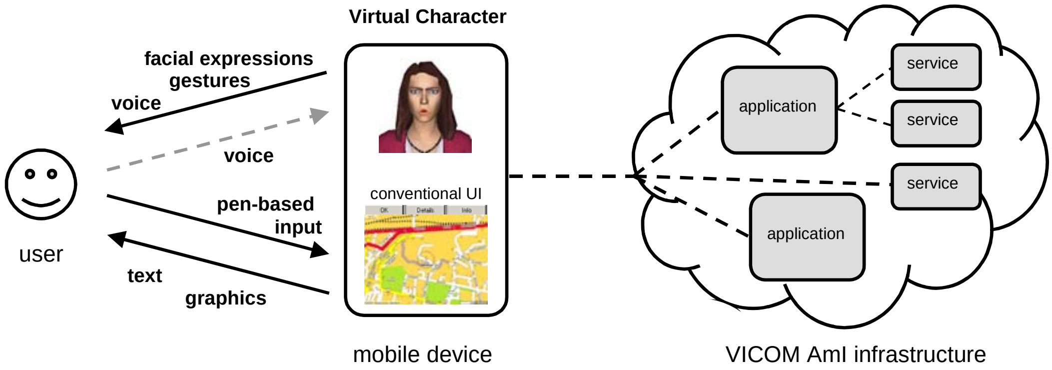

Figure 6.1 The considered Virtual Character based interaction paradigm: a Virtual Character, displayed on the user’s terminal, can act as a Virtual Assistant and provide a consistent front-end to the applications and services

The Virtual Character can address users by means of dynamically generated multimodal utterances including speech, synchronized lip movements, facial expressions and gestures.

This contributes to make the Smart Space interface more natural and intuitive: as an example, even a child can immediately detect and understand an expression or a disapproval gesture performed by a Virtual Character.

Two other positive effects of using of Virtual Characters to create personified interfaces is that they can solicit the cooperation between the user and the machine [20], building trust on the part of the user [27] and positively affecting memorization and learning [20].

6.2.3 User-Centered Adaptive Interaction

Both in our work activities and free time, we naturally tend to personalize the surrounding environment, in order to make it more suited to our skills, preferences and working style.

People prefer to interact with a personalized environment, rather than an anonymous and impersonal one [25]. This also applies to our digital working environment: everybody, using a PC has experimented the customizability of mainstream Graphical User Interfaces, where many parameters affecting the appearance and behavior of user interface elements such as I/O devices (pointing devices, keyboards as well as monitors) menus and desktop appearance. As soon as the user makes a frequent use of these systems, he feels the need to exploit this customization features.

The more Ambient Intelligence systems pervade our daily work and personal life, the more they need to take into account the distinctive characteristics and needs of users.

Designing an architecture which allows for a *user-centered* interaction, requires supporting a high degree of flexibility and adaptation. In user-centered paradigm it is the system that tries to meet the user personal interaction style and not vice-versa as it often happens. This adaptation has to be applied to a wide range of the system's functionalities, including:

- User input collection
- Decision-making
- Message content selection
- Multimodal message presentation
- Message delivery.

In this work, we focus on the *system-to-user* side of interaction, and specifically on the issues related to the last tree items. Overall, the fundamental objective of message customization is to maximize the message usefulness, in terms of its effectiveness in directing user behavior in a constructive way. Usefulness may be defined in terms of a number of objectives, such as maximizing ambience safety and security, or minimizing waste of time for users.

Another reason to adapt a message to user's expectations can simply be the will to make the system itself more acceptable for the user: the growing complexity of such technologies must be hidden beneath the easiest interaction possible. If the interface is tailored to the needs of a given user, the more he/she will be encouraged to accept and use it.

In the choice of message content, the system must also be careful not to overload the user's attention, especially if it understands the user is busy with another task. To maximize the effectiveness of the message, also the presentation modalities (text, graphics, speech, facial expressions of the Virtual Character) must be chosen so that user can rapidly and correctly understand them. For instance, sending a voice message to an user who is in a noisy room, or simply when the message content is private would be inappropriate.

Conversely, in an abnormal situation, the user is likely to perceive more immediately the danger from a surprised and worried expression of a Virtual Assistant than from a long textual message.

Additionally, the system should make it easy for the user to remember and reuse the information if needed. As an example using voice messaging to communicate a phone number or an address can be a waste of time because the user will probably need to recall it later.

Finally, both the choice of presentation modalities and of their encoding in a specific form must be adapted to the characteristics of the terminal on which the message will be displayed, so that its effectiveness is not limited by a degraded reproduction. As an example, if the user terminal is able to display only a low resolution map of the path to the desired destination, a textual description of such path should be used. The need for terminal adaptation clearly appears by considering the main characteristics of user terminals that are being considered in the VICOM project (Table 6.1.).

The approach that has been chosen in this work is that of exploiting the robustness and flexibility deriving from the possibility of conveying the same information through different modalities, and particularly though the verbal and non-verbal behavior of the Virtual Character.

Table 6.1 In order to effectively deliver an informative message to the user, it is necessary not only to choose the right communication modality, but also to take into account the characteristics of the terminal he/she is interacting with.

	2G Phone	**2.5/3G Phone**	**PDA**	**Multimedia PC**
Display	100x100	240x160	320x240	1280x1024
Color depth	b/w	4K/65K	65K	16M
Graphic capabilities				
• 2D graphics	x	x	x	x
• 2D animation		x	x	x
• 3D graphics		x	x	x
• 3D animation		varying	simple	complex
Video reproduction		x	x	x
Audio reproduction		mono	x	stereo
Available storage	<64k	2-16MB	32-128MB	several GB
Connection type	GSM	GPRS	WLAN	Fast Ethernet
Connection bandwidth	9.6k	38.4k	10Mbit	100Mbit
Round trip delay	~2s	~100ms	<330ms	<50ms

To be able to consistently manage the acquisition of context information from the environment, and adaptively generate and deliver multimodal feedback to users, an Ambient Intelligence system must be designed around a well-defined and flexible architecture. The reference architecture of the presented work is described in the following sections.

6.3 Inspiring Awareness Model

In order to validate the design of a general structure to manage the main issues related to a Smart Space, our approach is the one to take inspiration from a theoretical human conscience model. The wanted inspiration can be found in the work of Antonio Damasio [14] where a model for brain and conscious reasoning is reported and motivated on neuro-physiological bases. This approach results particularly useful in the design of a complex distributed systems characterized by high-level analysis and interaction capabilities.

Systems of this kind need to be *aware* of the behaviour of all objects acting in their scope of sensing as well as of its own components' status and reactions. In this sense awareness of the context in which the system act is not enough, the system has also in some way to be *conscious* of its own internal state.

In this section we thus introduce in details the cited model in order to use it as the root for our architecture design.

As a first definition, consciousness pertains to the knowledge of any object or action attributed to a self. The search for the "self" starts from the following question: "how do we ever know that we are seeing a given object?". The answer can be formulated in terms of two key players, the *organism* and the *object* and in terms of the *relationships* those players hold in the course of their natural interactions. Here the organism in question is the AmI system whereas the object is any entity that gets to be known by the system; the relationships between organism and object are the contents of the knowledge we call consciousness. Seen in this perspective, consciousness consists of constructing knowledge about two facts:

- The organism is involved in relating to some object
- The object in the relation causes a change in the organism [14].

While the human perceiving system changes dutifully at the mercy of the objects it interacts with, a number of brain regions whose job is to regulate the life process do not change at all in terms of the kind of object they represent. The degree of change occurring in the object –the body– is quite small. This is because only a narrow range of body *states* is compatible with life and the organism is genetically designed to maintain that narrow range and equipped to seek it. This fact leads to the observation that some parts of the brain are free to map whatever object the organism's design permits to map. On the other hand, other parts of the brain, those that represent the organism's own state, are not free to change at all. The body's internal state must be relatively stable by comparison to the environment surrounding it. So the deep roots for the 'self' are to be found in the ensemble of brain devices that *continuously* and *non-consciously* maintain the body state within the narrow range and relative stability required for survival. Following the Damasio model, the state of activity within the ensemble of such devices is defined *Proto-Self*, whereas the "non-conscious" forerunner for the levels of self which appears in our minds as the conscious protagonists of "consciousness" represents the *Core Self* and *Autobiographical Self.* These three components are the main parts concurring to form the human consciousness :

- The *Proto-Self* is an interconnected and temporarily coherent collection of neural patterns which represent the state of the organism, moment by moment, at multiple levels of the brain. We are *not* conscious of the Proto-Self
- The *Autobiographical Self* is a conscious part of self, based on autobiographical memory which is constituted by implicit memories of multiple instances of individual experience of the past and of the anticipated future. This memory grows continuously with life experience but can be partly remodelled to reflect new experiences
- The *Core-Self:* is generated for any object that provokes the core-consciousness mechanism that is to say the continuous environment awareness due to the analysis of stimulating objects. Because of the permanent availability of provoking objects, it is continuously generated and thus appears continuous in time. The mechanism of Core-Self requires the presence of Proto-Self. Core-Self can be triggered by any

object. The mechanism of production of Core-Self undergoes minimal across a lifetime. We are conscious of the Core-Self.

In Figure 6.2. Damasio's model is depicted, as it can be seen Proto-Self constitutes the basis for the Core-Self whereas the Autobiographical Self has got inferences coming from the Core-Self and the Autobiographical Memory. This means the Autobiographical Self, that can be seen as the short term memory, is continuously composed by recalled elements coming from the long term memory (Autobiographical Memory).

6.4 Mapping the Model into an AmI Architecture

As previously stated our aim is to try and exploit the abstract concepts proposed by Damasio and translate them in the architecture of an artificial system able to be aware of the context it is working in. Proto-Self functionality can be realized through a system's *internal state context analysis*: the non-conscious human monitoring on the inner parts of the organism can be translated into the use of internal sensors. In both cases the aim to control critical variables and to analyse the relationship between environmental (external) events and internal state conditions.

Core-Self is instead represented in its artificial intelligence counterpart, as the *physical context representation*, namely the collection and association of the continuous observation data coming from the senses (for the human brain) or the external sensors (for what concerns a context awareness based system).

Autobiographical Self, based on the use of Autobiographical Memory represents the conscious part of self. The two concepts are strictly connected and can be seen as the short-term memory and the long-term memory respectively in the model of the human brain.

They both constitute the idea of growing *experience* that can be implemented with associative techniques working on a knowledge base.

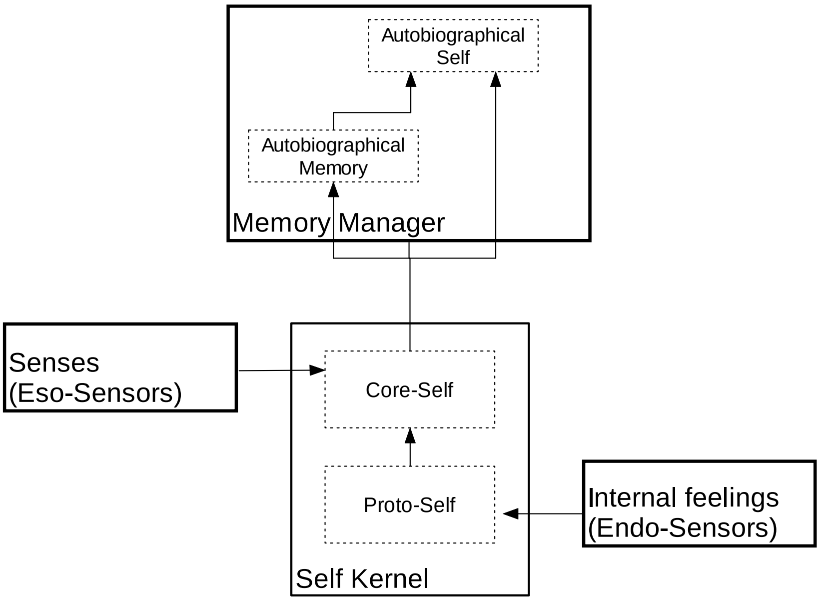

Figure 6.2 Damasio's Human Conscience Model. The described components are depicted in the diagram along with their connections. The Proto-Self and Core-Self are linked to senses and to internal organs by means of nerves connections and electric messages

6.4.1 Artificial Senses: Endo-Receptors and Eso-Receptors

Sensors own a key role in the definition and in the input of both Proto-self and Core-self: in an intelligent system they represent the bridge between the 'brain', namely the intelligent core of the organism, and the world, being it internal or external reality. Sensors (or receptors) are used by the system to keep contact to the interesting data and variables of the working environment (*External World*) as well as of its own internal state (*Internal World*). Considering this kind of distinction we can classify these devices or software agents into two groups that do not differ for technological aspects but for the aim of their observation, that is to say their observed domain. This means we can distinguish between *Endo-receptors* and *Eso-receptors* and associate them respectively to Proto-self and Core-Self.

In particular, as suggested by their name, Endo-receptors are devoted to the observation of the internal state of the system: that is to say devices fit for analysing internal components or variables proper of devices making part of the whole organism (system).

Instances of sensors belonging to this class are: computational units (i.e.: Desktop PCs present in the environment available for users), devices' status sensors, thermal sensors, safety-oriented sensors (smoke, gas, fire, water infiltration, etc.), lighting sensors and so on. Endo-receptors are what is needed to realize the concept of the *Proto-Self*, as described in the introductory part.

With *Eso-receptors* we refer to all the devices used by the structure to keep track of the events occurring in the observed domain and to collect data about the target of the analysis, being humans or other external interacting objects. Eso-receptors are the counterpart of the human senses, in this category fall sensors such as video sensors (working in visible or infra-red wave length fields), Global Positioning Systems (GPS), standard or directional microphones, weight sensors, fingerprints readers, electro-magnetic waves emission scanners, photoelectric cells, etc.

6.4.2 Proposed Logical Architecture

The complete structure of an AmI system here proposed (Figure 6.3.) is defined in terms of interconnected logical modules. Each module implements a particular functionality such as *sensing* (i.e.: Endo/Eso receptors), *analysing* (i.e.: Context Manager and Memory Manager), *deciding* (i.e.: Decision Manager) and *acting* (i.e.: Actuators and Multimode Communicator).

In this sense, the system steps through the aforementioned functionalities with inferences on an Internal World (i.e.: the "artificial organism", the system itself) and on an External World (i.e.: the "objects", users). The latter environment represents those elements not under the direct control of the system. This means the system is aware of the objects interacting in its domain, but it cannot apply a direct inference on them. For example, in the realized test-bed, we apply the interaction towards the external world (the users) by adapting message communication.

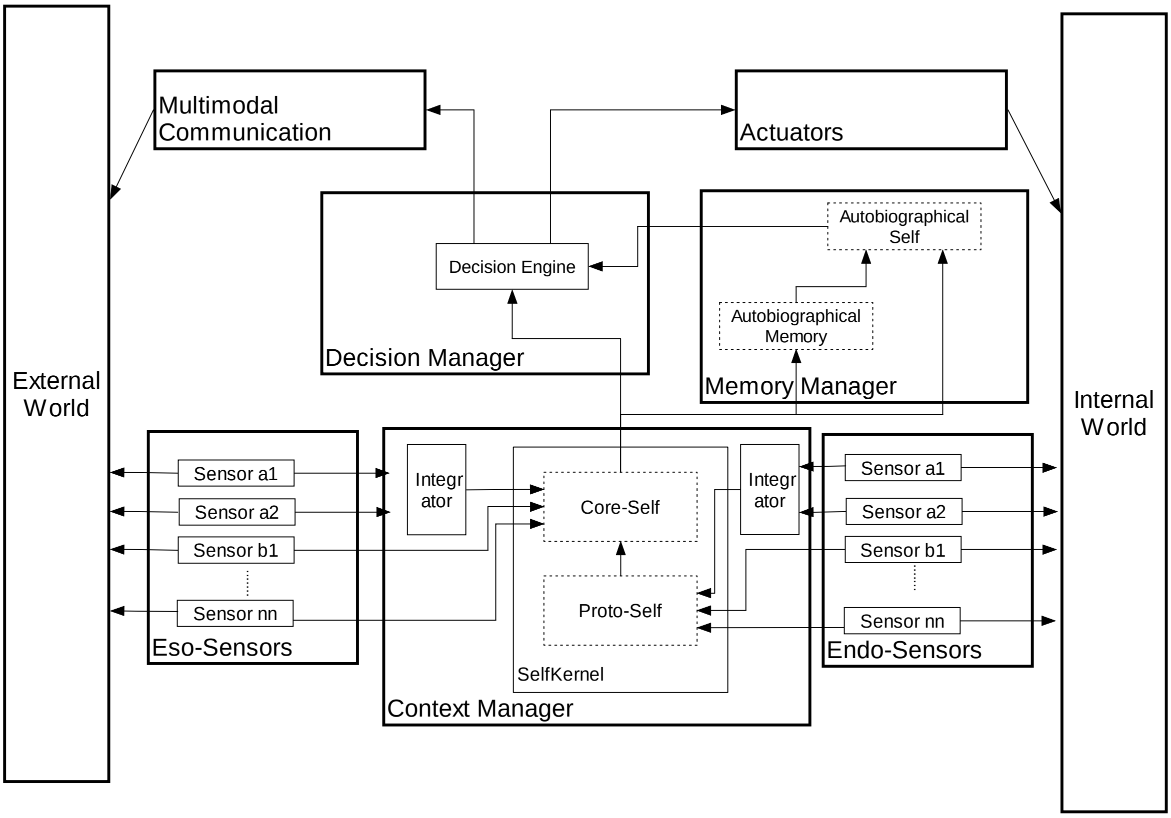

Figure 6.3 The proposed AmI logical architecture. Internal and External World represent the system itself and the objects and users interacting with the Smart Space, respectively. System adaptivity is explicitly expressed in the Decision Manager and in the Multimodal Communication modules

The only bridge keeping a connection between the intelligence of the system and the external world is expressed, as previously stated, by the pro-active functionalities it is provided with. Referring to the model, this is what represents the source of the continuous stimuli received by the Core-self. On the other hand, the Internal World is every thing the system can directly control: physical devices, internal software parameters but also physical components the system can apply its inference on.

As it can be seen in Figure 6.3. the core part of the architecture that will be further described in next paragraph, takes origin from the *Self Kernel* sub module introduced in the description of Damasio's model. In particular Proto-Self and Core-Self are encapsulated in a Context Manager (CM) devoted to the analysis of heterogeneous data and to the generation of contextual information.

The ensemble of internal and external state information is encoded by the Context Manager, a hierarchical representation where a *super-state* (e.g. ARRIVAL, meaning a new moving object entering the scene) gives the highest-level summarization of the context, while lower-level attributes describe the position and behavior of individual people, objects and resources [7]. The reference hierarchy is summarized in Table 6.2.

Table 6.2 Hierarchical representation of the context

Level		Description
1	*Event level*	super-states
2	*Behaviour level*	trajectories of users
3	*Object level*	identity of users
4	*Feature level*	number and position of users, system components status
5	*Signal level*	raw, non processed data

In such a structure, lower levels contribute to the estimation of the higher ones, a super state (an *event*) can be evaluated by considering objects and related features (i.e., the object's position). Event and Behaviour levels define the context in which objects and users act, whereas the remaining levels contribute to the definition of more specific and personalized description for objects.

Autobiographical Self with *Autobiographical Memory*, respectively embody system's long and short-term memory; they constitute the Memory Manager module (MM) devoted to the storage under the form of symbolic metadata of events produced by Context Manager. In addition it forms a Knowledge–Base for higher level modules. In an artificial system this concepts represents a static memory in which all the information needed to take decisions are stored and a buffer keeping trace of the instantaneous variables used to manage data processing.

The *Decision Manager* depicted in the diagram uses the context awareness represented by the Self Kernel output to make the system reacting and have its influence on the AmI domain, as either internal or external world. There are many possible approaches to the implementation of this module; our current implementation employs a rule-based system described in next section. To provide its results, the decision logic also relies on the direct observations of the receptors and on the past data provided by the Memory Manager. The decision can affect the Internal and External Worlds in different ways:

- Internal World: directly through some physical or software *Actuator Modules* (i.e.: *Actuators* in the diagram) such as mechanical tools which close or open windows and doors, as well as thermostatic devices affecting the rooms' temperature or in general controlling and modifying parts of the system
- External World: through *Multimodal Communication* channels: the decision concerns the choice of the proper information to be delivered and the choice of the best communication channel to address the user.

In the scope of this work we further describe in the following sections the problems related to the Decision Manager and the Multimodal Communication modules, particularly dealing with message communication and interfaces issues while the Actuators part is not here explored.

6.5 The Decision Manager Module

In the proposed architecture, the Decision Manager module is mainly responsible for determining which actions or communications must be performed in response to ambience events or, possibly, to direct user requests. Beside this, it is also responsible for exploiting knowledge about the current context to adapt its response to the user, and, together with the Multimodal Communication module, to his/her terminal.

Initially, we encountered remarkable difficulties in implementing adaptive behaviors into an application logic written in procedural programming languages. In a further refinement of our architecture, we decided to follow an alternative approach and structure the Decision Manager around a *rule engine*, to allow for a declarative definition of the system's behavior and particularly of the context-aware adaptation logic.

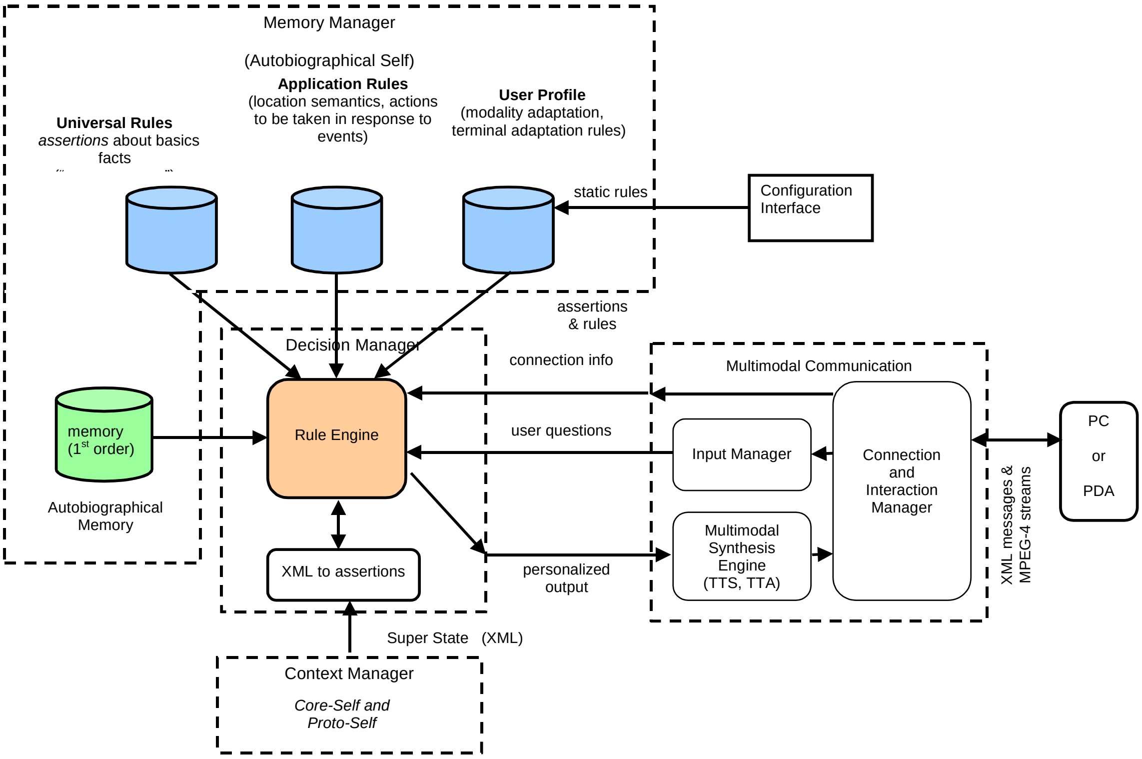

Figure 6.4 The internal structure of Decision Manager is based on a rule engine which exploits a wide set of heuristic rules to react to detected events and user queries, and to adapt to context attribute.

While, with respect to more sophisticated Artificial Intelligence techniques, the “reasoning” capabilities of a rule engine are limited to applying simple heuristics to a dynamic set of *facts* the engine knows about, this approach has two main advantages. On one hand, efficient implementations of rule engines, which can effectively be integrated with conventional object-oriented programming languages, such as Java, exist. In our current implementation, the *Java Expert System Shell* (JESS) has been used [9]. JESS allows for the specification of the rules in a high-level format and employs an efficient algorithm (called RETE) for continuously applying an high number of rules to input data.

On the other hand, rule-based systems such as JESS allow for expressing those heuristics in a powerful notation.

Figure 6.4. outlines the resulting internal structure of the Decision Manager. The rule engine continuously applies its knowledge to the events asserted by the Context Manager and the feedback provided by the Communication module’ if the result is a specific message for the user, a text-production function is activated, which, starting from a generic template for that message, uses adaptation rules to derive a textual representation of the personalized message for the specific user to which the message is addressed.

The rule engine receives two kinds of inputs. The Context Manager provides a high level hierarchical representation of the system state and of its understanding/estimate of the ambience context (see section 6.5), in XML format. As the hierarchical description is periodically updated by the Context Manager, the state tree is converted into a set of *assertions* of facts by the rule engine, which express both the values of context attributes and the relationships between them.

The second type of input is represented by direct user queries which are collected by the user terminal (PC or PDA) and received by the Multimodal Communication module. In the current implementation, these queries take the form of the selection from a dynamic menu of possible requests (e.g. *where am I?*, or *which PC can I use in this lab?*). In future releases of the system, these requests will also be collected by means of a speech recognizer. Again, those inputs are converted into the assertion of a fact (e.g, *(user-query where-am-I)*).

Since we are currently focusing on the communication with the user, the output of the Decision Manager is a complete specification of multimodal feedback to be provided to the user. In the future, the output of the Decision Manager will also include a set of commands to be forwarded to Actuator Modules.

The output specification includes the text of the message, the list of modalities in which the Virtual Character must deliver the message (text, voice, facial expressions, gestures).

The message personalization is achieved by expressing the message in the Virtual Human Markup Language[11]. VHML is an XML-based language, which allows for fully characterizing both the verbal and non-verbal behavior of a Virtual Character. In this way, not only the words spoken by the Virtual Character, but also his expressions and gestures can be varied depending on the personalized conditions detected in the left part of the rule.

The intensity and duration of the character expressions can also be varied by means of VHML tags: for instance, in the case of the system disappointment for the user performing a dangerous action, the intensity of the worried expression can be increased each time the message is repeated.

The rules, which in fact form the “experience” of the AmI system, are stored in a knowledge base maintained by the Memory Manager, and represent the Autobiografical Self, according to the inspiring model. The Memory Manager is also responsible for remembering past descriptions of the system and ambience context, which are stored in the artificial Autobiographical Memory (the short-term memory of the system). In a first, simplified approach to dialog and interaction management, the Autobiographical Memory stores the previous state of the system to allow for the detection of simple behavioral patterns in users.

In order to simplify the customization of the rule set to a specific ambience to be augmented, and increase rule reuse, the Knowledge Base is divided into three different rule sets.

- A first database contains *universal* rules expressing facts that hold in any environment, such as mappings between timestamps and time ranges (morning/afternoon), or spatial relationships (wether an object is near/far from another one starting from 3D coordinates), that allow other rules to be specified at an higher level of abstraction (e.g., *if isMorning()* instead of *if timestamp()<13.00.00*)
- A second database contains *application specific* rules, which describe the services to be provided to various classes of users in a specific location, together with lower level knowledge such as an association of location names to 3D coordinates and the description of the resources present at each location
- Finally, an *user and terminal profile* rule base contains a description of user groups and their permissions to access to information, together with terminal classes and their capabilities.

6.5.1 Context-based Message Adaptation

Within the rule engine, these two flows of input data are processed by means of a set of rules that define message selection and personalization strategies. Acquired contextual knowledge about the environment and the user is represented in terms of concepts (*facts*) and the responses to those events in terms of *rules.* A rule is activated when one or more specific *patterns* (the so-called left part of the rule) are detected in the asserted facts. In turn, this can trigger the assertion of new facts (making for a limited but useful *forward* *reasoning* mechanism) or the call of *functions*, which allow for interfacing rules with external code/modules (Figure 6.5).

Taking into account the context attributes provided by the Context Manager, we identified 7 main attributes of the context tree according to which the message can be personalized [22]: time, location, situation of a room, resource status, other users' activities, user behavior (cooperative/non-cooperative), number of times the same situation is repeatedly detected. Context-based adaptation is achieved by inserting these attributes in the activation patterns on the left part of the rule (Figure 6.5, right).

```
( rule-name
    ;left hand side
    (pattern 1)

        …

    (pattern N)
=>
    ;right hand side
    (function call)




    (assert fact 1)
        …
    (assert fact N)
)
```

```
( welcome-student-rule
    ;left hand side
    (context-event ARRIVAL
        (user-name ?userName)
        ;other context attributes  )
    (lab-is-open)
    (workstations-available false)
=>
    ;right hand side
    (send-message
        ?username
         (modalities Virtual Character speech facial expressions)
         ("Welcome to the lab" ?username "<sorry>There are no
workstation available</sorry>. Please come back later")
    (assert user-welcomed ?userName)
        …
        …
)
```

Figure 6.5 The general structure of a rule (left), and an example rule in simplified JESS notation (right)

6.5.2 Adaptation to the User's Terminal

The adaptation of the messages to the user terminal is managed by defining several profiles (represented as set of JESS rules), one for each terminal class available. The modality selection rules, however, are not univocally defined by the available terminal, but can take into account also user preferences which can be inserted off-line (a deaf user might prefer a text-based message).

We verified that the terminal-based modality selection rules cannot be completely decoupled from the context-based selection rules, e.g. the user activity (a supervisor walking around the campus might prefer a voice message over a text message, while a student attending a class might prefer avoiding voice messages).

In case context adaptation and terminal adaptation rules suggest different results, conflicts can be solved by exploiting a mechanism specific of the JESS rule engine which allows for the definition of the priority (the so-called *salience*) of a rule.

6.6 The Multimodal Communication Module

The Multimodal Communication module works on two levels. At the lower level it needs to detect the presence of user terminals within the coverage of the Smart Space network, to receive and maintain reliable connections with them. Since the Smart Space needs to be able to communicate with its users pervasively, independently of where they are within its area and whether or not they are provided by a smart terminal, the Multimodal Communication module is able to reach either by wired and wireless networks, or by means of public displays and speakers.

At the higher level, the Multimodal Communication module is responsible for dynamically synthesizing multimodal messages and collect explicit inputs from the users.

Starting from the high level representation of the Virtual Character behavior, generated by the Decision Manager the Multimodal Communication module either selects, in the simplest cases, one from a set of predefined messages, or in more complex cases it is able to dynamically generate the multimode feedback that is then presented by the Virtual Character on the user's device.

The Multimodal Communication functionality is typically implemented in a distributed way. On the server-side, the real-time generation of multimodal output is performed by a *Multimodal Synthesis Engine* [22]. The MSE is composed of several tools, each of which generates a specific communication modality. A Text-to-Speech engine synthesizes the character's voice, while a Text-to-Animation engine generates the character's facial movements, including both lip movements corresponding to speech, and expressions that reflect the system classification of an event (e.g., happiness, worry, surprise). The animation is encoded in a high-level parameter stream, which is sent together with the voice stream and other conventional modalities (text, graphics) to the user's device.

On the client-side, an Animation Engine running on the user's terminal synchronously renders the message by animating and displaying a specific character model. Currently, the client module has been ported to two main platforms: Windows- based PCs and handheld devices running Microsoft Windows Mobile 2003 (Figure 6.6). On desktop workstations and kiosk-like public terminals, the multimodal interface is based on embedding the animation player on a web browser, so that it can complement the information directly presented by character's voice and facial expressions with HTML-based text, graphics and forms. In order to adapt the multimodal presentation to terminal capabilities, a simplified Virtual Character model is used on the PDA.

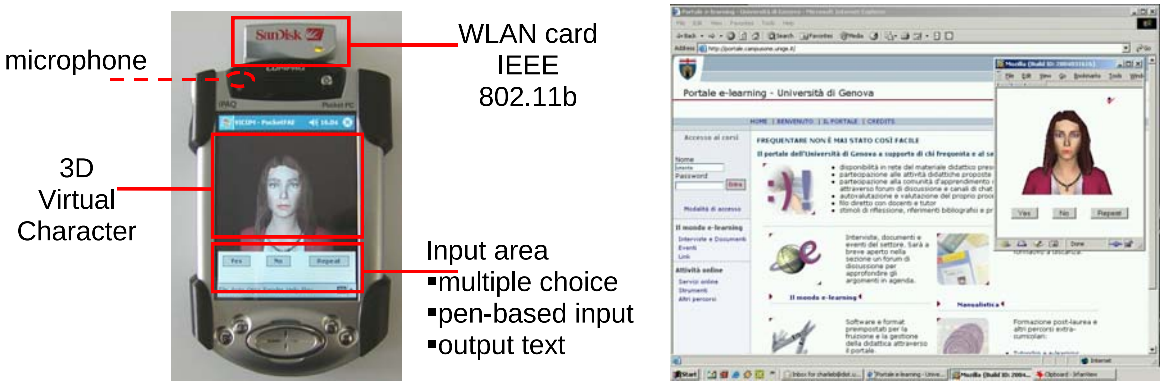

Figure 6.6 With the goal of providing a pervasive multi modal interface to users, the client part of the Multimodal Communication module, including the animation player, has been made available on both PDA and PC devices. The figure shows the multimodal interface running on two example terminals, an iPaq 3970 PDA equipped with a 400MHz RISC CPU, 320x240, 65k color screen, and IEEE 802.11b wireless LAN card (left), and a Microsoft Windows-based workstation (right)

6.7 Experimental Test-Bed and Results

In order to validate the concepts introduced by the proposed general architecture, we chose to realize a test-bed in a sub-area of university campus. The first step towards the implementation of a large scale Intelligent Campus has been turning a laboratory into a

Smart Space able to monitor its internal state, including its resources and users acting inside the lab.

Referring to the VICOM scenario presented in the introduction, the student can be informed about a particular event of interest (i.e.: a class he is going to attend has been relocated to a different room) in the most direct modality in relation to his/her location and state. If the student is working at his Desktop PC an animated Virtual Character can appear on the screen interrupting his/her current tasks and playing an alarm message, whereas if the student is walking in a corridor but is equipped with a mobile phone, the system can page him with a voice-only message.

Progressively, the system being developed will be extended with the integration of new services.

6.7.1 Physical Architecture

A set of two cameras has been installed to ensure a complete coverage of the room. Several *state*-sensors, including both hardware devices (such as a badge reader at the entrance of the room) and software agents running on the lab PCs (to detect user activities such as logins/logouts, and to measure CPU or network load) were also connected to the analysis module.

Operatively, this theoretical model has been implemented in a system following the logic-functional architecture outlined in Figure 6.7. different computational units (PC 1-2-3-4 and TERMINALs) have been used to support SS functionalities. A network infrastructure, based on wired LAN and an 802.11 WLAN, connects the different physical units and allows real-time data exchange.

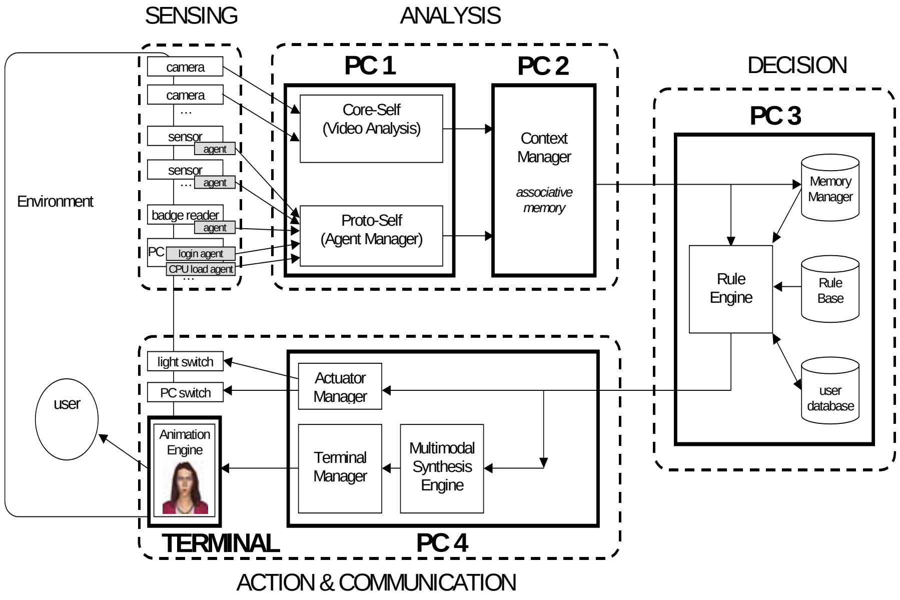

Figure 6.7 In the physical architecture of the test-bed, the Smart Space functionalities are distributed among a set of machines and terminals.

In particular tasks are allocated in 5 different computational units:

- **PC 1**: Core-Self (based on video analysis tools) and Proto-Self (based on a set of software agents running on the lab workstations) implement the Sensing capabilities of the Smart Space
- **PC 2**: Context Management is completed by Information Conversion and Data Fusion modules

- **PC 3**: hosts the rule-based Decision Manager and the Memory Manager
- **PC 4**: Multimodal Communication server that manages the generation of the multimodal output and the interaction with the mobile terminals
- **TERMINAL**: Personal Digital Assistant or PC maintaining the communication to the SS infrastructure and displaying the audio/visual feedback to the user.

This configuration, reduced with respect to the high development possibilities offered by the defined architecture, is a good example to demonstrate the interactivity and adaptativity features a Smart Space needs to present. The implemented system is in a continuous growing phase but this first realization has the power to demonstrate the theoretical concepts hidden behind the definition of the general architecture. The system is in fact able to perceive several internal and external status variables and to find relations among them, defining the events connected to the current context. The usefulness of this contextual information is then expressed in the influence on the adaptive communication tasks.

6.7.2 Definition of the Rule-Base

In the current testing phase, the Decision Manager has been programmed with a simple but varied set of rules targeted at verifying the adaptive interaction and context-awareness capabilities of the system. The set of corresponding multimodal message templates is defined so that the Virtual Character can act as an assistant to the human tutor who normally supervises student activities. In this line, the functionalities offered by the intelligent laboratory aim at supporting students in accessing lab resources and at providing the human supervisors with warnings on abnormal situations within the lab by reaching them even when they temporarily move to the surrounding offices.

The development of the rule base clearly outlined the need for expressing the adaptation rules at the highest possible level of abstraction, possibly by means of the introduction of multi-level rules that process events at increasingly higher level of abstraction.

In order to generalize the definition of adaptation rules, the informative messages to be provided by the Smart Space have been grouped in categories (Table 6.3.). In this way, it was possible to achieve a more effective and reusable definition of adaptation rules.

Table 6.3 Feedback messages to be provided to students and supervisors have been grouped in 7 main categories

Categories	**Examples**
Welcome messages	You are/are not authorized to enter the lab
Report about the situation in the lab	Number of people present in the lab
Availability of resources in the lab	Number of available workstations
Advertisement of available services	Workstation booking, printing facilities
Guidance information	How to reach teacher offices
Danger warnings	
Miscellaneous	

In the first place, the message must be varied according to the role of the target user within the Smart Space (e.g., student or supervisor), as configured *a priori* by a system administrator. Supervisors must be able to get access to any information produced by the surveillance system (e.g., the number of people present in those rooms), including information on the behaviors of other users (e.g., being notified when an user enters the lab after hours).

Additionally, an orthogonal categorization of user has been adopted, between *collaborative users*, who are equipped with multimodal communication devices already registered and enabled to get its services, and *non-collaborative or guest users*, which are not equipped with, or owns unknown devices.

In the definition of context-awareness rules, we followed a set of common guidelines to give the system a consistent behavior.

In the first point, the system must be able to react differently according to the current *time*.

For instance during work hours, guidance information could be provided to all users.

After-hours, users could only be able to get guidance to emergency exits.

Concerning the *user's location*, both at high level (building/room he is located) and at low-level (proximity to a resource, door, and so on), a general criteria is that of giving higher priority to personalization in term of lower-level location features, which means that "local" rules related to a specific place override default rules declared at the room or building level.

The most important criteria related to the monitoring of *room situation* is the distinction between normal conditions and danger/emergency condition. For example in an emergency condition, all messages not relevant to the user safety will be hidden not to distract the user's attention.

For what concerns the awareness of *available resources*, the system should try to maximize first the user productivity, then the resource usage efficiency. A key criterion is obviously the one of minimizing interference with *other user's activities*, particularly in the choice of presentation modalities. This means avoiding voice-based messages or talking-messages when the user is in a room where several people are working.

Repetition of the same event, then, is used to increase the weight of an action. This concept can be connected to the following last criterion: providing a *proportional feedback* to users, especially in cases when the system must outline an incorrect or forbidden behavior on behalf of the user. The feedback must be, on one hand, proportional to the severity of the user's misbehavior with respect to his own safety and to the overall system security. On the other hand, the intensity of the message needs to be progressively scaled according to the user's reaction to previous suggestions. This can be effectively achieved using non-verbal modalities, and particularly facial expressions of the Virtual Character: the first time the user contradicts the character, he can just show some surprise for the user not following, for instance, the suggested path. The following times the Virtual Character can show an increasing degree of angriness.

6.7.3 Example Interactions

Table 6.4. summarizes three instances of message adaptation. In the first case, the system detects the arrival of a student in the lab, and, according to the time of arrival and internal situation of the lab either welcomes the student or ask him to leave.

In the second case, an user ask whether there are workstations available in the lab; the answer delivery modality changes according to whether the user is standing at the lab entrance or approaching the lab from a nearby corridor covered by the Wireless LAN.

Finally, the third example shows how the system progressively varies the facial expression of the Virtual Character as the student repeats the same question more than once, and eventually redirects him to a human supervisor.

Table 6.4 Three examples of message adaptation in the considered Intelligent Laboratory scenario

Message description	**Context Attributes**	**Cases**	**Personalized message**	**Presentation modalities**
Welcome to the department	Superstate Time of the day	*ARRIVAL morning*	Welcome to the Biophysical and Electronics Engineering Department. Press the help button on your PDA if you need assistance	smiling and talking character
		ARRIVAL night	Welcome to the department. Please make sure that you have the after-hours access permit with you at all times	serious talking character
	Superstate	*Danger detected*	You must not enter the building as an emergency situation has been detected	text message
Number of PC available in the lab	User's location User's activity Resources	*User standing in front of the lab*	There are 2 PCs available, please log on workstation 3 on your left	text-message and map of the lab on a public display at the entrance of the lab
		User walking in a corridor	There are two PCs available, do you want me to reserve one for you?	voice message on the user's PDA
How to reach the lab	Number of times question is asked	*0*	Path to the lab	2D map
		1	I'm sorry I was not clear, please follow the corridor and turn right at the end	talking character with sorry expression
		2	I have alerted the lab supervisor, so that he can come to assist you.	talking character

6.7.4 Open Issues

The development of the test-bed and preliminary evaluations provided interesting feedback about the issues to be faced in the effective implementation and further refinement of the proposed Smart Space Architecture.

An open problem is represented by the consequences of errors in delicate tasks such as user position tracking or event recognition. These errors have a direct influence in triggering the system reaction but have to be considered unavoidable because of the nature of these tasks. Development in these fields works in the direction of a better reliability and better decision robustness. At the moment no error management is considered: if sensors fail, system's reactions may be wrong. The definition of metrics to manage errors should be useful to try and reduce effects of these problems.

Conversely, the definition of message selection criteria outlined that achieving a useful level of message personalization requires the adaptation to a number of cases that grows exponentially with the number of available communication modalities, user terminals, and, notably, with the increase in the complexity of the situation that can be estimated by the context-assessment module.

In a word, the more the Smart Space infrastructure is able to extract fine-grained information about the monitored environment, the more the system must be able to filter and adapt its feedback to the user. This means that in the Smart Space architecture should avoid using hard-wired message adaptation strategies, and exploit as much as possible automatic or semi-automatic tools for configuring, storing, managing modality selection and message selection/generation rules.

Finally, a limitation of Virtual Character based interfaces, that must be taken into account, is the fact that speech and most non-verbal modalities are non persistent [24].

While this is very useful to focus the user's attention on a specific element at a time, conventional text/graphics displays, where available, are anyway more efficient for presenting large amounts of information, or information that needs to be reviewed and modified by the user.

6.8 Conclusions and Future Work

The evolution of Ambient Intelligence systems is pursuing the dream of a digitally augmented environments where user seamlessly access information and perform tasks with the support of a variety of information appliances, such as PDAs, smart phones, Desktop PCs or public displays. In order to avoid that the user gets lost in the complexity of the available services and functionalities, however, this evolution must be supported by corresponding advances in the context-sensing and adaptation capabilities of the Smart Space, in the capability of interacting with the user by means of complementary communication modalities, and particularly in adapting the interaction to the user's characteristics and needs.

This chapter has presented the issues and rationale of the design of a flexible architecture for Ambient Intelligence systems able to adaptively interact with their users. A theoretic approach to the design of such an architecture has been described: a neurobiologically inspired human self consciousness model is the foundation for the artificial analysis and decision core, allowing the system to acquire and manage a deeper understanding of context information.

The proposed design employs a rule-based adaptation module where acquired contextual knowledge about the environment and the user is represented in terms of concepts and facts and is exploited to personalize the multimodal feedback for the user, which is displayed by a Virtual Character. The system exploits animated characters acting as Virtual Assistants, displayed on a range of different devices, to offer a consistent and pervasive interface and reduce learning and adaptation effort on the user side.

Within two national research projects, the architecture is being implemented to create an intelligent student laboratory. While the system has been tested in a small-scale environment, preliminary results confirm that the architecture allow for a high degree of personalization of the interaction between the user and the Smart Space.

However, in order to achieve the vision of an Ambient Intelligence infrastructure which *seamlessly* augment the activities which are possible in a real environment, making our interaction with the virtual environment no more cumbersome than that with the physical environment, further research is needed. We plan to continue our research along four main directions.

The perception tasks need to be powered with new sensors to provide a more complete picture of the instantaneous context. In the meantime the intelligence to extract useful information from these receptors must be enhanced to guarantee correct data to context awareness based decision tasks.

The scalability of the proposed architecture must be tested against significantly more complex environments.

The effectiveness of personalized interaction will be evaluated by collecting user's responses, and comparing two alternative scenarios: a setup where the Smart Space which provides only uniform textual/graphical feedback, and a setup where it provides personalized multimodal communications to the user.

Finally, a challenging task will be that of increasing the automatic adaptation capabilities of the system. While the proposed architecture supports the notion of learning by means of a dynamic Memory Management module, currently all the adaptation rules must be defined *a priori*. Future research will also concern the progressive introduction of simple learning capabilities taking advantage from the fact that the rule-engine allows for the dynamic definition of new rules at runtime [10].

6.9 Acknowledgments

This work was performed under co-financing of the Italian Ministry of University and Research (MIUR), within the project FIRB-VICOM [3] and the project PRIN PER2 [4].

References

[1] ISTAG, Scenarios for Ambient Intelligence in 2010. Online: http://www.cordis.lu/istag.htm.

[2] R. A. Brooks, with contributions from M. Coen, D. Dang, J. DeBonet, J. Kramer, T. Lozano-Perez, J. Mellor, P. Pook, C. Stauffer, L. Stein, M. Torrance and M. Wessler, The Intelligent Room Project**,** *Proceedings of the Second International Cognitive Technology Conference* (CT'97*)*, Aizu, Japan, August 1997.

[3] VICOM project, Virtual Immersive COMmunications, 2002-2005. Online: http://www.vicom-project.it.

[4] PER2 project, Distributed Systems for Multisensor Recognition with Augmented PERception for Ambient Security and PERsonalization, 2002-2004. Online: http://ginevra.dibe.unige.it/ISIP/Projects/miur02_en/main_en.html.

[5] T. E. Starner , Wearable Computing and Contextual Awareness, PhD Thesis, Program in Media Arts and Sciences, School of Architecture and Planning, MIT, June 1999.

[6] J. Pascoe, Adding Generic Contextual Capabilities to Wearable Computers, in *Proceedings of 2nd International Symposium on Wearable Computers,* October 1998, 92-99.

[7] L. Marchesotti, S. Piva, C.S. Regazzoni, Structured Context Analysis Techniques in a Biologically Inspired Ambient Intelligence System, *Proceedings of the IEEE on System Man and Cybernetics, Special Issue on Ambient Intelligence*, 2004 (to appear).

[8] C. Pelachaud, V. Carofiglio, B. De Carolis, F. de Rosis, I. Poggi, Embodied Contextual Agent in Information Delivering Application, Proceedings of AAMAS 2002, Bologna, July 2002.

[9] The Java Expert System Shell. Online: http://herzberg.ca.sandia.gov/jess/.

[10] E. Friedman-Hill, *Jess in Action: Java Rule-based Systems*, Mannings, July 2003.

[11] A. Marriott, S. Beard, J. Stallo, and Q. Huynh, VHML - Directing a Talking Head, *Proceedings of The Sixth International Computer Science Conference. Active Media Technology*, Hong Kong, 2001.

[12] The Virtual Human Markup Language. Online http://www.vhml.org/.

[13] R. Pockaj, M. Costa, C. Bonamico, and F. Lavagetto, Real-time MPEG-4 Facial Animation with 3D Scalable Meshes, In *Image Communication Journal*, **17**, (2002), 743–757.

[14] R. Damasio, The Feeling of What Happens-Body, *Emotion and the Making of Consciuousness*, Harvest Books, September 2000.

[15] Benoit, J. Martin, C. Pelachaud, L. Schomaker, and B. Suhm, Audio-Visual and Multimodal Speech Systems, in *Handbook of Standards and Resources for Spoken Language Systems*, D. Gibbon, Ed., 1998.

[16] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan, More Than Just a Pretty Face: Conversational Protocols and the Affordances of Embodiment, *Proceedings of Knowledge Based Systems*, 2001.

[17] J. Cassell, T. Stocky, T. Bickmore, Y. Gao, Y. Nakano, K. Ryokai, D. Tversky, C. Vaucelle, and H. Vilhjálmsson, MACK: Media lab Autonomous Conversational Kiosk, *Proceedings of Imagina '02*, Monte Carlo, 2002.

[18] B. De Carolis, C. Pelachaud, and I. Poggi, Verbal and Non Verbal Discourse Planning, *Proceedings of the Workshop on Achieving Human-Like Behavior in Interactive Animated Agents*, in conjunction with *The Fourth International Conference on Autonomous Agents*, Barcelone, Spain, 3 June, 2000.

[19] D. Garlan, D. Siewiorek, A. Smailagic, and P. Steenkiste, Project Aura: Toward Distraction-Free Pervasive Computing, *IEEE Pervasive Computing*, April-June 2002.

[20] L. Gong, Towards a Theory of Social Intelligence for Interface Agents, *Proceedings of the Virtual Conversational Characters: Applications, Methods, and Research Challenges Workshop,* in conjunction with HF2002, Melbourne, Australia, 2002.

[21] M. W. Kadous and C. Sammut, Mobile Conversational Characters, *Proceedings of the Virtual Conversational Characters: Applications, Methods, and Research Challenges Workshop*, in conjunction with HF2002, Melbourne, Australia, 2002.

[22] L. Marchesotti, C. Bonamico, C. Regazzoni, and F. Lavagetto, Video Processing and Understanding Tools for Augmented Multisensor Perception and Mobile User Interaction in Smart Spaces, *International Journal of Image and Graphics*, 2004 (to appear).

[23] A. Mehrabian, Nonverbal Communication, Chicago, Aldine-Atherson, 1972.

[24] J. Nielsen, Voice Interfaces: Assessing the Potential - AlertBox 27, 2003. Online: http://useit.com

[25] A. Nijholt, Disappearing Computers, Social Actors and Embodied Agents, *Proceedings of 2003 International Conference on CYBERWORLDS*, Singapore, 2003, 128-134.

[26] G. Niklfeld, M. Pucher, R. Finan, and W. Eckhart, Steps Towards Multimodal Data Services in GPRS and in UMTS or WLAN Networks, *Proceedings of ISCA Tutorial and Research Workshop on Multimodal Dialogue in Mobile Environments IDS-02*, Irsee, Germany, 2002.

[27] J. Ostermann, E-Cogent: An Electronic Convincing aGENT, in I. S. Pandzic and R. Forchheimer, (Eds.) *MPEG-4 Facial Animation: The Standard, Implementation and Applications*, 2002.

[28] S. L. Oviatt, Mutual Disambiguation of Recognition Errors in a Multimodal Architecture, *Proceedings of the Conference on Human Factors in Computing Systems*, CHI’99, Pittsburgh, PA, 1999, 576–583.

[29] R. Raisamo, Multimodal Human-Computer Interaction: a Constructive and Empyrical Study, PhD Thesis, University of Tampere, Finland, 1999.

[30] B. Reeves and C. Nass, Perceptual Bandwidth, *Communications of the ACM,* **43** (3), (2000) 65-70.

[31] SmartKom project, 2002.Online. http://www.smartkom.org

[32] K. Thorisson, Face-to-Face Communication with Computer Agents, Working Notes, *AAAI Spring Symposium on Believable Agents*, 1994, 86-90.

[33] K. R. Thórisson, Computational Characteristics of Multimodal Dialogue, *AAAI Fall Symposyum Series on Embodied Language and Action*, 1995.

[34] A. Van Dam, Post-WIMP User Interfaces, *Communications of the ACM*, **40** (2), (1997) 63-67.

[35] F. Vatalaro, Dalle Telecomunicazioni alla Telepresenza Immersiva, Notiziario Tecnico Telecom Italia, anno 11, n°3, Dicembre 2002.

[36] M. Trivedi, K. Huang and I. Mikic, Intelligent Environments and Active Camera Networks, *IEEE Transactions on Systems Man and Cybernetics*, October 2000.

[37] L. Marchesotti, L. Marcenaro and C. Regazzoni, Heterogeneous Data Collection and Representation within a Distributed Smart Space Architecture, Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems), Ghent, Belgium, 2002.

[38] L. Marchesotti, G. Scotti and C. Regazzoni Issues in Multi-camera Dynamic Metadata Information Extraction and interpretation for Ambient Intelligence, Nato Asi 2003, NAREK center of Yerevan University, Tsakhkadzor, Armenia, 18-29 August 2003.

[39] J. L. Crowley, Context Aware Observation of Human Activities, Proceedings of the IEEE International Conference on Multimedia and Expo, ICME-2002, Lausanne, Aug 2002.