Interacting with Embodied Agents in Public Environments

Addolorata Cavalluzzi, Berardina De Carolis, Sebastiano Pizzutilo and Giovanni Cozzolongo
Intelligent Interfaces
Department of Informatics
University of Bari
{cavalluzzi, decarolis, pizzutilo, cozzolongo}@di.uniba.it

ABSTRACT
In this paper, we present the first results of a research aiming at developing an intelligent agent able to interact with users in public spaces through a touch screen or a personal device. The agent communication is adapted to the situation at both content and presentation levels, by generating an appropriate combination of verbal and non-verbal agent behaviours.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Interaction Techniques.

General Terms
Human Factors.

Keywords
Interface agents, personalization.

1. INTRODUCTION
Interaction with services provided by "active" environments may be transparent to the user, by envisioning a world of omnipresent but invisible information technology embedded into products and everyday items [15]. Alternatively, the environment may manifest itself by means of an intelligent virtual entity with which users may interact in a socially engaging conversation like with a human companion. Embodied Conversational Agents (ECAs) are software agents with a more or less realistic and 'human-like' embodiment, which exhibit many of the properties of humans in face-to-face conversation, including the ability to provide a natural interaction and enforce social involvement[6]. They provide a type of multimodal interface where the agent can dialogue with the user by employing verbal and non-verbal expressions showing human-like qualities such as personality and emotions [4,11].

In this view, the Agent is displayed in a public interaction space, which takes the form of kiosk, interactive screen or LCD projection. However, even if interacting with an ECA is more engaging and natural than with a touch based interface [1], this metaphor poses problems related to the nature of the interacting media itself (difficulty in perceiving voice output in a noisy environment) and related to the public nature of interaction (privacy issues, lack of user-related information needed for personalizing information presentation). In this case, as far as personalization is concerned, the agent may only provide information of general interest, by tailoring it to the common target audience which is expected in the space in which the service is located. To this aim, the agent may adapt its embodiment, communication style and personality to the role it is expected to play in that context and to the place culture [5]. If, on the contrary, the physical space is designed so as to 'sense' the particular users being in the environment and to 'know' their interests and preferences, the Agent can use this information to create more specifically targeted presentations. In this case, if supported by an appropriate technology, interaction may become personalized and private even in a public space. Users may move around this kind of spaces with their own personal device: when in proximity of an information access point, the agent is activated: user-related information may be transferred to it and, therefore, information provided may be adapted to that user, who can handle interaction through her/his device.

This paper describes our work in this direction. Our research was originally developed in the context of a EU Project (Magicster)[1] that focused on the development of an ECA which shows some of the properties that are typical of human face-to-face conversation. In particular, ability to provide a natural multimodal interaction, to adapt information presentation to the user and, when needed, to increase the sense of social relation with the user by showing an empathic attitude [4]. The idea of how the agent behaviour changes in a pervasive computing context was not considered in that Project and was the subject of a new research, whose first results are described in this paper. Since our approach is general enough to be applied in a technologically evolving context, we will focus our description on how the conversational aspects of interaction are modeled rather than on technical features. After a brief description of the agent architecture, we show an example of its application in which the agent acts as a travel agent. Conclusions are outlined in the last Section.

2. SYSTEM ARCHITECTURE
The functional architecture of our system (Figure 1) includes two main components: the Environment Agent (installed on a server) and the Personal Interface running on a PDA.

2.1 The Environment Agent
The Architecture of an ECA is related to the model of a Natural Language Generation (NLG) system [12]. Given a communicative goal to be achieved, the Agent plans the communication content and renders it according to the expressive capabilities of its "body" and to the conversational context. In this view, the Agent is seen as an entity including two main components (a '**Mind**' and

[1] Magicster: IST-99, contract n 29078

a '**Body'**) which are interfaced by a common I/O language, so as to overcome integration problems and to allow their independence and modularity. In a given phase of the dialog, the **Agent's Mind** decides what to communicate according to its mental state and to the goals which are active in that phase. The *Dialog Manager*, built on top of the TRINDI architecture [13], computes the dialog move according to the current situation that is described in the Agent Information State (IS). The IS stores i) the *environment model* (information about the application domain, type of interaction, objects or points of interest in the digital or physical world to which the Agent may refer), ii) the *agent model* (agent *beliefs*, *goals* and relations among them), iii) the *user model* and iv) the *dialog history*.

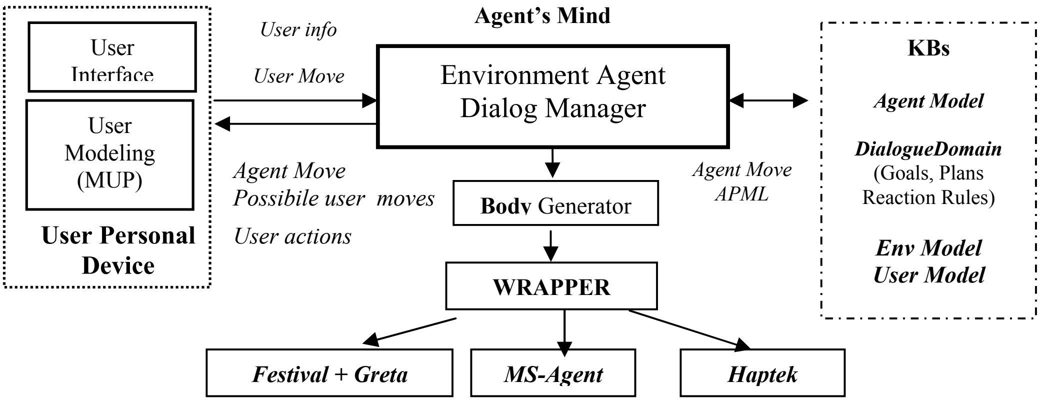

Figure 1: Functional components of our conversational system

According to these features, the Dialog Manager selects a set of goals to be achieved during the conversation. The agent knows how every goal may be achieved in a given context: it selects, from a library, the *plan* whose application conditions (user characteristics, dialog history, and so on) match to the context. The set of initial goals may change during the dialog. When a user move activates one or more 'urgent' goals, the goal priorities are revised: some of the active goals may become less 'urgent' and/or 'important' while inactive ones may become active. The plan that best suits to achieve the most urgent and important goal in the given context is then applied. Once the agent move has been planned, every communicative act is rendered with an XML string compliant to APML (Affective Presentation Markup Language: [3]). This string combines the text with tags which specify the 'meanings' that the Body will have to attach to each part of it. These meanings include the communicative functions that are typically used in human-human dialogs: emphasis, affective, meta-cognitive, performative, deictic, adjectival and belief relation functions. Once the APML string has been generated, the meanings expressed in it can be interpreted by the **Body Generator Module** that decides which combination of body signals (verbal and non verbal behaviours) to use to convey every meaning specified in the APML move. As mentioned previously, the agent embodiment can differ according to several factors like the type of environment, the role, the target audience (average age and cultural factors). We may couple the tagged agent move with different bodies, by using appropriate wrappers. So far, we developed APML-wrappers for Greta [9], Haptek [7] and MS-Agent technology [10].

2.2 The Personal Interface Component

This component enables the user to interact with the environment agent by means of a personal device. It includes two sub-components, aimed at modeling the user and providing an interface for exchanging messages (dialog moves) with the agent.

As far as user modelling is concerned, we adopt a mobile approach [15], in which the user information is stored and modeled on the personal device. When the user approaches one of the active points in which the environment agent is displayed, this component provides information about the user preferences (xml-annotated user info) that are relevant for that particular domain. The agent can then make its own reasoning about these preferences and adapt interaction accordingly. This approach has the advantage to leave data about the users under control of the users themselves.

During the interaction, the user may ask information to the agent in that particular domain, by using the interface of his/her personal device. The Agent answers by using a multimodal presentation strategy in the kiosk and by sending, at the same time, to the PDA its output move in written form. Possible user answers or reactions, which are enabled in that phase of the dialog, are sent as well to the PDA to be displayed as interface icons; the user may also make a different move by means of a text-based input, which is parsed by a simple keyword matching method.

3. AN EXAMPLE OF INTERACTION

Let us describe how the system works, using as an example the Travel Agency domain. In particular, we make as an hypothesis that the system is running in a Southern Italy Tourist Information Point providing tourist information about the region.

Initially, according to the location and the environment settings, the initial environment and agent models are triggered (Table1). The Environment model stores information about the type of interaction space, the list of objects and places that can be referenced within the digital space (the agent background scene or window panels) or the real one (front desk, etc.) and the list of people working at the current time in the environment. This knowledge about the environment enables the agent to use, for instance, deixis to refer to objects or human colleagues mentioned in the dialog.

Agent features are set to meet the requirements of the expected user group (adults with interest in visiting the region, knowing about suggested itineraries, local attractions, food and so on) and of the cultural behaviour which is typical of the agency location. These features influence the agent embodiment and behaviour. We selected a dark-haired female face showing a "warm" friendly behaviour. The role of a travel agent does not require a strong empathic attitude; she has rather to establish a social relation with the user by showing that it understands the user feedback.

Table1: a synthetic description of the Environment and the Agent Model in the Travel Agency example

Environment Features	
Type	Travel Agency
Interaction Space	Public
People References	Front-Desk: Maria((Mon, Thurs; 8-13)(Sat, Sun;13-19)) Rocco((Mon, Thurs; 13-19)(Fri, Sun; 8-19))
Space References	(infopanel,coord1), (front desk,coord2), …
Agent Features	
Role Personality	Travel Agent Friendly
Gender Culture	Female Southern Italy
Emphatic Attitude	Medium
Default Agent Goals	
Describe(Role(Agent)) ∀x | x = Destination(travel): Suggest(x) DescribeinGeneral(x) Describe(SelectedItineraries(x)) Describe(Art&Culture(x)) Describe(Nature(x)) Describe(Accomodation(x))	

The agent's domain-knowledge (travel info in this case) is directly related to the environment in which interaction occurs. As a consequence, corresponding default goals and communicative plans of the agent are activated as shown in the last section of Table1: first of all, the agent *introduces herself*; then, the presentation of *suggested destinations* starts, by using a web window panel in the agent background. If no interruption occurs from the users, the agent describes, for each mentioned destination, the most *popular itineraries* in the current period, the *artistic and cultural attractions* and so on. We defined these presentation plans as defaults after analyzing information on the web pages of some on-line tourist information providers. During interaction, the agent's goal can be updated according to explicit users requests. An example of generated dialog, illustrating this scenario, is shown in Figure 2.

Agent 1> Welcome to Sud Vacanze travel agency! I'm Mary and I will illustrate you the most attractive places in this region.

Panel 1> Suggested Destination

Agent 2> Puglia, the region where you are, offers a combination of artistic places to visit and a beautiful nature and sea. … Are you interested in some place in particular?

User 1> Selects artistic places icon.

Agent 3> Fine! These are the most popular itineraries in this period of the year.

Panel 2> Suggested Artistic Itineraries in Puglia

Agent 4> You can get more information by selecting icons in this panel. If you want to buy or make a reservation for one of them, you can ask to me or to Maria at the front desk ….

Figure 2. An example of generated interactive presentation in the Travel Agency Domain

As mentioned before, the user move causes changes in the agent's goals and plans. In this example, when the user shows his/her interests in artistic itineraries (User 1 move in Figure2) the agent revise her goals and after showing her agreement with the user choice, the agent starts illustrating some selected artistic itineraries in the region. She then suggests to the user how to buy a ticket or an holiday package by using the kiosk facility or from the front desk where Maria is working (as stated in Table 1).

Each Agent's move is tagged according to APML (Figure 3). The Body Generator reads this annotated string, interprets it and renders it according to the available communicative channels: different bodies may have different expressive capabilities and therefore may use different channels.

```
<APML> <turn-allocation type="take turn">
<performative type="inform"> You can get more information by selecting
<emphasis x-pitchaccent="Hstar" deictic="infopanel"> icons </emphasis>
and, if you want to buy or make a reservation for one of those, you can ask
to me or to <emphasis x-pitchaccent="Hstar" deictic="front desk"> Maria at
the front desk </emphasis> .</performative></turn-allocation> </APML>
```

Figure 3. APML tagged move.

This is the general agent behaviour, let's see what changes when the user interacts with the system through his/her personal device.

3.1 Personalizing Interaction

As we mentioned in the Introduction, if the user interacts with the Agent using his/her personal device, when he/she is in proximity of an information access point where the agent is displayed, user-related information are transferred to it. Therefore, information provided may be adapted to that user, who can handle interaction with the PDA.

In our system, information about the user (Mobile User Profile: MUP) are represented in a XML based language since it has to be shared and understood by different environments. In particular, we specify information in the MUP according to the *UbisWorld* language [14]. This language is able to integrate user features and data with situational statements and privacy settings in order to support human-computer ubiquitous interaction.

Table 2: a fragment of MUP

```
<SituationalStatement version="Full_0.1">
<content> <subject><UbisWorld:Dora /></subject>
<predicate><UserOL:holiday /></predicate>
<predicate-range><UserOL:sea,art,mountain/>
</predicate-range><object>art</object>
<predicate><UserOL:accomodation/></predicate>
<predicate-range><UserOL:cheap,resort,2stars,3stars,.../>
</predicate-range><object>3stars</object>
</content>
<restriction><location>italy</location></restriction>
<meta>
<owner><UbisWorld:Dora /></owner>
<privacy><UbisWorld:friends /></privacy>
<purpose><UbisWorld:commercial /></purpose>
<retention><UbisWorld:short /></retention>
<viewer><UbisWorld:SudVacanze /></viewer>
<evidence>not-specified</evidence>
<confidence>0.75 </confidence>
</meta>
</SituationalStatement>
```

Table 2 represents a simple fragment of the MUP of a tourist named Dora who is highly interested to artistic places. When the conversation starts, the Agent asks to the user modeling component the information it needs for personalizing its presentations; the UM component sends back these data in an XML string. The Agent may use this information to improve its knowledge about the user needs, goals and preferences.

In our example (Figure 4), the fact that the user likes artistic places when she is on holiday will allow to focus the Agent's suggestions on artistic itineraries in Apulia by adapting them to the user characteristics.

Agent 1> Welcome to SudVacanze travel agency! I will illustrate you some of the most interesting itineraries and possible accommodation solutions in the artistic town of this region. **Itinerary 1.** ....
Screen1> Suggested Itinearies Panel
Agent 2> Which itinerary do you prefer?
User 1> Selects Itinerary 1 icon
Agent 3> Do you want suggestions about possible accomodations?
User 1> Yes please
Agent 3> These are the 3 stars hotels in Bari.
Screen 2> Suggested Accomodations
...

Figure 4. An example of dialog adapted to user preferences.

Figure 5 shows the interaction with the user personal device. In order to avoid problem due to noise in the environment and to have more privacy during the interaction, the user answers to the Agent question by selecting the icon corresponding to his/her choice. These icons are generated dynamically according to the dialog history. In this case the user may select *Castel del Monte* icon, corresponding to *Itinerary 1,* or *Trulli* icon (Itinerary 2).

User answers and actions may produce changes in the user model: when new information about the user can be inferred, it updates or adds a new slot in the MUP and sets the "confidence" attribute of that slot with an appropriate value. This is calculated as a function of all the previous user actions which may have an impact on that slot and is categorized as being 'low', 'medium' or 'high'.

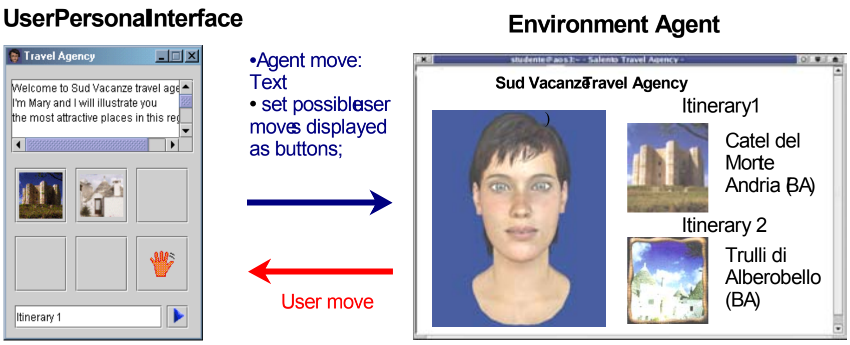

Figure 5: Personal Device Interaction

4. CONCLUSIONS

This research builds on prior work of our research group on dialog simulation [4], emotion modelling and personalization in ubiquitous computing [2]. In our opinion, an ECA represents a creative expression of a natural and engaging interface between users and services of smart environments. In this paper we discussed, in particular, the personalization of conversations with an embodied intelligent agent in public spaces, where interaction may be performed using a public touch screen or a personal device. The last option allows to overcome to limit typical of public space interaction such as lack of personalization in presenting the information and I/O problems due to the noisy nature of the environment.

5. ACKNOWLEDGMENTS

We thank our Magicster project partners and also those who cooperated in implementing the prototype described in this paper: in particular, Giuseppe Cellamare, GianLuigi Del Vecchio, Giuseppe Grassano and Ignazio Palmisano. Finally, we thank Fiorella de Rosis to which we owe several fruitful ideas underlining this work.

6. REFERENCES

[1]Cassell J., Stocky T., Bickmore T., Gao Y., Nakano Y., Ryokai K., Tversky D., Vaucelle C., Vilhjálmsson, H. (2002). MACK: Media lab Autonomous Conversational Kiosk. *Proceedings of Imagina02*. February 12-15, Monte Carlo.

[2]De Carolis B., Pizzutilo S. and Palmisano I. D-Me: Personal interaction in Smart Environments.In P Brusilovsky, A Corbett and F de Rosis (Eds*): Proceedings of User Modeling '03*.Springer LNAI 2702. 2003.

[3]De Carolis B., Pelachaud C., Poggi I. and Steedman M. APML, a Mark-up Language for Believable Behavior Generation. *In H Prendinger and M Ishizuka (Eds): "Life-like Characters. Tools, Affective Functions and Applications"*. Springer, in press.

[4]de Rosis F., De Carolis B., Carofiglio V. and Pizzutilo S. Shallow and inner forms of emotional intelligence in advisory dialog simulation. *In H Prendinger and M Ishizuka (Eds): "Life-like Characters. Tools, ffective Functions and Applications"*. Springer, in press.

[5]de Rosis F., Pelachaud C. and Poggi I. Transcultural believability in Embodied Agents: a matter of consistent adaptation. In S. Payr and R. Trappl (Eds): "*Agent Culture: Designing Human-Agent Interaction in a Multicultural World."*Laurence Erlbaum Associates, 2003.

[6]Gratch J., Rickel J., Elisabeth Andre, Norman Badler, Justine Cassell, Eric Petajan (2002) Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems* 17(4): 54-63.

[7]HAPTEK: http://www.haptek.com

[8]Kobsa A., Generic User Modeling Systems. *UMUAI vol. II* nos.1-2 pp.49-63. Kluwer Academic Publisher. 2001.

[9]Pelachaud C., Poggi I., Subtleties of Facial Expressions in Embodied Agents, *Journal of Visualization and Computer Animation*, to appear, 2002 .

[10]MS Agent:http://www.microsoft.com/msagent/

[11]Picard W. R., Klein J. (2002*), Computers that Recognise and Respond to User Emotion: Theoretical and Practical Implications, Interacting with Computers*, 14, 2 (2002), 141-169. TR 538 .

[12]Reiter E. and Dale R. *Building Natural Language Generation Systems*. Cambridge University Press. 2000.

[13]TRINDIKIT 2.0 Manual: http://www.ling.gu.se/research/projects/trindi

[14]UbisWorld home page :http://www.u2m.org/

[15]Weiser M. *The Computer for the 21st Century. Scientific American*, september 1991.