Co-Habited Mixed-Realities

Walter Van de Velde
Artificial Intelligence Laboratory
Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels
walter@arti.vub.ac.be

Abstract

This paper proposes the concept of Co-Habited Mixed Reality, and argues how it serves as a basis for large-scale social interaction systems and communityware. The technology that it implies is aimed at focusing a large potential for interaction such that effectiveness of participation to large scale events is enhanced. We illustrate the idea with scenarios from an ongoing project (COMRIS: co-habited mixed reality information spaces), implementing the cohabited mixed reality for a large conference site or fare. Several of the key issues involved in this realization are briefly summarized. Work from other projects that demonstrates the feasibility is also described.

This paper is submitted to the IJCAI-97 workshop on Social Interaction and Communityware, and contains original and previously unpublished work.

1 Co-Habited Mixed Reality

In this paper we are concerned with concepts and technologies that can enhance a variety of social processes in large and informally structured groups. The applications that we have in mind involve large information intensive events in which the possibilities for interactions among participants are many, but not every encounter leads to a productive social process. For example, at a major fair there may be many potential buyers and sellers, but only some will come to a productive interaction, i.c. a business deal. Or, at a conference center like the Nagoya IJCAI-97 site, the scale and great diversity of topics and purposes that are being addressed make it difficult to get everything done, i.e., each of one's interests satisfied.

Our results must help *to focus a large potential for interaction such that effectiveness of participation to such large events is enhanced.* To this end, the idea of cohabited mixed reality brings together three visions that are driving much work on collaborative systems. We share with many the vision of *shared virtual spaces* (e.g. [Fahlen, 1993]), projecting that the future open and wide-area networks will have the characteristics of large structures of interconnected places, in which participants can navigate, act and interact. Within such places, information provides a (shared) context for the different participants to access and change. The structure and organization of a place enables (and possibly regulates) those interactions that serve the place's purpose, like entertainment, commerce, learning, and so on. We are confronting this vision with a second one: the *agent-based approach* (e.g. [Genesereth and Ketchpel, 1994]). Here one imagines the future information systems as a huge collection of software agents, each of them offering or autonomously pursuing a service on behalf of users or other agents. A third trend that is basic to our view is that of a *radical information push.* Information push is contrasted with information pull models, in which the user has to take the initiative to pull the information that she thinks she needs to her. Current information push models provide profile-based pro-active information delivery, such as in customized news filtering and delivery applications (also known as 'publish and subscribe'). Our radical version goes much further as we pursue context- and situation-based push: at each moment the information that is deemed relevant for the user's current minute situation is pushed toward her.

What these three trends amount to is a view of the network turned inside out: rather than viewing it as a large information store from which users retrieve what they need, the network itself is a huge virtual society of agents, trying to reach out to the users. The basic challenge behind pulling this together is one of *scalability.* For example, with virtual spaces within which millions of agents are actively trying 'to do their thing', and are competing to get to potential users, there is a question how this can still behave meaningfully toward the user, how she will be protected from information overload or pollution. The conceptual and technological solutions

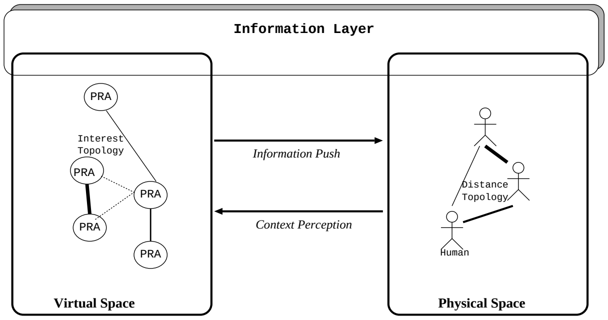

Figure 1: The basic COMRIS vision of Co-Habited Mixed Reality Information Spaces.

adopted for Co-Habited Mixed Realities are basically aimed at dealing with this problem of scalability.

Figure 1 illustrates the basic vision of co-habited mixed reality. Our interpretation of the notion of mixed reality space is that it consists of a couple of interleaved spaces, one being a real space, the other a virtual one. The real space is populated by real agents, the human participants in some social processes that are going on in that real space. For example, a conference center is populated by participants, organizational staff, exhibitors, and so on. These are engaged in numerous social interactions: listening to talks, discussing specific topics, negotiating job-opportunities, defining joint projects, clearing-up misunderstandings, or just reflecting on 'the good times' over a drink. The virtual space, on the other hand, is populated by virtual agents (software agents), that act as the virtual representatives of the real agents. Each such agent actively represents an interest of a real agent. For example, a speaker may have as interest that many people come to listen to her talk, or maybe she wants a specific person to attend. One of 'her' agents would represent this interest by actively promoting it, e.g. by personally inviting people, by offering them a targeted teaser or sneak preview, by reminding people in time that the talk is due to begin, and even ultimately by luring into the conference room those participants that happened to walk past its door a few minutes before the start. Of course such an agent would be only one of several that our speaker uses. Another one would pursue her interest in, for example, optical computing, yet another one could attend to the problem of finding an interesting southern European partner for this project that she is currently setting up. Taking the standpoint of another participant, then, it is clear that many virtual agents may be trying to catch my attention. Here lies the obvious critique on information push models, as well as the biggest scalability problem. We will return to these issues as we describe our model of 'competition for attention'.

The peculiarity of our concept of mixed-reality is in the relation between the real and the virtual space. In particular, we require no mapping between both, neither do we envisage the perceptual integration, as it would typically be done in augmented reality [Azuma, 1997]. Instead, we contend that the processes in the two spaces must only serve each other, without there being a need for the real participants to have a feeling of virtual presence, or the other way around. Our mixed-reality is thus closer to a 'dual-reality', in which the coupling between the two spaces can be much looser. This has, for us, two advantages. First we are relieved of the demanding task to fully integrate the real and virtual space in one augmented (and interactively responsive) perception. After all, one can not expect real conferences to take place soon in immersive virtual reality environments. Instead we must only provide a personal link between the two spaces, that assures some form of 'synchronization' so that the activities in one space further the ones in the other (and *vise versa*). This link is in a portable and affordable device that can be worn like a (large) badge on a person, with through ear-phone speech output. This device – the 'parrot' that could be sitting on ones shoulder – provides a bi-directional link between the real and the virtual world. On the one hand it perceives what is going on around its host (context perception, in Figure 1). On the other hand it points her to relevant information so that participation in the ongoing event is enhanced (information push). While the participant is participating in real space, gently directed by its parrot, a similar event is going in virtual space. In the virtual conference center the virtual agents are the participants. It is their activities and encounters that produce the information that is relevant to further the real participant's interests, even those that the latter is not explicitly attending to. For instance, a participant may ask, after the event, one of its virtual representatives a summary about an interest that was 'delegated' to it (see Section 2). Note that some real participants may not use virtual representatives, and that some software agents may have physically remote owners that may or may not be on-line. The scheme thus allows for different degrees of mixed-ness, distributed-ness and synchronization.

The second advantage arrizing from the loosely coupled spaces is that the structure of the virtual space does not need to map in any way to the structure of the real space. Its virtuality can be exploited more freely. Our idea here is to view space as a potential for interaction. Potential for interaction in real space is determined by physical space: things and persons that are closer have a higher potential for interaction. We define the topology of the virtual space in the same way, i.e., as a potential for interaction. However the interaction here does not rely on physical space. Instead we call it interest based. Two agents for which some social process furthers their interests must be considered to be close in virtual space. Closeness of interest is the simplest case of this. When two agents have similar interests than they may want to talk to eachother about it. However, also different interests may become close to eachother. For example, the interests of a buyer and a seller are obviously not the same, yet there exists a social (commercial) process that furthers both. What we implement in the virtual space, then, is a way for agents to perceive potential for interaction. This perception is essentially subjective, and it is only when two subjective perceptions are close enough to eachother that a successful encounter may be initiated. We call 'interest based navigation' this process of agents moving 'closer' into regions and toward eachother such that potential for interaction is increased. We will return later to explaining the ways in which potential for interaction can be represented and perceived.

The use of a portable personal assistant is kin to work on wearable computing (e.g. [Starner *et al.*, 1995]). The innovation in COMRIS is that it serves as an active window into another world, from which information trickles through to the extent that it is relevant to the situation in the real world. The COMRIS parrot is not just a memory aid, but constantly on the watch to let relevant information through, while blocking of the irrelevant or unwanted, at that moment in time. The effectiveness of this idea depends on (1) the capability to detect the context, situation and interests of the host on a minute basis, (2) the capacity of the virtual space, its inhabitants and the processes going on in it to focus on activities that are relevant for the user's interests, and (3) the capability to exploit structured and unstructured information sources related to the event and its participants.

2 Application Scenarios

The *conference center* is a structure of places for registration, presentation, refreshment, and so on. At a conference, like the IJCAI, people gather to show their results, see other interesting things, find interesting people, meet officials or celebrities in person, or engage in any kind of discussion. The COMRIS system can be used to improve the participation in such an event. Each participant has its personal agent (in the form of an electronic badge and ear-phone device, hooked by IR-beacon to an Intranet) that informs her about potentially useful encounters, ongoing demonstrations that may be worthwhile attending, even places to avoid, and so on. The personal assistant communicates with a range of personal assistants that 'live' in a virtual conference center. The virtual space is thus populated by the personal representatives of the participants, but linked by the user's personal device to the real center, which is populated by the real participants. In this way perception in the one reality augments the perception of the other. The topology of the virtual space is interest-based, while the one in real space is the normal physical space. The personal agents perform interest based navigation, while the feedback to the user is managed by competition for attention (see Section 4). Note that some real participants may not use virtual representatives, and that some software agents may have physically remote owners that may or may not be on-line. The scheme thus allows for different degrees of mixed-ness, distributed-ness and synchronization.

The conference center is clearly an information intensive space, in which a variety of social processes are going on and in which different interests play a role. Moreover it provides a potentially overwhelming potential for interaction, requiring mechanisms to channel it into a manageable stream. It also allows for various degrees of complexity along different dimensions: number of agents, complexity of interest, number of participants, organization and nature of information, and so on.

To make this more concrete we describe in the following three simple scenarios.

Matchmaking. The COMRIS system can help to identify people with a common interest. The software agents that are representing particular participants interests will, by interest-based navigation, get close to each other in virtual space. They will possibly exchange information about their respective hosts, on the basis of which the relevance of drawing the attention of this host will be evaluated. Especially when two real participants with similar interests are detected close to each other, the parrot may inform its user of a useful encounter. This shows how the COMRIS system can initiate and support wanted communication between humans in the physical space, because it improves the chances of starting conversations with people that share common interests. The real facilitation of this interaction is based on an interest-based matchmaking mechanism that creates social groups of personal representation agents in a virtual space.

Agenda management. The COMRIS system can help in undertaking interesting activities. In the virtual space there are software agents that advertise particular events (e.g. talks or demonstrations during a conference). Via the interest based navigation mechanism they will create a common interest relationship with the personal representative agents of a user. Depending on the current situation and context of the user in the physical space, the right personal representative agent will get the attention of the parrot of this user. This parrot will then inform the user at the right time about an interesting event that might be worth participating. This shows how the COMRIS system relieves users from having to keep upcoming activities in the back of their mind; they can concentrate fully on the activities they engage in, and will be notified at the right time of activities that interest them. Whereas the first scenario focuses on managing the activities that are going on around a real participant, this one aims at directing this participant in choosing activities.

Parallel conferences. The mixed reality approach in the COMRIS system allows for real and virtual activities to go on in parallel. While a participant is fully focusing on one interest (for instance to learn about some topic), he or she may have delegated interests to other agents. These software agents then autonomously try to further that interest by seeking contacts with other agents, scanning relevant information that they provide, and so on. At no point in time during the conference is the real participant disturbed by this representative. It is only after the conference that the user decides to ask his personal representative for a report on its activities related to that particular interest. This will be in the form of a topical summary with pointers to concrete information pages that have been found relevant. This shows how the COMRIS system supports tracing a kind of parallel virtually attending a conference via software agents. Therefore this scheme also allows for purely virtual applications.

3 Technology overview

The technical framework for COMRIS, the acronym that we'll use to refer to the system[1], is sketched in Figure 2.

Information Layer: The information layer comprises distributed multimedia stored in the World Wide Web, and in, via Internet technology accessible, databases and legacy applications. It also comprises non-electronic information which may not be accessible from virtual space, just as some electronic information may not be accessible from the real conference space. The architecture presents this information as a largely shared accessible information layer for the virtual and the physical space.

Virtual Space: The virtual space is where software agents reside. The software agents that operate here are called Personal Representative Agents (PR-Agents). These agents have access to the information layer for retrieving and storing information that is relevant for their agent related activities. Figure 3 shows the conceptualization of our PR-agents. Basically an agent encapsulates an interest. Its performance is a measure for its accumulated success in representing or furthering an interest. Its relevance measure express how relevant its actions are with respect to a particular situation. Its competence measure on the other hand, expresses how good it is at doing an action, irrespective of whether it is

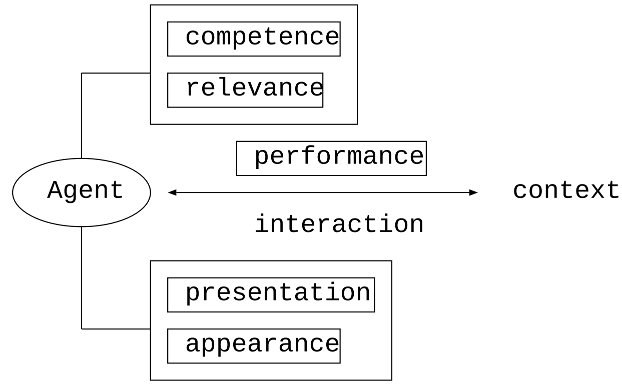

Figure 3: A model of software agent

relevant or not. Performance, relevance and competence are the elements that regulate the interactions through 'competition for attention'.

Finally, its appearance and presentation constitute its perceivable parts (appearance is perceivable for all other agents, presentation requires an explicit interaction, like giving a business card). These are used, basically, for agents to estimate potential for interaction and thus define the (subjective) structure of the virtual space.

The PR-agents have a hook to the physical agent layer via a communication channel with their associated personal assistant agent (PA-agent). The communication from the PR-agents to the PA-agents is based on the the information push model. In the virtual space agents exploit the interest-based navigation and competition for attention mechanisms for realizing their activities. This dynamically evolving structure brings agents together who's interests are likely to come together in productive interaction.

Physical Space: The physical space is inhabited by human agents. In this space all participating humans wear their Personal Assistant or parrot. Furthermore context beacons are placed here, in order to make it possible for the Personal Assistants to recognize the kind of Physical Space they are currently operating in. The people in this layer also have access to the information layer which contains the same information stores as those which are available to the software agents in the virtual space.

Personal Assistants: The Personal Assistant (PA) bridges the gap between the virtual and the physical space. Therefor it is partially embedded in both spaces. A PA is a hardware device carried by a human agent which monitors the physical space and interacts whit its human carrier. On the other hand the personal assistant is wirelessly connected to the Internet, and it contains software that communicates to the software agents in the virtual space. It is through a PA that PR-agents have to compete for the attention of the user. This compe-

[1]COMRIS is also the name of project LTR 25500 within the EU Long-Term Research initiative focussed on Intelligent Information Interfaces (I3).

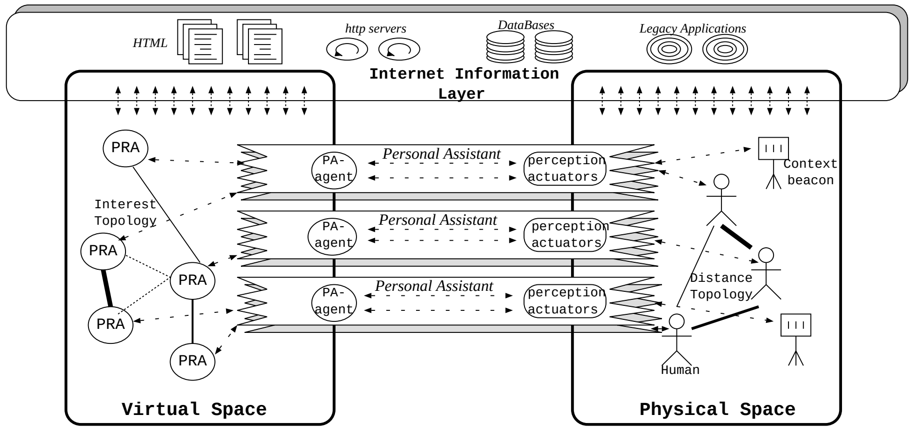

Figure 2: Co-habited mixed reality technical framework.

tition is biased by the interests that are perceived most relevant to the user at each moment. We identify and shortly describe the three main interaction components that constitute a personal assistant.

1. A perception system to the physical space in which humans operate. This perception module contains a (primitive) vision component that can capture the context of the current situation in physical space (where, - but without absolute positioning - and what is the host doing (walking, listening, talking, eating,...). Moreover it allows to recognize other PA-devices. It serves the need of an input communication channel for the PA which will be carried by a human participatory in the physical space.
2. An actuator component that passes information to its human carrier. This is mainly a speech module that serves the purpose of an output communication interface from the PA to the human, as well as simple display modes (e.g. LED-bars).
3. A Personal Assistant Agent. This is the inside cognitive part of the personal assistant. It controls the perception and actuator module and communicates with the PR-agents in the virtual space. For the latter it basically manages the competition for attention that all PR-agents are engaged in toward the parrot. Therefor this component makes the link between the current situation of its human carrier in the physical space and the activities of its associated PR-agents in the virtual space.

It is important to be clear about the role of natural language and speech in this system. It is not our intention to use natural language for all information that the system may handle and want to communicate to the users. We claim, however that the role of natural language in mixed-reality information spaces is to provide, at each moment, a context sensitive information indicator, i.e. a short text in which the relevance of a particular recommended action or source of information is justified with respect to the context of the interaction [Geldof and Van de Velde, 1997]. The actual exploration of these sources is left to the discretion of the user.

4 Key issues and challenges

To fully develop the COMRIS technology requires to face several research challenges, some of which are elaborated upon below.

4.1 Interests and Agents

We have proposed a model of co-habitation in which real participants inhabit the real space, while virtual agents inhabit the virtual space. The challenge is to come up with a model of virtual agents that serves the needs of enhancing social interactions in the real space. Software agents as active representatives of a user's interest are a means to achieve this. Agents will constantly push a user's interest, in competition with other agents pushing their own interest. What the actual actions are that an agent will take depends on the interest that it encapsulates. Interest profiles have been used mainly in two ways. First to manage similar interests (e.g. profile based information filtering, match makers, data warehouse technology). Secondly interest models have been used in situations of, generally speaking, total quality management with a focus on detecting conflicting interest and subsequent conflict resolution (e.g. EU GEOMED project mediation system [Van de Velde, 1996], or in collaborative engineering applications).

We propose another interpretation of this notion of interest that comprises the previous ones, but significantly extends the scope. We view social processes as being founded on constellations of related interests, which must not be matching or conflicting, but be in some sense compatible. For example, in a commercial scenario, the buyer and seller obviously do not have the same interest, but their interest can be combined to further one another.

The point that we want to make here is that it is not sufficient to directly work at the level of interest models. Several options present themselves: explicit modeling of interaction protocols as frameworks in which slots can be matched against specific interests, or implicit modeling of such protocols by imposing structure on a society so that potential participants in a social process can gradually focus in on the act of interacting. An advantage of the second approach is this zooming in which allows for more flexible match making, and thus also for deferred commitment. The COMRIS research explores the second option. This is technically realized through the techniques of interest based navigation.

4.2 Interest-based Navigation

A key feature of our virtual space is that its topology is not like real space. Rather it is structured to reflect potential for interaction, not physical, as in the real space but interaction that is likely to further the agent's interests. Interest based navigation is the process by which agents find out about useful places, or encounters with other agents. Agents perceive the space as a potential for interaction. By analogy to physical space, where closeness corresponds to increased potential for physical interaction, the spatial concept of a software agent is also based on estimating potential for interaction. This is the structure we want to implement in virtual space. We claim that social structures (and thus opportunities for useful social interactions) reflect meaningful constellations of interests. Agents want to evaluate the potential for interaction so that their interests get served. The complexity of finding out about useful partners can be managed by the notions of *appearance* and *presentation*. Appearance is the equivalent of the set of bodily features by which humans create expectations about persons. Presentation is analogous to a business card. The point of working with appearance and presentation, rather than directly with interest models was argued for in the previous section. Our research applies ideas from social anthropology (i.c. [Keane, 1997; Synnott, 1993]) to exploit appearance and presentation for focusing interaction. Once an agent has, by interest based navigation, sufficient evidence that an encounter with another one is likely to further its interest, it engages in the second step toward an actual encounter: competition for attention.

4.3 Radical Information Push

We aim to move the prevailing information pull model toward an information push model. In the information push model, the information is actively pushed into the concrete context and situation, and at the right time. The term 'information push' is starting to become a new buzzword. For example: "TIBCO Inc., and more than a dozen Internet companies have endorsed a proposed new industry standard for the "push" model of information distribution over the Internet. The proposed standard, called publish and subscribe, will reduce Internet traffic and make it easier to find and receive information online." (AgentNews, WebLetter, Volume 1,Number 17 December 16, 1996 Baltimore, MD). Note however that our model of information push goes much further in that it pushes information not based on interest alone but on relevance of that interest in the minute situation of a user.

The challenge in this is to make sure that systems have a good idea of the context of its users. The concept of 'place' provides a first important hook here. A 'place' is an environment for a variety of participatory processes. But also the notion of 'interest' must be taken into account. In our view presence in places, interaction with other agents, and interaction with users is based on structures of interests that connect the different players. Understanding these interests that play a role in a participatory process is the second way to focus the information push process. Finally, of course, the broadest perception as possible about the actual situation and activities of the user is a third source of information to focus the information push process.

4.4 Competition for Attention

A key problem in a radical information push approach is to avoid an overload of pushed information. In COMRIS this is achieved by a model of *competition for attention*. The user's attention is limited, so the agents need to compete for it. The vision might be grasped by imagining the net being turned inside out: rather than users looking for services, we think of it as a huge pool of services looking for users. The analogy of a bunch of children stretching their necks to be in the viewer of a camera is also illustrative: the users attention is limited just as the size of the camera field is. In a similar fashion, the user's attention is limited and software agents need to compete for it. It is this competition model that we are exploring in the COMRIS design.

Competition for attention is based on three values that characterize an agent's behavior: a relevance measure, a competence measure and a performance measure. The *relevance measure* R(agent, context, act) captures the relevance of an agent action in a particular (information) context. The *competence measure* C(agent, context, act) captures how good an agent is in doing a particular action. We thus take into account that, depending on the situation, an agent may feel more or less competent for doing a particular action. Thus, although an action may be relevant, the agent may choose not to do it because of

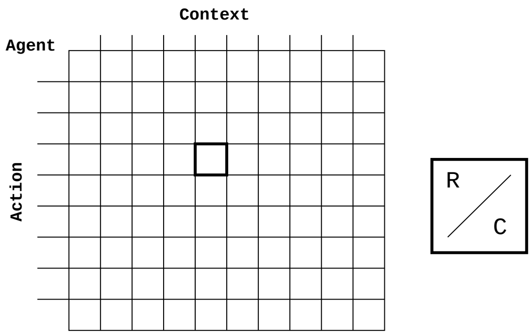

Figure 4: A table to estimate relevance and competence

lack of competence. In that situation another action may be chosen. The *performance measure*, finally, reflects the accumulated success of an agent's actions over time. A success is, roughly, the case in which the agent manages to capture the user's attention. In screen based interaction this might by done by observing clicking-behavior. With our interaction the subsequent perceived contexts and the impact on the interests can be used to estimate performance.

In its current version the COMRIS agents maintain an action-context matrix with two values, one for competence and one for relevance (Figure 4). These values are updated with each action to reflect better estimations of these values. Different updating strategies determine agent 'character':

1. affirmative: the context is the problem
2. uncertain: the competence is the problem
3. and mixtures - survival based on performance

Accumulated success of an agent is reflected in a high performance value. Thus, even if an agent constantly overestimates its relevance and competence value, poor performance will degrade its chances of actually doing good in the competition for attention.

4.5 The Parrot

We are developing a wearable personal assistant. It is inspired by the image of a parrot sitting on one's shoulder, looking around with you and whispering relevant information in your ear. Of course, we hope the COMRIS parrot will be more intelligent than the average parrot who only repeats what you said...

The objective is to produce a lightweight device that can capture as much as possible of the context and situation of its host wearing it. It shall be able to locate the host with respect to the space he or she is in. No exact positioning is required, but presence in a room or closeness to some other must be detected. The device shall also be able to locate other parrots, on other participants, and identify them. Finally it shall communicate wirelessly with an Intranet host where the main cognitive part of the personal assistant is running. The device produces natural language speech as output. The text input for this is provided by the mechanisms of competition for attention.

The main problem that we focus on here is the identification of the user's context. A person carrying a Personal Agent Device is mainly interested in his direct neighborhood. Therefore, limited visibility of a device is acceptable. There are two potential bases for scanning: ultrasound and Infrared. The computational core of the Personal Agent Device has to be specialized to do mainly sensor-fusion and output management (speech component). Existing Wearable Computers are not suited as they are intended to be general-purpose devices. Therefore, they are designed for various input/output devices and without a main concern for speed. The location-problem for Personal Agent Device in contrast requires few specialized input/output channels and a fast computational core. This is mainly due to the need of high sampling rates. Nevertheless, the outcome of this project can serve as a special additional device that can be connected to a Wearable-Computer. As a basis for the computation core we propose to use a sensor-motor brick, originally developed at VUB AI-Lab for small mobile robot applications.

In order to realize various forms of feedback to the parrot, an interesting option for the Personal Agent Device is a vision-module which allows to track hand and body movements of a person opposed to the devise. For example if somebody points to a third person, this module can detect and identify - with the help of the other parts of the Personal Agent Device - the third person. The main part of the module is a camera mounted on a pan-tilt-unit in form of two servos. A prototype of this module exists at the VUB AI-lab which is capable of the described tracking. Its disadvantage is its size as it requires a PC. But the whole systems can be embedded with a small frame-grabber- and DSP-card (size 8x5 cm), making prototyping and the realization is series easy (once the programming problem has been solved) and fast (frame-grabber- and DSP-card are off-the-shelf).

4.6 Text and speech output

The parrot's main output channel is through natural language. The challenge consists in generating and producing 'meta text' about considered information in a very efficient and straightforward way. Text and speech production consists of 3 subtasks: deciding what to say (a), how to say (b), actually say(c) . The main constraints in our setting is that the text should be extremely short (contain only the essential information) and should be produced almost in real-time. In all 3 steps, there are basically 2 options, depending on the requirements of 'efficiency' versus 'flexibility'. The 2 options can be contrasted as 'composition' versus 'synthesis'. The first option uses templates and pre-existing building blocks, the latter consists in performing full-fledged text generation, all the way from concepts to sound signals. With regards to (a) and (b) we opt for the compositional approach. This is justified both by the pragmatic constraints of the application (efficiency reasons) and by the fact that the application domain is well known and fairly delimited. Our strategy consists in creating an interest structure that organizes the possible topics in the domain. 'What to say' is determined by the context in which the user finds himself, this context is interpreted in terms of that interest structure. The context is monitored so as to find out which parts of the interest structure are predominant, the result is a weighted interest structure. In turn, this information serves as input to the module that determines how to say (b). Again, the interest structure is used as a back-bone, a structural organization of template items to be combined to text and encapsulates knowledge about how to produce text for that particular application. For what concerns (c), a major factor in the acceptability of the end result is the quality of its 'prosody', i.e. speech melody, accentuation, rhythm. Best results can be achieved if the application contains a language generation module (a &b) that produces the text to be spoken itself, so that the linguistic and phonetic properties of the message are explicitly known and can serve as 'correct' input to the synthesizer. This means that our templates should also encapsulate the knowledge about phonetic and prosodic properties of the text to be produced. This amounts, among other things, to defining an application-specific lexicon for the pronunciation of all lexical items, to define prosodic rules for the particular types of messages (prompts, monologues, user-system dialogues) that have to be spoken, to build a database of appropriate concatenation units for the synthesizer (diphones, polyphones, demi-syllables, words,...).

5 Discussion

Mixed realities can have several interpretations, for instance augmented reality, or augmented virtuality. In both cases it is required that, in order to achieve a coherent layering of the two (real and virtual), there is some straightforward mapping between them. We can hardly imagine in medium term to apply this concept to an application like a fair or conference. For one, AR applications of this kind require special interaction devices that do not accommodate easily any other kind of normal behavior. Moreover this technology is still prohibitively expensive. Including presentations of other actors in this kind of displays is only starting to happen, let alone the use of active representatives in it. Finally, this direction typically emphasizes synchronized multi-user scenarios, whereas we want to include, in addition, asynchronous features of interaction.

We have proposed our concept of co-habited mixed-reality as a coupled pair of a real and a virtual space, each with its own structure and inhabitants. We think that this concept is an original one and accommodates various degrees of scale, mixed-ness, distributed-ness and synchronization. Our work develops interaction tools and techniques that illustrate the minimal awareness of the one space from within the other. This has the potential of removing one major bottleneck in the large scale pick-up of mixed reality applications, namely price and 'heavyness' of solutions.

Related work that is going in a similar direction is on wearable computing [Starner *et al.*, 1995]. The aims of that work are more ambitious in one sense, less ambitious in another. The capacity of a wearable computer is envisaged to include sophisticated support for memory, and image analysis. On the other hand it does not emphasize the multi-user possibilities of such devices, even if they are simple. This is what COMRIS is focusing on in particular. An experiment that is in the direction of COMRIS is reported on in [Borovoy *et al.*, 1996]. Here computationally augmented batches indicate closeness of user profile of two people standing in front of each other. It illustrates feasibility but lacks the back-up of the virtual world and the agents in COMRIS. In most work on wearable computing is a sink, a dead-end into memory. In COMRIS it is a window on a different world.

Scalability is in some sense, the basic challenge faced. The underlying motivation for introducing such concepts as spaces, interest based navigation and competition for attention is to channel an overwhelming potential for interaction into a manageable stream. Competition for attention as a paradigm for interacting with large agent societies is, to the best of my knowledge, a new idea. The need for an 'economy of attention' is stressed in speak about multi-media and internet (e.g. Esther Dyson, at a symposium on Internet and Politics, 1997), but no computational realization has been proposed. The basic model of competition for attention has been experimented with in a series of WWW servers for a Brussels' movie festival, Ecran Total [Geldof and Van de Velde, 1997]. In this application 6 software agents compete for the attention of the user. Each of them is trying to push a service or subset of information toward the user. The context, in this application, consists of the real-world situation, teleological features and user characteristics. User interests are tracked solely by observing clicking behavior of the users. Although this application is very small in comparison to the scale that COMRIS is aiming at, it has broken ground to demonstrate how competition for attention can be used, first as a conceptual design paradigm, and second as a technology for integrating information and agent-based services (see also [Schrooten, 1996]).

Interest based navigation is also aimed at managing the complexity of figuring out with whom to interact in large scale societies. An important piece of work in this area is around the concepts of *Aura, awareness, focus and nimbus* [Fahlen, 1993], which are also about finding those with which one wish to communicate. It is used to prevent both cognitive and computational overflow caused by group interaction. Aura intersection is necessary but not sufficient for interaction. Once auras collide, each entity calculates its awareness of the other. Awareness is determined by a combination of the focus of the observing entity and the nimbus or presence of the observed entity. No doubt these concepts are related to interest based navigation. The difference is that we do not consider space as a given, but as subjectively constructed by the agents.

Another approach to dealing with complexity, and one which we have not fundamentally incorporated in COMRIS, is to incorporate emotion [Picard, 1995], [Pfeifer, 1988], [Bates, 1994]. Emotion allows for the communication of higher order information. We think that it would be useful both as part of interest based navigation, and as an element of expression of the parrot toward its host. For example, [Canamero, 1997] has developed a agent control modules that take into account elements of a hormonal emotional basis of control, which could be linked to appearance and presentation features. Emotional expression is also the subject of research on 'affective computing' [Picard, 1995] and believable agents [Bates, 1994]. Taking into account emotion in the speech prosody of the parrot would be a challenging direction.

The basic text generation technology has been experimented with in the latest version of the Ecran Total server. In that application a context-sensitive navigation point is presented to the user in the form of a short text (in fact, a meta-text), that provides justified pointers to the pages are most likely to be relevant to the user. The technology is template-based, with dynamic aggregation of parameterized templates. A series of simple tricks improve the quality and readability of the produced text [Geldof and Van de Velde, 1997]. The speech component has not been used so far, but we expect little difficulty in using commercially available speech synthesis techniques, starting from the texts generated.

The parrot hardware technology derives from previous research at the VUB AI-Lab on lego-vehicles. Lego-vehicles are small autonomous robots with on-board computing facilities. For these a brain-brick and sensory-motor brick has been developed. The first one is specialized to compute robot behavior. The second on handles sensor and motor IO of a wide variety of types (analog, digital, IR, light, motor, and so on). The most recent versions are equipped with a radio-link that allows the communication to a (fixed) host computer. This technology can be used for the purposes of COMRIS, in which the robot body is now, in a sense, provided by a human. At least active IR, low-level vision, audio, and the radio-link will be used. The actual construction of a prototype parrot has only started recently.

The virtual space is a large scale society of agents. Our work in this area hinges on both multi-agent systems and on artificial life. In its use a large scale evolving societies it is similar to [Axtell and Epstein, 1997] who study the development of artificial societies under various heterogeneous and locally defined conditions (health, wealth, hazards,...). Although in terms of complexity and scale this is what we need, the users are out of the picture (single-reality). From multi-agent systems we take an additional emphasis on coordination and cooperation patterns [Wooldridge and Jennings, 1994]. We view social processes as based on structures of interests that, although dissimilar or incompatible by themselves, can be furthered when brought together in a particular social process (a preliminary exploration in [Van de Velde, 1996]).

The COMRIS technology is under development in a joint effort of 6 European research groups, partially funded by the European Union's Long-Term Research initiative on Intelligent Information Interfaces. In terms of scale the project aims at demonstrating its results with a series of 100 parrot devices by the year 2000.

Acknowledgment

Special thanks to Ronald Schrooten, Andreas Birk, Dolores Cañamero and Kerstin Dautenhahn for helping to shape these concepts. This work is supported by project COMRIS (LTR 25500), within the the EU Long-Term Research initiative focussed on Intelligent Interactive Interfaces (I3). Additional funding comes from the EU Telematics Information Engineering projects MAGICA and GEOMED. GEOMED (IE 2037) stands for 'Geographical mediation systems'. MAGICA (IE 2069) stands for 'Multi-Media Agent-based Interactive Catalogs'. Additional funding is provided by the project SACEA 'Software Agents for Cooperative Environment Administration' funded by the Belgian Office for Scientific, Technical and Cultural Affairs as project IE/XX/A10, and the project NMS (Network Multi-Media Server), funded by the Flemish IWT and Alcatel. The author is a senior researcher for the Science Foundation, Flanders, Belgium (FWO).

References

[Axtell and Epstein, 1997] R. Axtell and Epstein. *Growing Artificial Societies.* MIT Press, Cambridge, MA, 1997.

[Azuma, 1997] Ronald T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 1997. To appear, Fall 1997.

[Bates, 1994] Joseph Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7), July 1994.

[Borovoy *et al.*, 1996] R. Borovoy, M. McDonald, F. Martin, and M. Resnick. Things that blink: Computationally augmented name tags. *IBM Sytems Journal*, 35(3&4):488–496, 1996.

[Canamero, 1997] Dolores Canamero. A hormonal model of emotions for behavior control. AI-Memo, Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, Belgium, 1997.

[Fahlen, 1993] L. Fahlen. The MultiG telepresence system. In *Virtual Reality International 93, Proceedings of the Third Annual Conference on Virtual Reality*, pages 78–86, Meckler, London, 1993.

[Geldof and Van de Velde, 1997] Sabine Geldof and Walter Van de Velde. Context-sensitive hypertext generation. In K. Mahesh, L. Carlson, S. Nirenburg, A. Ram, and P. Resnik, editors, *Working Notes of the AAAI-97 Pring Symposium on Natural Language Processing for the World Wide Web*, pages 54–61, 1997.

[Genesereth and Ketchpel, 1994] M.R. Genesereth and S.P. Ketchpel. Software agents. *Communications of the ACM*, 7(37):48–53, 1994.

[Keane, 1997] Webb Keane. *Signs of recognition: powers and hazards of representation in an Indonesian society*. University of California Press, Berkeley and Los Angeles, California, 1997.

[Pfeifer, 1988] Rolf Pfeifer. Artificial intelligence models of emotion. In V. Hamilton et. al., editor, *Cognitive Perspective on Emotion and Motivation*, pages 287–320. Kluwer Academic Publishers, Amsterdam, 1988.

[Picard, 1995] Rosalind. W. Picard. Affective computing. Technical Report 321, MIT Media Laboratory, Cambridge, MA, 1995.

[Schrooten, 1996] R. Schrooten. Agent-based electronic consumer catalogs. In B. Crabtree and N. Jennings, editors, *Proceedings of the First International Conference on The Practical Application of Intelligent Agents and Multi-Agent Technology*, pages 543–556, Blackpool, UK, 1996. The Practical Application Company.

[Starner *et al.*, 1995] T. Starner, S. Mann, B. Rhodes, J. Healey, J. Levine, and A. Pentland. Wearable computing and augmented reality. Vision and Modeling Technical Report 355, MIT Media Laboratory, Cambridge, MA, 1995.

[Synnott, 1993] A. Synnott. *The Body Social: Symbolism, Self and Society*. Routledge, London and New York, 1993.

[Van de Velde, 1996] W. Van de Velde. Mediation agents in multi-party decision making: An experiment. In B. Crabtree and N. Jennings, editors, *Proceedings of the First International Conference on The Practical Application of Intelligent Agents and Multi-Agent Technology*, pages 629–654, Blackpool, UK, 1996. The Practical Application Company.

[Wooldridge and Jennings, 1994] M. Wooldridge and N.R. Jennings. Agent theories, architectures, and languages: A survey. In M. Wooldridge and N.R. Jennings, editors, *Intelligent Agents*, Lecture notes in Artificial Intelligence 890. ECAI-94, Springer-Verlag, August 1994.