Towards rich context virtual worlds

Mehdi El Jed, Bernard Pavard, and Nico Pallamin.

IRIT-GRIC
Groupe de Recherche en Ingénierie Cognitive,
118 route de Narbonne, 31062 Toulouse Cedex.
{eljed, pavard, pallamin}@irit.fr
http://www.irit.fr/GRIC/VR/

Abstract. This paper discuss modelling context in collaborative virtual environments. Our approach provides the user with 'rich' contextual information immerging him in a 'situated' interaction (cognitive, social and cultural). We propose a design framework in order to improve the richness of the contextual information. In particular, we will consider two basic socio cognitive mechanisms related to multi-user interaction in real world: indexicality and reflexivity.

Keywords: virtual reality, rich context, situated interaction, emotion, social behaviour.

1 Introduction

Context is a key notion in modelling human activities (reasoning, perception, knowledge, comprehension of the language, etc.). The traditional approaches of Artificial Intelligence had early highlighted the theoretical difficulties to formalise this concept. As an example, Bar-Hillel pointed out the fact that, in general, it is not true that every communicative situation can be represented by a non–indexical language (a language that does not have to refer to external context to be understood (Bar-Hillel, 1954)). Indeed, within the framework of the representational AI, any formal notion of context can be hardy identified by a simple set of elements: all contexts can always be related to another more general context in an endless indexical process (Guha, 1991), (Mc Carthy, 1993).

This phenomenon called 'qualification problem' in AI is also found in social sciences and particularly in modelling complex work environments. The choice made by a human operator to select a context related to his decision-making is opportunist and generally non predictable: an operator 'chooses' the context elements according to the problem to be solved and the environmental external elements (Karsenty and Pavard, 1997). The references used to identify this context are at the same time internal (memory, knowledge, etc.) and external (environment of work, organisational rules, etc.). As an example, Salembier showed that air traffic controllers facing exactly the same aircraft configuration never choose the same solution due to unpredictable factors like hour of the day, origin of the aircraft, nationality of pilots, etc (Salembier, 1996).

Many formal solutions have been presented in order to overcome this difficulty without however solving definitively the so called qualification problem (Benerecetti, Bouquet and Bonifacio, 2001) (Mc Carty and Hayes, 1969). Finally, this problem may arise more from the representational nature of the AI & Cognitive theories[1] than from the choice of a good information processing theory. Indeed, this theoretical difficulty tend to 'disappears' with constructivist, connexionnist or post struturalist approaches. These approaches deny the traditional 'one to one' relationship between signifiers (like objects used as belonging to a context) and signified (their meaning for the action) to state a more interactive view where signifiers and signified are in permanent mutual relationship[2].
This means that the meaning of an object is not predefined (stated in a specific way) but determined by its relationship to other objects. This property relaxes the constraint to explicitly list all attributes that constitute a context. Contextual attributes can then be seen as distributed elements over the environment in a mutual relationship: any change in the environment (at the signifiant level) may change the value of other relationship between signifiants and signified.

The aim of this paper is to describe the conceptual and design framework we used in order to implement this concept and to model a 'rich context interaction' between characters cooperating in a virtual world. As in a multi-user video game, these characters are animated by 'real' humans but their gestures, facial expressions and emotional behaviours are generated by an emotional and social model. This model, take into account the intentional user decision to control his avatar (ex: walk, speak, etc.) and try to identify the right context of the interaction to produce the most meaningful social interaction (ex: gaze direction towards others avatars when walking, gesture related to speech production, etc.). The user is thus supposed to be "cognitively and socially situated" in his interaction with others actors (cf. Figure1).

Fig. 1. Overview of virtual interaction between several avatars in the virtual world. Rich context simulation is expected to generate reflexive behavior between users as in a real situation.

[1] According to which the thought can always be represented in terms of linguistic expressions and formalized by predicates leading to a strict correspondence between the code and its meaning.

[2] This point of view has been initially developed in the Saussure model of language (Saussure, 1916). Language is seen as a system of relationship that transcends the notion of user. This concept has been restated in post structuralist approaches.

Within this framework we tried to:
- keep as open as possible the contextual elements of the virtual world that could be used by humans in order to animate their avatar and to build up the right cognitive, social and cultural context,
- minimise the representational dimension of the emotional and social model that generate the character gestures and emotional behaviour.

In order to better understand the influence of the context on virtual interactions, we will first describe the aim of the application then we will focus on two important characteristics: indexicality & reflexivity in communication.

Following sections will be devoted to the description of how we defined contextual information and how our design framework has been specified in order to improve the 'situatedness'[3] of the interactions.

2 Situated virtual interaction

Aim of the application

From a pragmatical point of view, the aim of this application is to develop a game engine able to reproduce a close to reality human-human interaction in a virtual space. As our final goal is to use this game engine to train professional people to interact in a complex cooperative world (firemen, control room, etc.), we have emphasised the fact that gestures as well as body movements and emotional expressions are cognitively, socially and culturally meaningful for the accomplishment of the task (what we called situated virtual interaction).

Indexicality and reflexivity in human communication

Indexicality and reflexivity of human behaviour are two important dimensions of situated virtual interaction: they contribute to the realism of the interactions and to the immersion of the user in the virtual world.
- Indexicality: This concept refers to the fact that human language, to be understood, is strongly dependant on the context of the situation. Indeed, the meaning given by a user to a linguistic expression is often related to several contextual elements such as the identity of the speaker, the form of the statement, the intonation of the voice, etc. Removing these contextual elements may drastically limit the comprehension of the expression and the attribution of a precise meaning to the interaction. Deictic (gestures with reference to external objects in the environment) are good examples of non verbal indexical behaviour and are permanently used by speakers in order to structure their discourse and to refer to a particular context.
- Reflexivity: This concept is based on the idea that any human action will modify the environment and then the context shared by all actors. Reflexively, this modified context will change the decision process for all other users. It is the dynamicity of this endless feedback loop between action and shared context that is supposed to give social consistency to human interaction (as it is stated in the situated cognitive paradigm). As the situation changes, all actors (avatars in co-

[3] By 'situated' we refer to Suchman work (Suchman, 1987) concerning the fact that action cannot be only generated from a predetermined plan but is also 'influenced' by external resources like world artefacts as well as internal resources (cognitive, social or cultural).

presence and objects) take into account these changes and modify their own behaviour. As consequence, interaction between users gives place to an always different comprehension of the situation since each one operates according to its personal experience. Indeed, the interpretation of a particular action is dependant on the culture, beliefs or the assumptions of each social actor.

Emotional and social model

The feeling of presence in a virtual environment is supposed to arise only if the emotional and social dimensions are taken into account. For that, we propose to enrich the avatar behaviour by an emotional and social model. By expressing emotional and social abilities, the avatar will be able to produce on the user a better feeling of social and cultural immersion. For this purpose, we developed an emotional model (El Jed et al., 2004) to evaluate and update the emotional states of each avatar (emotions, stress and mood) according to its interpretation of the virtual world events. Reciprocally, this interpretation of the virtual world will be influenced by these emotional states. A behavioural model for the social interaction was also developed for the selection of the most adapted gestures and body expressions according to the decisions taken by the user.

3 Context analysis

As we stated in the introduction, the notion of context is present in several disciplines and has multiple more or less precise definitions. Several works focus in user context classification (Kaenampornpan et al., 2004) (Dey and Abowd, 2000). We will take as a starting point these classifications in order to structure our context description.

The levels that we have identified are the following:

- Task context: It represents the goals to be carried out and the constraints to be satisfied. We usually make a distinction between the task to be achieved and the activity done to carry out this task.
- Physical context: It represents all the aspects of the external environment relevant to the interaction: physical configuration of the environment (types of the objects, positions and orientations, avatars in the field of view, etc) and various environmental parameters (time, noise, luminosity, etc).
- Emotional context: It represents the internal emotional states that may influence the user in decision-making. By emotional states, we mean the emotions felt (sadness, satisfaction, anger, fear, surprise, etc), the mood (good, bad, etc), the stress and the personality (open, conscientious, extravert, etc.).
- Cognitive context: It gathers the subjective perceptions the user acquires from his environment. It concerns the beliefs, assumptions, intentions, attitudes or needs that take part in the human interpretative processes for the analysis and the comprehension of a situation.
- Social context: It is about the interlocutors' social identity (roles, status, hierarchy, etc). Social context may influence the comprehension of the situation, the way social interactions are carried out (formal, informal, etc.) and the roles played by the participants in a particular situation.
- Cultural context: It concerns the meaning given by an actor to a situation and also relies on implicit beliefs which form the cultural background. The cultural context

is a set of shared and mutually accepted rules of behaviour (ex: specific practices, habits, languages, gestures, etc).
- Linguistic context: It is constituted by three elements: the semantic contents of a dialogue, the communicative function of the sentence and the form taken by the statement.

These various levels of context characterize a situation of virtual interaction. They will be the reference structure that we will follow to model the avatar context.

4 Context perception

In the following part we make a distinction between the context of the user and the one of the avatar. The user's context is defined as all contextual elements of the situation which let the user make a decision and modify the virtual world. These elements can be implicit (beliefs, assumptions, practices, etc.) or explicit (types of the objects, localizations, emotional states, etc).
Avatar context refers to all information accessible in the virtual world that can be used by the software (our model) to produce a non intentional action (complementary to the user intentional action). Unfortunately, these two kinds of context cannot be matched. Many implicit contextual elements relevant to cognitive mechanisms cannot be identified by the software (like user's intention). To avoid this difficulty, we will only focus on the explicit contextual elements (observable ones) to represent the avatar context (cf Figure 2). Although this solution is a simplification of the user's context, we think that such representation is still rich enough to catch the contextual elements necessary to allow the avatar to produce meaningful body expressions (gestures, facial expressions, etc) adapted to the situation.

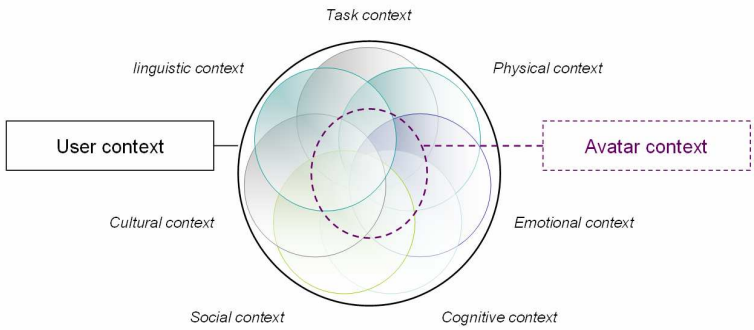

Fig. 2. User context versus avatar context

4.1 Approach

In order to produce meaningful gestures, our model must identify the context of the situation which is (at best) a subset of the real user's context. The contextual elements making possible to constitute the avatar context change from a situation to another. To produce a meaningful behaviour, an avatar considers different contextual elements according to its current situation (for example the elements relevant during a dialogue between several actors are different from the ones relevant when taking care of a victim).

The context so constructed strongly depends on the situation (what we call in our model the task context). The avatar context is constituted by the various elements belonging to the contextual levels (identified in section 3) chosen according to their relevance to the task context.
In our model, we define a set of task contexts (situations) that happen in the scenario and the relative contextual elements to be identified.
For example, the context of an avatar in communication with three other ones can be characterized by the following elements:

Task Context	Physical context	Social context	Emotional context	Cultural context
Dialogue	1st Interlocutor, Position1 , etc.	Role1=chief	Emotion1, stress1, mood1.	Activate a set of behaviour rules and animations to play during the dialogue.
	2nd Interlocutor, Position2, etc.	Role2=colleague	Emotion2, stress2, mood2.	
	3rd Interlocutor, Position3, etc.	Role3=colleague	Emotion3, stress3, mood3.	

Table 1. Example of contextual elements of a dialogue situation

Although the contextual elements characterizing a given situation are predefined, the human user is still able to catch other contextual elements in the environment (in particular the implicit ones) and to exploit them for the decision-making.
As an example, an avatar during a dialogue, identifies its context of communication with 3 interlocutors and decides to distribute alternatively his gaze between his listeners. The meaning of the discussion is not taken into account by the avatar in the construction of its context but the user remains free to control his avatar according to the meaning of the dialogue. He can thus intervene on his avatar to look towards a direction indicated in the discourse or fix gaze on a particular listener.
Such action will have a double consequence: First, the other avatars will simply update their context taking into account the performed action to produce a consistent response (looking towards the pointed direction). Second, the others users will use this gesture as a supplementary cue to get a better interpretation of the situation.

4.2 The 'context perception' process

The context 'perception' process is presented in the figure 3. It is based on an emotional and a perception module.

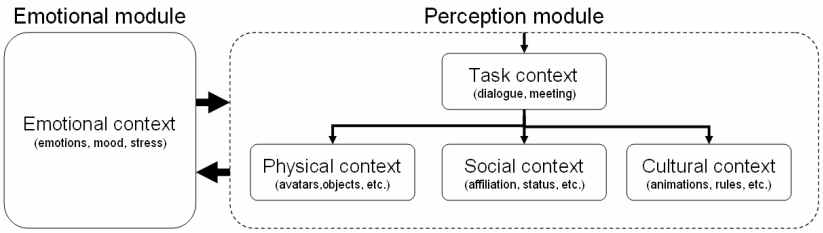

Fig. 3. Process of the context perception

The perception module includes a set of sub-modules. The 'task context' sub module identifies the situation of the interaction. According to this identified situation, the sub module 'physical context' identifies the relevant environmental elements (as stated in our model). The 'social context' sub module recognizes the social role of the actors

and the 'cultural context' module will select the relevant behavioural rules and animations to play. The emotional module will determine the emotional context based on the information received by the perception module. Reciprocally, the perception module is continually influenced by the 'emotional context'.

- **Identification of the task context**

The avatar task context is the situation faced by the avatar in the virtual environment. Taking into account of all the possible situations in a dynamic and unpredictable environment is a complex task. To simplify it, we define a set of possible situations of interactions (situation of dialogue, special events, achieved tasks, etc.) which we will attempt to identify in the simulation.

We are interested in particular in situations which let the users be socially situated in the interaction through verbal (dialogue through the communication network) and nonverbal communication (deictic to indicate objects). As an example, we can have the following situations:

– Dialog situation

A dialogue situation is characterized by the form of the communication (verbal or non verbal) and the role played by each actor in the interaction. The context perception module makes it possible to characterize each dialogue situation by identifying the user who perform a communicative act (verbal or non verbal) and associates to each avatar a role in the dialogue (listener or speaker). This mechanism enables us to detect and identify at any moment of the simulation the occurrence of dialogue situations.

– Encountering situation

An encountering situation is characterized by the proximity in the virtual space of the avatars. The context task sub module detects this proximity and announces that there is an encounter between two or several avatars. This situation is then identified as dialogue if an actor starts a communicative act (speech, deictic, etc.).

These virtual interactions allow the users to react according to their own emotional states, expertise and history. From this point of view, they represent an interesting study framework for emotional and social behaviours.

- **Identification of the physical context**

It is the capacity of an avatar to recognize, for each situation, the associated contextual elements of the environment (objects in the proximity, their types, compositions, positions and orientations). Given an avatar identity, we can obtain many pieces of related information such as name, position, activity, etc.

In a dialogue situation as well as in an encountering situation, an avatar can identify its interlocutors, their positions and orientations to produce the most adapted behaviour.

- **Identification of the social context**

The knowledge of the social identities of the avatars in interaction makes it possible to produce non intentional behaviours adapted to the situation according to actor status in co-presence.

For example an avatar which meets another one with the same social status will look at it a few moments. When the social identity of the avatar met is identified as hierarchical superior in the organization of the actors, a gesture of greet will be produced.

- **Identification of the emotional context**

Each avatar interacting in the virtual environment can evaluate its emotional states. By emotional states, we mean the set of probable emotions felt (satisfaction, disappointment, anger and fear), the state of the mood and the degree of stress. The emotional module calculates and updates the emotional context according to the events perceived by the avatar (as described in the OCC model) (Ortony, Clore and Collins, 1988) and moderated by its personality (open, conscientious, extravert, pleasant and anxious) (Mc Crae and John, 1992).
The emotional context influences both the perceptual process and the production of the behaviours according to the situation. For example, in a meeting situation, an avatar whose degree of stress exceeds a certain threshold will not look at the avatar met (suppressing the natural social response).
An avatar can also perceive the emotions of the other avatars and consequently change its behaviour by contagion of emotions. For example, staying close to stressed avatars it can become stressed too.

- **Identification of the cultural context**

The cultural context sub module allows the avatars to produce body expressions adapted to the implicit interactions rules shared by the members of a group. For example, some conventional gestures are shared by the community of firemen to indicate the end of an intervention or to communicate when the environment is noisy.

4.3 Example of situated interaction

As an example of processing the context perception in the virtual environment, we will analyse the behaviour of two avatars which represent the members of a firemen team.

Fig. 4. Two avatars walking in the virtual world.

The figure 4 shows an avatar (A) animated by a human user walking towards an avatar (B) controlled by another user. The sub module ‘task context’ of the avatar (A) starts by identifying the situation as an ‘encounter’ and inform the other sub modules about the occurrence of the ‘encountering’ situation. The sub module ‘physical context’ identifies the met avatar and the associated environmental elements (position, orientation, etc).

Fig. 5. Social and emotional interaction in a meeting situation.

In figure 5, the sub module of the 'emotional context' (on the avatar A) provides information about its emotional states, stress and mood. It identifies that it is not stressed so an action to look at the meted avatar (B) can be produced. The same behaviour is also produced on the avatar B.
The sub module of social context is then activated to identify the social identity of the meted avatar. The avatar B identifies the avatar A as a hierarchical superior. A gesture of greet will be produced. The sub module of the cultural context makes it possible to identify the rules to be applied at this interaction. It will further identify the set of adapted animations to be played in that situation (animation of official greet).

5 Discussion

In this research, we proposed a general methodological and conceptual framework in order to design rich context virtual worlds. Our strategy was twofold. On one side, we allowed users to identify by themselves the right context taking into account all their social and cultural knowledge (what we called situated interaction following the ethnomethodological paradigm as well a as the Suchman work). On the other side, we designed a game engine that takes into account a model of interaction between user, emotion and avatar behaviour.
Our aim was to try to reproduce close to reality social interaction and to minimise the fact that context in real situations is often too rich to be listed or predetermined (even inside a very specific scenario). We stated in the introduction that the contextual dependence phenomena is difficult if not impossible to formalise in close to reality situation due to the richness of all possible indexical references as well as the reflexivity between actions and context modification.
We stated that this difficulty was mainly due to the fact that traditional AI as well as cognitive sciences are based on a representational paradigm. If we tried to escape this problem putting the users in open and rich virtual worlds, the game engine we designed in order to produce socially and emotionally meaningful gestures had nevertheless to identify the right context (related to the task, personality of the user, social status of interacting characters, etc.). At this point, we were not able to fully avoid the qualification problem because our model tries to simplify the situation by categorising the situation (based on the nature of the scenario, the different phases of the task, etc).
With this approach we tried to identify the context from a combination of parameters. Thus, gestures as well as emotional behaviour calculated from our model may induce wrong reflexive interactions between users. Nevertheless, users may also adjust their

own behaviour in order to produce coherent and meaningful interactions. At this point of the research, we cannot give answers to this question because we need more experiments in order to analyse if such regulation mechanism arise. Future multi-users sessions with professional actors (firemen) will probably give more insight on this point.

References

Bar-Hillel, Y. (1954) Indexical expressions. Mind, 63:359-379.

Benerecetti, M., Bouquet, P., Bonifacio, M.,(2001) "Distributed context-aware systems", Human-Computer Interaction 16, special issue on Context-aware computing, 213--228, 2001.

Dey, A.K., Abowd, G.D. (2000) “Towards a Better Understanding of Context and Context-Awareness”, at the CHI 2000 Workshop on The What, Who, Where, When, Why and How of Context-Awareness, April 1-6, 2000.2.

El Jed, M., Pallamin, N., Dugdale, J., Pavard, B. (2004) "Modelling character emotion in an interactive virtual environment". In proceedings of AISB 2004 Symposium: Motion, Emotion and Cognition. 29 March – 1 April 2004, Leeds, UK.

Guha, R. V. (1991) “Contexts: A Formalization and Some Applications”, Stanford PhD Thesis, 1991.

Kaenampornpan, M., O’Neill, E., Kostakos, V. and Warr, A. (2004). “Classifying Context Classifications: an Activity Theory Perspective”. 2nd UK-UbiNet Workshop, 5-7th May 2004, University of Cambridge, UK.

Karsenty, L., Pavard, B., (1997). « Différents niveaux d’analyse du contexte dans l’étude ergonomique du travail collectif ». Réseaux, 85, 73-99.

Mc Carty J. and Hayes P.J. (1969). “Some philosophical problems from the standpoint of artificial intelligence”. In D. Michie (ed), Machine Intelligence 4, American Elsevier, New York, NY.

Mc Carthy, J., (1993) "Notes on formalizing context". In Proceedings of the thirteenth international joint conference on artificial intelligence, 1993.

McCrae, R. R., and John, O. P. (1992). An introduction to the five-factor model and its applications. Special Issue: The five-factor model: Issues and applications. Journal of Personality 60: 175-215, 1992.

Ortony, A., Clore, G., and Collins, A., (1988).Cognitive Structure of Emotions, Cambridge University Press.

Salembier P. (1996) « Cognition(s) : Située, Distribuée, Socialement partagée, etc ». Bulletin de l'ENS.

Saussure, F. (1916). Cours de linguistique générale. Paris : Payot, 1995 (Première édition 1916).

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communications. Cambridge, UK: Cambridge University Press.