

1) NWO/GBE grant request: **IMMEDIATE**

1a) Project: *Intelligent MultiMEDIA Transactions Environment*

1b) Project Acronym: **IMMEDIATE**

1c) Principal Investigator: **Dr. A. Eliëns**

Address: Dr. A. Eliëns, Faculty of Sciences, Div. of Mathematics and Computer Science, De Boelelaan 1081, 1081 HV Amsterdam, email: eliens@cs.vu.nl

1d) Online Information:

- IMMEDIATE proposal: www.cs.vu.nl/~eliens/research/immediate.html
- research dr. A. Eliëns: www.cs.vu.nl/~eliens/research
- research dr. Z. Huang: wasp.cs.vu.nl
- DLP – www.cs.vu.nl/~eliens/dlp
- RIF – www.cs.vu.nl/~eliens/research/rif.html
- WASP – www.cs.vu.nl/~eliens/research/wasp.html

2) Summary:

IMMEDIATE, which stands for *Intelligent MultiMEDIA Transactions Environment*, concerns the development of a collection of *transaction primitives* for embodied conversational agents in (3D) virtual environments. Such transaction primitives must include facilities for the *interpretation* of questions and commands from a (human) user, *learning* functionality (so that the agent can acquire, in the process of interaction, knowledge about the virtual world and the user), as well as a repertoire of *actions* (that is movements and gestures) with which the agent can respond in an adequate way.

In the IMMEDIATE project, we aim at developing two pilots and two target applications. The first pilot (1) is a *domestic servant*, that inhabits a virtual house and is able to move objects around the house and perform simple duties, such as cleaning the house. The second pilot (2) is a *virtual sales person*, that is able to demonstrate products for sale on demand and may negotiate about particular attributes of the product, such as price and color. As target applications we plan to develop: (1) a cultural heritage application, in particular a virtual environment for material about contemporary art, in which a presentation agent selects material and creates a presentation based on the interests of a user; and (2) a *social role playing* application, where the user may interact with an agent that may take a particular role. For the *social role playing* application, we plan to develop a system with which public transport attendants can gain experience with passengers with a variety of attitudes. For example, passengers might be compliant and hand over their ticket, or aggressive and start vandalizing the transport vehicle.

The IMMEDIATE project must result in a collection of reusable primitives to build applications that require complex interactions with embodied conversational agents.

The software platform used for the realization of the target applications of the IMMEDIATE project is the DLP+X3D platform developed by the *Intelligent Multimedia Group*¹ of the Vrije Universiteit. The DLP+X3D platform combines distributed logic programming (DLP) with 3D virtual environments (X3D/VRML) and allows for the development of multiuser virtual environments with autonomous (conversational) agents. See Eliëns (1992), Eliëns et al. (2002), Huang et al. (2002) and Huang et al. (2003).

3) Classification:

Following NOAG-*i* 2001-2005, the research proposed is related to the following areas:

- IS – Intelligent Systems – adaptive presentation, multi-agent systems

¹ The *Intelligent Multimedia Group* is an informal subgroup of the IM&SE section of the Informatics department of the Faculty of Sciences of the Vrije Universiteit.

- MM – Multimedia – integration of media, user interfaces
- SE – Software Engineering – architecture, object-technology

In the online version links are provided to information about these areas (in Dutch) taken from NOAG-i 1997. As an estimate in percentages, the contribution of the respective fields might be summarized as: 50 (IS), 30 (MM), 20 (SE).

Based on the rather outdated classification of (sub)disciplines of Computer Science, our research could be classified as belonging (to some extent) to: (1.2) distributed systems; (2.3) information retrieval and presentation; (2.7) expert systems; (3.1) software architecture; (3.3) object technology; and (3.9) interoperability.

4) Composition of the Research Team:

| name | expertise | affiliation | hours/week |
|-------------------|------------|-------------|------------------|
| Dr. A. Eliens | multimedia | VU/IMSE | 8 (coordination) |
| Dr. Z. Huang | agents | VU/AI | 8 (WASP/RIF) |
| Drs. C. Visser *) | DLP | VU | 18 (programmer) |
| M. Hildebrand **) | AI/CS | VU | 32-36 (OIO) |
| ... ***) | ... | ... | 2 (promotor) |

The research will be executed within the *intelligent multimedia* group at VU, under the supervision of Dr. A. Eliens² and dr. Z. Huang. *) Drs. C. Visser will provide programming support during the full four years of the project, of which two years will be covered by the requested funding (see 10). **) M. Hildebrand is currently a student at the University of Utrecht and is the proposed candidate for the assistant research position of the IMMEDIATE project. ***) The promotor will be from the SIKS research school.

5) Research School: SIKS

The partners are members of SIKS (the Dutch Research School for Information and Knowledge Systems, www.siks.nl).

6) Description of the Proposed Research

The IMMEDIATE project aims at a reusable collection of *transaction primitives*. By *primitives* we mean concepts that can be realized in identifiable pieces of code, that is objects and functions, that can be reused in a variety of applications. Transaction primitives, in this context, are primitives that support the interaction between a human user and a software agent, or possibly multiple software agents.

We distinguish between three levels of primitives, covering respectively:

- the interpretation of questions and commands – to communicate with user
- learning capabilities – to increase knowledge about world and user
- repertoire of actions – movements and gestures

Together, these primitives should allow for the construction of applications with intelligent conversational agents, that can interact with the user, learn about the user and the (user's interest in the) virtual environment, and perform meaningful actions in that environment.

To develop the collection of transaction primitives, we will start with a pilot application:

- *domestic servant* – with learning capabilities
- *virtual sales person* – with presentation and negotiation skills

In addition, continuing on the results of the pilot applications we will work on two additional target applications, respectively:

² See www.cs.vu.nl/~eliens/cv for his CV.

- *cultural heritage* – presentation agent
- *social role playing* – public transport training

These target applications serve as a vehicle for identifying the major issues in our research, and as demonstrators illustrating the effectiveness of our approach.

research topics

In the interaction with a human user, a software agent is confronted with three problems: the *interpretation* of the user's question and commands, *learning* about the information the virtual environment contains and how the user identifies objects in that environment, and to respond to the user in a meaningful way, using text and *actions* (that is gestures and movements).

interpretation When the user enters a question or command, the agent must use natural language processing to identify what objects the user refers to and what manipulations to those objects are requested. For the correct interpretation of queries and commands, we need to develop a component for parsing and assigning the proper denotations.

learning When the agent is not able to identify the object the user intends or to execute a command requested by the user, feedback and correction is needed from the user. Based on this feedback, the agent must adapt its vocabulary so that future requests can be met immediately. Initially, the agent need not have complete information about the virtual environment, but may instead build up its knowledge in the process of interaction by exploring the virtual environment, that is by searching the information contained in the environment.

actions To respond in a meaningful way, the agent must not only produce text but create a presentation, making use of the expressive facilities of rich media virtual environments. The presentation must include appropriate movements and gestures and possibly the manipulation of objects in 3D space. The development of a suitable repertoire of actions will be based on the work reported in Huang et al. (2003).

target applications

We will use a limited number of target applications as a starting point for identifying the major research issues and also as demonstrators to illustrate the effectiveness of our approach.

We will start our investigations by developing two pilot applications, featuring a *domestic servant* that may move around the house and perform simple tasks, and a *virtual sales person* that may present products and negotiate about price and product-attributes on demand. The next two target applications, respectively a *cultural heritage* application and a *social role playing* application, are increasingly more complex and require considerably more effort to accomplish.

pilot 1 – *domestic servant*: The principal challenge in the *domestic servant* application is to acquire knowledge about the environment and the meaning of commands. We assume that the agent has knowledge about the existence of objects as well as a basic repertoire of actions it can perform to respond to a command. Gradually it will learn the precise denotations and be able to perform simple actions, such as moving an object to another place in the house. Since the number of objects in the house is finite and the repertoire of actions is finite, simple trial-and-error learning with feedback should suffice in this case. For natural language understanding, we will use the NLP-component developed by the proposed candidate during his master thesis project.

pilot 2 – *virtual sales person*: A *virtual sales person* needs to have sufficient presentation skills on the one hand, and must be able to negotiate with the user about price and product-attributes on the other hand. Presentation skills include the ability to demonstrate the product, as well as providing information about the product and its attributes, possibly in a multimedia format (such as digital video). Negotiation may be required when settling a price with the user, in particular when the user asks for modifying product-attributes such as the color. Dependent on the product, the agent must closely interact with the environment to manipulate (the attributes of) the product as requested by the user. Typically, the number of possible

requests in such situations will be limited, so that it suffices to endow the agent with a limited repertoire of actions with which it may respond to the user.

target application 3 – *cultural heritage*: In cooperation with the Dutch Institute for Cultural Heritage³ we will develop a *cultural heritage* application for INCCA⁴ (International Network for the Conservation of Contemporary Art). The INCCA has developed a (multimedia) repository about contemporary art, containing interviews (auditory material), photos and drawings (images), and documents and other written material. The application will support a conversational agent that aids the user in navigating the information space and that will create presentations about the information contained in the INCCA repository on demand, making use of the rich presentation facilities of 3D virtual environments. We will, again, apply feedback learning to adapt the functionality of the presentation agent to the user's requirements.

target application 4 – *social role playing*: Virtual environments with conversational agents provide an excellent vehicle for training social skills. See the section on related work below. We plan to investigate the requirements for a *social role playing* application that allows public transport employees to experience potentially dangerous situations in a virtual environment. The challenge for such applications is to develop a suitably rich repertoire of actions (for the conversational agent) to suggest, for example, potentially aggressive behaviors. These actions include body movements, as well as facial expressions and manual gestures.

implementation platform

In our group we have developed a platform for *intelligent multimedia*, that is a platform for virtual environments based on agent technology, supporting embodied conversational agents, Eliëns et al. (2002).

To effect an interaction between the 3D content and the behavioral component written in DLP, we need to deal with *control points*, and (asynchronous) *event-handling*. The control points are actually nodes in the VRML scenegraph that act as handles which may be used to manipulate the scenegraph.

Our approach also allows for changes in the scene that are not a direct result of setting attributes from the logic component, as for example the transition to a new slide. An event observer is may be used to detect changes in the virtual environment, and to invoke appropriate actions.

The DLP+X3D platform may also be used to realize multi-user virtual environments, Huang et al. (2002).

Recently, we have developed the scripting language STEP to specify humanoid movement and gestures. STEP is based on dynamic logic and implemented on top of the DLP+X3D platform, Huang et al. (2003).

agent model Our agent model is based on an extension of the BDI cognitive model with sensors and effectors, that allow agents to perceive events that occur in the virtual world as well as to operate on the virtual world by sending events. See Huang et al. (2000), Huang et al. (2001). In this way agents can not only control the 3D world they live in, but also their own presentational characteristics, that is their appearance and attributes they possess. For the research described above, we need to augment the agent model with characteristics of a virtual actor for effectively deploying embodied conversational agents in 3D virtual environments.

related work

Our work is related to or a number of other projects. We will discuss:

- Vcom3D⁵ – Signing Avatar and Virtual Role Playing
- PAR⁶ – Parameterized Action Representation
- Parlevink⁷ – agents in VRML worlds
- Alice⁸ – 3D interactive programming environment

³www.icn.nl

⁴www.incca.org

⁵www.vcom3d.com

⁶hms.upenn.edu/software/PAR/par.html

⁷parlevink.cs.utwente.nl

⁸www.alice.org

Vcom3D is a small company that developed technology for 3D character animation. This technology provides interactive, 3D animated characters that communicate through lip-synched speech, body language (including gesture and facial expression), and action. They also developed *virtual role playing applications*, in particular a training applications for police officers to deal with high school students. The use of embodied characters for possibly aggressive kids allowed the developers to include characteristic behaviors, such as squirky movements and turning away the eyes, to prepare the police officers for potential conflicts. (This information is not on the Vcom3D web site but is obtained by personal communication with the staff.)

The PAR project investigates *parameterized action representations* for virtual humans in 3D space, Badler et al. (1998). They introduced the notion of an *Actionary* as a repository of basic actions, parametrized by the executing agent(s) and the objects involved. One application of their approach is a multi-user virtual environment where users control the behavior of their avatars by giving simple commands. These commands are then effected by instantiating one or a collection of the primitives from the *Actionary* (which may be regarded, according to the authors, as a *dictionary* for actions). Despite the apparent similarities, our approach is significantly different from the PAR approach by providing support for generic actions expressed in a logic-based formalism, thus allowing for learning capabilities in addition to the interpretation of commands and the execution of (instantiated) actions.

The Dutch Parlevink project, at the University of Twente, has developed a virtual theater, which demonstrates that agents may usefully be employed in (VRML-based) virtual environments. The focus of the Parlevink project, however, is more on cognitive models and natural language processing. Our goal is, in addition, to develop a suitable library of transaction primitives, including movements and gestures, to improve the communication capabilities of our conversational agents.

Finally, we want to mention Alice, a 3D interactive programming environment for virtual worlds, developed at Carnegie-Mellon University. One interesting feature of Alice is that it allows for programming generic behavioral properties in the object-oriented scripting language Python. Nevertheless, we believe that due to the knowledge-intensive nature of the programming tasks required for developing real applications in rich media virtual environments, a logic-based language such as DLP will in the long end be much more effective.

embedding in research: (intelligent) multimedia

Over the past six years, our research efforts have focussed on developing models and software architectures for multimedia and hypermedia applications. (A full version of this description, including publications, is available online in the SIKS report (2001)).

This research has resulted in two Ph.D. theses:

- 10/4/2001 – Structured hypermedia – a matter of style
- 8/5/2001 – Diva: Architectural Perspectives on Information Visualization

For both theses, prof. dr. J.C. van Vliet was promotor, and the principal investigator had daily supervision and acted as co-promotor.

The hypermedia work was done in collaboration with dr. L. Hardman and dr L. Rutledge from the CWI Multimedia group. This cooperation resulted in the formalization of the Amsterdam Hypermedia Model, an extension of the Dexter Hypertext Reference Model. The hypermedia project also resulted in a software framework for developing web-based hypermedia applications, the *hush* library and its music and video extensions. See Eliëns (2000), Eliëns (1995), Eliëns *et al.* (1997b) and Eliëns *et al.* (1997).

The visualisation project concerned the use of animations and visualisation to display business process simulation results in a hypermedia context. During the project the focus shifted towards visualisation, in particular business visualisation. See Schönhage and Eliëns (1998) and Schönhage *et al.* (1998). Also, explorations were done to investigate interactive visualisation in 3D. See Schönhage and Eliëns (1999) and Schönhage et al. (2000).

Our current research efforts are directed towards developing a high-level platform for rich media 3D virtual environments. Our goal is to study aspects of the deployment and architecture of virtual environments as an interface to multimedia information systems. See van Ballegooij and Eliëns (2001).

Our research has been supported by two NWO projects:

- WASP – Web Agent Support Program – framework for Web Agents built on DLP
- RIF – Retrieval of Information in virtual worlds using Feature detectors – information retrieval and delivery in 3D multi-user virtual environments

The combined effort of these projects led to the DLP+X3D platform and the development of the STEP language. See Huang et al. (2002) and Huang et al. (2003).

The *intelligent multimedia* research theme may be regarded as continuing the hypermedia and visualisation projects described before. Our efforts are directed towards realizing the technology needed for developing *intelligent multimedia* applications. In particular, we aim for developing demonstrators in the area of *embodied conversational agents*. See Eliëns et al. (2002). This work is being done in cooperation with dr. Z. Ruttkay from CWI. The work on the *cultural heritage* application will be done in cooperation with drs. T. Scholte from the ICN (Dutch Institute for Cultural Heritage).

Apart from publications in international conference proceedings, we have demonstrated our work on the ICT Kenniscongres 2002 in the theme *intelligent multimedia*.

7) Work Programme

A brief summary of the issues that will be tackled and the deliverables that we expect to produce during the four years of the project looks as follows.

- year 1: identification of major issues – *pilot applications 1 & 2*
- year 2: *interpretation and learning* – target application 3
- year 3: *repertoire of actions* – target application 4
- year 4: wrap up & thesis

In the first stage of the research, the work to be done by the proposed candidate, Michiel Hildebrand, will be a follow-up on his master thesis, which describes a first exploration of the use of a natural language interface, and interpretation and learning mechanisms for embodied conversational agents developed with the technology of our *intelligent multimedia* group. The initial idea is to endow agents with a basic repertoire of actions and a set of capabilities that allow the agent to understand commands from the user and explore and manipulate properties of the world. Capabilities may be represented as (*condition, action*) pairs. To deal with the world, the agent must also maintain a set of beliefs about the world, recording the knowledge of the agent at a particular point in time. However, to be able to acquire information about the world, the objects in the world must be annotated with meta-information, which must somehow be accessible to the agent.

In his master thesis, Michiel Hildebrand explored a sequence of scenarios of increasing complexity, starting with a scenario which allowed only direct commands and a fixed number of objects about which the agent had full knowledge, to scenarios that required interpretation and desambiguation to understand commands, exploration of the world to identify objects, and learning to be able to respond to commands. In the most complex scenario, capabilities themselves are considered information (or beliefs), thus allowing for acquiring dynamically new behaviors in the course of interacting with the user. In analogy with the *level of detail* rendering technique (which results in rendering distant objects with less detail), we also explored the notion of *cognitive level of detail*, which amounts to giving access to more information about objects with increasing proximity (relative to the agent).

The candidate researcher will work in close cooperation with the other members of the *intelligent multimedia* research group (consisting of dr A. Eliëns, dr. Z. Huang and drs. C. Visser), to explore the topics mentioned and to further enhance the *intelligent multimedia* technology that is being developed. [STEP2], Eliëns et al. (2003)

education track

The advanced education of the candidate researcher will mainly be taken care of by the standard courses offered by the SIKS research school. (See www.siks.nl) In addition other courses may be advised when the occasion arises.

8) Instrumentation

Not relevant.

9) Literature

The online version has links to online copies of most of the cited references. For a full list of references, see below.

10) Requested Budget

| personell | period | euro |
|------------|--------|---------|
| OIO | 4 year | 135.762 |
| programmer | 2 year | 112.138 |
| | | 247.900 |

Remarks:

- the amount of euro 135.762 includes the benchfee for travel expenses and other support.
- the VU will provide an additional 2 years of programmer support.

References

[XXX] STEP2 missing reference

- [Badler et al. (1998)] Badler N., Bindiginavale R., Bourne J., Palmer M., Shi J., Schuler W. (1998), A Parameterized Action Representation for Virtual Human Agents, Workshop on Embodied Conversational Characters, WECC98, Lake Tahoe, CA, Oct 12-15, 1998.
- [Eliëns (1992)] Eliëns A. (1992), DLP – A language for Distributed Logic Programming, Wiley
- [Eliëns (1995)] Eliëns A. (1995), Hush – a C++ API for Tcl/Tk, *The X Resource*, Issue 14, April, pp. 111-155
- [Eliëns (2000)] Eliëns A. (2000), *Principles of Object-Oriented Software Development*, Addison-Wesley Longman, 2nd edn.
- [Eliëns et al. (2003)] Eliëns A., Dormann C., Huang Z. and Visser C. (2003), A framework for mixed media – emotive dialogs, rich media and virtual environments, Proc. TIDSE03, 1st Int. Conf. on Technologies for Interactive Digital Storytelling and Entertainment, Gööbel S. Braun N., n Spierling U., Dechau J. and Diener H. (eds.), Fraunhofer IRB Verlag, Darmstadt Germany, March 24-26, 2003
- [Eliëns et al. (2002)] Eliëns A., Huang Z., and Visser C. (2002), A platform for Embodied Conversational Agents based on Distributed Logic Programming, AAMAS Workshop – Embodied conversational agents - let's specify and evaluate them!, Bologna 17/7/2002
- [Eliëns et al. (1997b)] Eliëns A., van Ossenbruggen J.R. and Schönhage S.P.C. (1997b), Animating the Web — an SGML-based approach . In: *The Internet in 3D – Information, Images and Interaction*, Academic Press, pp. 75-96
- [Eliëns et al. (1997)] Eliëns A., Welie M. van, Ossenbruggen J. van and Schönhage S.P.C (1997) , Jamming (on) the Web. In: Proceedings of the 6th International World Wide Web Conference — Everone, Everything Connected, O'Reilly and Associates, Inc. April 1997, pp. 419-426
- [Huang et al. (2001)] Huang Z., Eliëns A., and De Bra P. (2001), An Architecture for Web Agents, Proceedings of the Conference EUROMEDIA 2001, 2001.
- [Huang et al. (2000)] Huang Z., Eliëns A., van Ballegooij A., De Bra P. (2000), A Taxonomy of Web Agents, IEEE Proceedings of the First International Workshop on Web Agent Systems and Applications (WASA '2000), 2000.
- [Huang et al. (2002)] Huang Z., Eliëns A., Visser C. (2002), 3D Agent-based Virtual Communities. In: Proc. Int. Web3D Symposium, Wagner W. and Beitler M. (eds), ACM Press, pp. 137-144

- [Huang et al. (2003)] Huang, Z., Eliëns, A., and Visser, C., *STEP: a Scripting Language for Embodied Agents*, in: Helmut Prendinger and Mitsuru Ishizuka (eds.), *Life-like Characters, Tools, Affective Functions and Applications*, Springer-Verlag, 2003.
- [Schönhage *et al.* (1998)] Schönhage B., Bakker P.P. and Eliëns A. (1998) , So many users - so many perspectives, *IFIP 12.2 Working Conference on Designing Effective and Usable Multimedia Systems*, Fraunhofer Institute, Stuttgart Germany, 9-10 September, Kluwer Academic, pp. 159-172
- [Schönhage and Eliëns (1998)] Schönhage S.P.C. and Eliëns A. (1998), Multi-user visualization: a CORBA/Web-based approach. Int. Conf. on Digital Convergence: the future of the Internet and WWW, British Computer Society, Bradford, UK, 20-23 april 1998, 9 pgs
- [Schönhage and Eliëns (1999)] Schönhage S.P.C. and Eliëns A. (1999), Dynamic and Mobile VRML Gadgets. In Proc. VRML99, 23 - 26 February 1999, Paderborn, Germany
- [Schönhage et al. (2000)] Schönhage S.P.C., van Ballegooij A. and Eliëns A. (2000), 3D Gadgets for Business Process Visualization: a case study, Proc. Int. Web3D/VRML Conference - 2000, Monterey CA, Febr 2000, ACM Press pp. 131-138
- [van Ballegooij and Eliëns (2001)] van Ballegooij and Eliëns A. (2001), Navigation by Query in Virtual Worlds, Web3D 2001 Conference, Paderborn, Germany, 19-22 Feb 2001