A framework for mixed media – emotive dialogs, rich media and virtual environments

Anton Eliëns, Claire Dormann, Zhisheng Huang and Cees Visser

Intelligent Multimedia Group
Vrije Universiteit, Amsterdam, The Netherlands
{eliens,claire,huang,ctv}@cs.vu.nl

Abstract. We present a framework for merging mixed media in a unified fashion. Our framework supports 3D virtual environments, rich media (such as digital video) and (emotive) dialogs presented by humanoid characters or simply text balloons. We speak of *mixed media* since each of these media may have a clearly distinct narrative structure. In this paper, we will discuss the background and motivation of our approach, which is the presentation of instructional material using (rich media) 3D slides, enhanced with (possibly ironic) comments presented by humanoid characters. And, we will explore the design space of such *mixed media* presentations, that is the issues involved in merging material with potentially conflicting narrative structures. In addition, we will look at the style parameters needed to author such presentations in an effective way, and we will briefly describe the implementation platform used to realize the presentations.

Keywords and phrases: presentation technology, persuasion, mixed media, dialogs, virtual worlds, rich media

1 Introduction

Some phenomena in the media are older than is generally recognized. As observed in [1], some of the conventions of the 20th century comic books draw directly or indirectly on an even longer visual tradition:*speech balloons can be found in the eighteenth century prints, which are in turn an adaptation of the 'text scrolls' coming from the mouths of the Virgin and other figures in medieval religious art.* In other words, mixed media have their origin in the medieval era, where both visual material and textual material were part of the rethorics of institutionalized religion.

Our interest in *mixed media*, that is the use of speech ballons with text together with rich media and 3D virtual environments, stems from developing instructional material using 3D technology. We enhanced the material, which was organized in sequentially ordered slides, with comments presented by humanoid characters. These comments would sometimes contain additional information, and were sometimes plainly ironic, anticipating on students' comments. Obviously, the dialogs were meant to draw the attention to particular aspects, and more generally to increase the emotional involvement of the students with the material. Experimenting with the way comments could be added to a slide, we found that the humanoids were often not necessary. Also, to avoid interference with the material presented in the slides, it was often necessary to place the speech ballons somewhat off-center, for example in the lower right corner.

In this paper, we explore the design space of mixed media presentations, which involve the juxtaposition of emotive dialogs, rich media such as digital video and virtual environments. We will discuss the issues involved in merging material with distinct, potentially conflicting, narrative structures, and we will look at the style parameters needed to author mixed media presentations in an effective way.

Structure The structure of the paper is as follows. In section 2, we will briefly present the background and motivation to our approach, with an example. In section 3, we will discuss narrative structure from the perspective of persuasion and emotional involvement and compare our approach to related work. In section 4, we will explore the design space of mixed media and study a variety of combinations at the hand of representative examples. In section 5, we will look at the authoring issues and discuss the style parameters we used. In section 6, a brief description is given of the implementation platform and in section 7 we will discuss research directions and open issues.

2 Background and motivation

Desktop VR is an excellent medium for presenting information, for example in class, in particular when rich media or 3D content is involved.

At VU, we have been using *presentational VR* for quite some time, in a course on Web3D technology and also in our *introduction multimedia* course.

Recently we included dialogs using speech balloons (and possibly avatars) to display the text commenting on a particular presentation.

A dialog is (simply) a sequence of phrases for two (virtual) speakers. Each speaker, alternately, may deliver a phrase. When delivering a phrase the speaker may step forward, dependent on the style of presentation chosen. The dialog text (and avatars) are programmed as annotations to a particular scene as described in more detail in section 5.

Each presentation is organized as a sequence of slides, and dependent on the slides (or level within the slide) a dialog may be selected and displayed.[1] To be more precise, when the presenter goes to another slide an *observer* object checks whether there is a dialog available, and if so the dialog is started. See section 6, discussing the platform used for realizing the presentations.

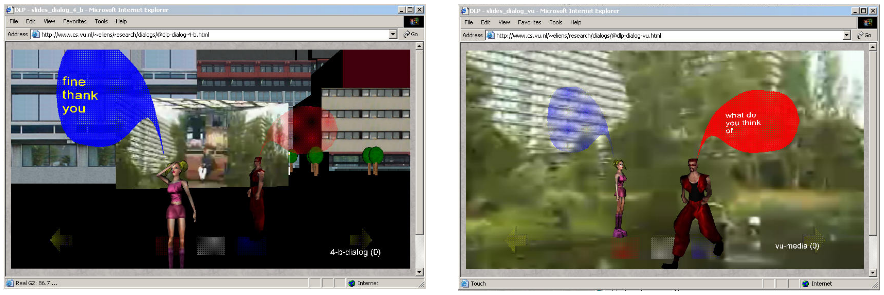
example – promotion video in virtual environment In this example, we used a promotion video produced to attract students to our university. In figure (a) the promotion video is embedded in a virtual environment

[1] See the appendix for a description of the slides format.

of our university campus. In figure (b) only the video is shown. The dialog used is, in both examples, a somewhat ironic comment on the contents of the movie displayed.

In figure (a), you see the left avatar (named *cutie*) step forward and deliver her phrase. This dialog continues until *cutie* remarks that she *always wanted to be an agent*. In figure (b), you see the right avatar (named *red*) step forward asking *what do you think of* to discuss what it takes to be a student.

Note that in example (a) the avatars are more or less natural inhabitants of the virtual environment, whereas in example (b) they are just lain on top of the movie. In the latter case, the positioning of the avatars may easily seem unnatural, although it sometimes works out surprisingly well.


DLP - slides_dialog_4_b - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.cs.vu.nl/~eliens/research/dialogs/@dlp-dialog-4-b.html
Go
fine thank you
4-b-dialog (0)
Real G2: 86.7 ...
Internet
DLP - slides_dialog_vu - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.cs.vu.nl/~eliens/research/dialogs/@dlp-dialog-vu.html
Go
what do you think of
vu-media (0)
Touch
Internet

(a) *dialog in context*

(b) *dialog on video*

3 General perspectives – narrative structure and persuasion

Mixed media, that is the combination of dialogs with rich media and virtual environments, may endanger the narrative structure of a presentation. Each of these media entities may have a distinct narrative structure. Slides are sequentially organized, and each slide may have levels that are displayed in a sequential fashion. The narrative structure of digital video may be arbitrarily complex, and may make full use of cinematographic rethorics. Navigation and interaction in virtual environments may be seen as a weak narrative structure, which may however be strengthened by guided tours or viewpoint transformations, taking the user to a variety of viewpoints in a controlled manner. Finally, dialogs have a well-defined temporal structure, with alternating turns for the two (virtual) speakers.

In the following we will compare our approach to Related research and investigate what possible advantages we may obtain from *mixed media* presentations.

consonant or dissonant comments

Our approach is clearly reminiscent to the notorious *Agneta & Frida* characters developed in the Persona[2] project.

The Persona project aims at: investigating a new approach to navigation through information spaces, Based on a personalised and social navigational paradigm, [2]. The novel idea pursued in this project is to have agents (Agneta and Frieda) that are not helpful, but instead just give comments, sometimes with humor, but sometimes ironic or even sarcastic comments on the user's activities, in particular navigating an information space or (plain) web browsing.

In contrast with the Personas project, our (dialog) comments are part of a presentation which of itself has a definite narrative structure, in opposition to the 'random' navigation that occurs by browsing 'information spaces'.

As a consequence, our comments may be designed taking the expected reaction of the audience into consideration. An interesting question is whether comments should be *consonant* with the information presented (drawing attention to perticular aspects) or *dissonant* (as with ironic or sarcastic comments).

engagement versus immersion

The characters and dialog text may be used to enliven the material. In this way, the students' engagement with the material may be increased, [3].

Clearly, there is a tension between engagement and immersion. Immersion, understood as the absorption within a familiar narrative scheme (in our case the lecturer's presentation), may be disrupted by the presence of (possibly annoying) comments, whereas the same comments may lead the attention back to the material, or provide a foothold for affective reactions to the material, [4]. Also, the audience might start to anticipate the occurrence of a dialog and possibly identify themselves with one of the characters.

emotional enhancement

In one of her talks, Kristina Höök observed that some users get really fed up with the comments delivered by Agneta and Frieda. Nevertheless, it also appeared that annoyance and irritation increased the emotional involvement with the task. For our presentations, we may ask how *mixed media* may help in increasing the emotional involvement of the audience or, phrased differently, how dialogs may lead to emotional enhancement of the material, [5].

An important difference with the Personas project is that our platform supports the actual merge of dialogs and the humanoid characters that deliver them in a unified presentation format, that is a rich media 3D graphics format based on X3D/VRML. As a consequence, the tension between immersion and engagement may be partially resolved, since the characters delivering the dialog may be placed in their 'natural' context, that is a virtual environment as in example (a).

[2] http://www.sics.se/humle/projects/persona/web

4 Design space – the juxtaposition of mixed media

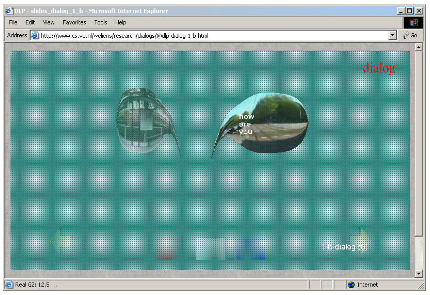
In this section, we will explore in a more systematic way what options we have in creating *mixed media* presentations. We have a division according to levels of complexity, with at level 0 the basic material, that is dialog text, rich media objects such as digital video and virtual environments. The other levels arise by adding dialogs to either the media object or the virtual environment. We also allow for the virtual speakers of the dialog to change attributes of the presentation, for example by depositing objects in the (virtual) environment,

In summary, we distinguish between the following levels of complexity:

- level 0: *basic material* – dialog text, media object virtual environment
- level 1: *combined* – media object + dialog
- level 2: *with avatars* – media object + dialog + avatars
- level 3: *with attributes* – media object + dialog + avatars + objects
- level 4: *with context* – media object + dialog + avatars + context

The ordering of these levels is not unique, but admittedly depends on our intuition of complexity.

For each of these levels we have a collection of representative examples. [3]

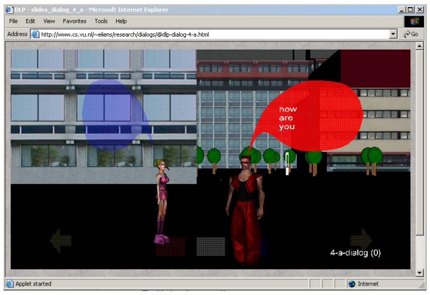


DLP - slides_dialog_1_b - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address
Go
dialog
how are you
1-b-dialog (0)
Internet

Address: http://www.cs.vu.nl/~eliens/research/dialogs/@dlp-dialog-1-b.html

Real G2: 12.5 ...

(c) *context in dialog*



DLP - slides_dialog_4_a - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address
Go
how are you
4-a-dialog (0)
Internet

Address: http://www.cs.vu.nl/~eliens/research/dialogs/@dlp-dialog-4-a.html

Applet started

(d) *dialog in context*

level 1 – combined:

When combining the media object with the dialog, we may simply superpose the dialog on the movie, in a similar way as shown in figure (b), or we may project the video onto the surface of the speech balloon, as illustrated in figure (c), which has a rather surprising effect.

level 2 – with avatars: It makes quite a difference whether humanoid characters are used to deliver the dialog or only the plain text balloons. The use of avatars seems to enhance the recognition of a particular role, that is the kind

[3] http://www.cs.vu.nl/∼eliens/research/media/paper-dialogs.html

of comments the character makes. Once the role is established, the avatars may dissappear and the (color and position) of the speech ballons suffice. Another difference is that the ballons should be positioned differently without the presence of an avatar.

level 3 – with attributes Apart from speaking their dialog text, the avatars may undertake autonomous actions. They might get bored and display ambient behavior, like looking on their watch. In addition, before or after speaking their phrase, they may change their position and modify the environment, for example by depositing (3D) objects to illustrate their comments. These actions may be arbitrarily complex, and for example result in going to the next (level of the) slide.

level 4 – with context:
The ultimate context of a slide may be a virtual environment, as in our examples (a) and (d) a virtual environment of our university campus developed by one of our students. Having such a context, the difference between a presentation and an information space becomes blurry since the user may interact with the environment and, for example, start a guided tour. In such cases the developer must decide whether the dialog takes place in a fixed position of the virtual environment or is fixed relative to the viewpoint of the user/audience.

Merging the dialogs with the virtual environment and the media object might, by the way, lead to overly complex presentations such as the one depicted in figure (a).

5 Authoring issues – style parameters

In the next two sections, we will look at the implementation of the dialogs, respectively from an authoring perspective and a system perspective.

The authoring of a dialog for a particular dialog or slide should be easy. The encoding of the dialog used in the examples discussed is illustrated below.

```
<phrase right="how~are~you"/>
<phrase left="fine~thank~you"/>
<phrase right="what do~you think~of studying ..."/>
...
<phrase left="So,~what~are you?"/>
<phrase right="an ~agent" style="[a(e)=1]"/>
<phrase left="I always~wanted to be~an agent" style="[a(e)=1]"/>
```

The phrases are (textually) included in a *slide*, which is itself indicated by appropriate begin and end tags. The alternation between speakers is indicated by the attributes *left* and *right*. Although detailed indications of (among others) when a phrase should be uttered are possible, these advanced options are hardly ever used, except for defining complex actions.

Furthermore, there are a number of style parameters that may be used to decide for example whether the avatars or persona are visible, where to place the dialogs balloons on the display, as well as the color and transparancy of the balloons. To this end, we have included a *style* attribute in the *phrase* tag, to allow for setting any of the style parameters.

style parameters

```
<phrase right="red" style="[p=(0.5,0,0),persona=0,balloon=0]"/>
<phrase left="cutie" style="[p=(-0.5,0,0),persona=0,balloon=0]"/>
<gesture right=1 style=default/>
<gesture left=1 style=default/>
```

Apart from phrases, we also allow for gestures, taken from the built-in repertoire of the avatars. In [6], we discuss how to extend the repertoire of gestures, using a gesture specification language.

Both phrases and gestures are compiled into DLP code and loaded when the annotated version of the presentation VR is started. See section 6.

6 Implementation – the DLP+X3D 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, [7]. Our platform merges X3D/VRML[4] with the distributed logic programming language DLP.

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 actually used to detect the transition to another slide. A dialog may then be started to comment on the contents of that particular slide.

The DLP+X3D platform may also be used to realize multi-user virtual environments, [8]. About how a multi-user environment might be used to explore narrative structures, we can only speculate!

7 Research directions – conversational agents

In the examples discussed, the agent avatars entered a dialog with one another to comment on a particular scene or slide, to augment a presentation. As a next step, we would like to extend our approach to allow for interaction in a virtual environment, that is to augment information spaces.

[4] http://www.web3d.org

virtual musea As an example, for a virtual gallery for Escher, we can use dialogs to enhance visitors understanding and enjoyment of the work of Escher on display. We can increase viewers' involvement by directing their attention to salient features of the painting or the exhibition. By giving information on the ideas that motivate the painting, its creation or the life of Escher, we can draw viewers more effectively into the world of Escher.

In addition, conversational agents can assist the user in finding a particular work and, in particular for the 'impossible spaces' of Escher, offer means to experiment with the graphical attributes of the world.

cultural heritage Another application where we can profit from the rich presentation facilities of desktop VR is the construction of a virtual environment for cultural heritage. As an example, think of the database of INCCA[5] (International Network for the Conservation of Contemporary Art), which contains interviews (auditory material), photos and drawings (images), and documents and other written material. We could well imagine to have (conversational) presentation agents to present the information to the user. Dialogs may be used to present various viewpoints on a particular oeuvre.

8 Conclusions

We have described a framework for *mixed media* that allows for the superposition of (text) dialogs delivered by humanoid avatars and/or speech balloons, on arbitrary rich media objects and virtual environments.We have looked at the design space of *mixed media* presentations, by discussing a number of representative examples, each illustrating a particular level of complexity. Also authoring issues were discussed, and an indication was given of the style parameters needed to develop effective presentations.We have further described the implementation platform used to realize the *mixed media* presentations and explored what new applications and extensions are feasible.

References

1. Briggs A. and Burke P. (2001), *A social history of the media – from Gutenberg to the Internet* , Polity Press
2. Munro A., Höök, K. and Benyon D.R. (1999), Footprints in the Snow. In: Social Navigation of Information Space, Springer
3. Morgan (2000), Theory and models for creating engaging and immersive ecommerce websites, Proc. of the 2000 ACM SIGCPR Conf. on Computer Personnel Research, April 2000, pp. 77-85
4. Dijkstra K., Zwaan R. Graesser A.and J. Magliano (1994), Character and reader emotions in literary texts, Poetics 23, pp. 139-157

[5] http://www.incca.org

5. Astleneir H. (2000), Designing emotionally sound instruction: the FEASP approach. Instructional Science 28, pp. 169-198
6. Huang Z., Eliëns A., Visser C. (2002b), STEP – a scripting language for Embodied Agents, PRICAI-02 Workshop – Lifelike Animated Agents: Tools, Affective Functions, and Applications, Tokyo, 19/8/2002
7. Eliëns A., Huang Z., and Visser C., 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
8. 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

Appendix: The slides format

In this technical appendix, a simplified description will be given of the slides format. Slides are fragments of a document that may be presented to an audience. Our approach allows for displaying slides in either dynamic HTML or VRML. Slides are encoded using XML[6].

slides in XML

```
<document>
...
<slide id=”1”>
<text>
<line>What about the slide format?</line>
<break/>
<line>yeh, what about it”?</line>
</text>
<vrml>Sphere { radius 0.5 }</vrml>
</slide>
...
<slide id=”2”>
<vrml>Sphere { radius 0.5 }</vrml>
</slide>
...
</document>
```

The first slide contains some text and a 3D object. The second slide contains only a 3D object. Inbetween the slides there may be arbitrary text. The slides are converted to VRML using XSLT, the XML transformation language.

VRML PROTOs

To support slides in VRML a small collection of PROTO definitions is used.

[6] http://www.xml.org

- *slideset* – container for slides
- *slide* – container for text and objects
- *slide* – container for lines of text
- *line* – container for text
- *break* – empty text

The slides contained in a document constitute a slide set. A slide set is a collection of slides that may contain lines of text and possibly 3D objects. For displaying 3D objects in a slide we need no specific PROTO.

The *slide* PROTO defines an interface which may be used to perform spatial transformations on the slide, like translation, rotation and scaling. The interface also includes a field to declare the content of the slide, that is text or (arbitrary) 3D objects.

The *slideset* contains a collection of slides, and allows for proceeding to the next slide. A *text* may contain a sequence of lines and breaks. It supports a simple layout algorithm.

annotated slides

Slides may be annotated with *dialogs*, as described in section 5. The annotation is compiled to DLP code, which is activated whenever the slide (or level within a slide) to which the annotation belongs is displayed.

To intercept the occurrence of a particular event, such as the display of a slide, we use an *observer* object which is specified as in the code fragment below.

observer

```
:- object observer : [actions].
var slide = anonymous, level = 0, projector = nil.

observer(X) :-
   projector := X,
   repeat,
     accept( id, level, update, touched),
   fail.

id(V) :-  slide := V.
level(V) :- level := V.
touched(V) :- projector←touched(V).
update(V) :- act(V,slide,level).
:- end_object observer.
```

The observer object has knowledge of, that is inherits from, an object that contains particular actions.

As indicated before, events come from the 3D scene. For example, the *touched* event results from mouse clicks on a particular object in the scene. On accepting an event, the corresponding method or clause is activated, resulting in either changing the value of a non-logical instance variable, invoking a method, or delegating the call to another object.