

A Dual-Mode User Interface for Accessing 3D Content on the World Wide Web

Jacek Jankowski and Stefan Decker
Digital Enterprise Research Institute, NUI Galway, Ireland
firstname.lastname@deri.org

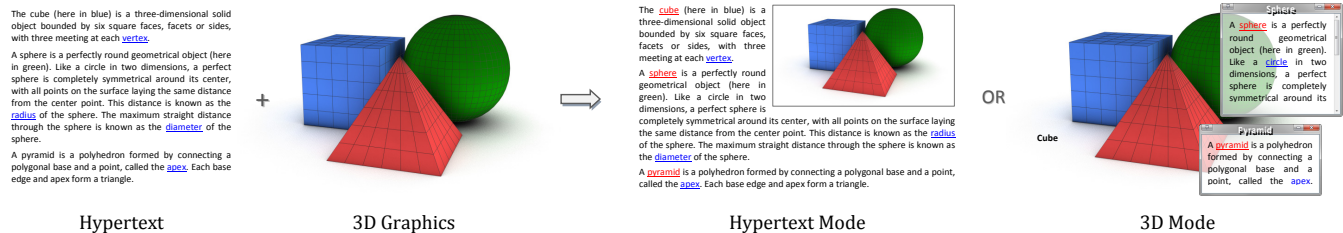


Figure 1: How to design a user interface that combines hypertext and 3D graphics? The integrated information presented in the hypertext mode (left) and the 3D mode (right) of the dual-mode user interface.

ABSTRACT

The Web evolved from a text-based system to the current rich and interactive medium that supports images, 2D graphics, audio and video. The major media type that is still missing is 3D graphics. Although various approaches have been proposed (most notably VRML/X3D), they have not been widely adopted. One reason for the limited acceptance is the lack of 3D interaction techniques that are optimal for the hypertext-based Web interface. We present a novel strategy for accessing integrated information spaces, where hypertext and 3D graphics data are simultaneously available and linked. We introduce a user interface that has two modes between which a user can switch anytime: the driven by simple hypertext-based interactions *"don't-make-me-think"* mode, where a 3D scene is embedded in hypertext and the more immersive 3D *"take-me-to-the-Wonderland"* mode, which immerses the hypertextual annotations into the 3D scene. A user study is presented, which characterizes the user interface in terms of its efficiency and usability.

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]: User Interfaces, Hypertext and Hypermedia; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

General Terms

Design, Human Factors

Keywords

Hypertext, 3D Graphics, 3D Web, User Interface

1. INTRODUCTION

The Web evolved from a text-based system to the current rich and interactive medium that supports images, 2D graphics, audio and video. The major media type that is still missing is 3D graphics. As computer graphics technology has reached the point where 3D models can be rendered, often in real time on commodity desktop and mobile devices, at a fidelity that is almost indistinguishable from the real thing, it should be possible to use 3D models rather than 2D images to represent various objects on the Web.

There have been a number of approaches over the last years to integrate 3D technologies on the Web and most of these systems and standards disappeared or barely survived (e.g., [23]). We argue that this is because of the fact that research was focused mostly on 3D graphics and 3D graphics alone. The focus of research did not include the search for user interaction techniques that are optimal for the hypertext-based Web interface. However, what good is a realistic environment if one cannot interact with it? As a result, hypertext (the ultimate product in symbolic communication) and interactive 3D graphics (the ultimate achievement in visual media) are at odds on the Web.

We believe that people can gain a lot from using integrated information spaces where hypertext and 3D graphics data are simultaneously available and linked. This paper focuses on user interface design that supports the integrated exploration of such environments; the design, where users can browse the text, look through general information and search for more specific information, and where they can also navigate freely through a 3D space, and examine and manipulate virtual 3D objects, to gain a better understanding of the data. The objective of our user interface is to pair interactive 3D graphics know-how with well established UI conventions of the Web to support all these user tasks. We believe that this issue is of a great importance, since there are upcoming new open (WebGL) and proprietary (Stage3D) proposals for 3D graphics APIs in the Web context.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2012, April 16–20, 2012, Lyon, France.

ACM 978-1-4503-1229-5/12/04.

2. FACTORS DRIVING THE DESIGN

There were five main driving factors in designing our UI:

2.1 User’s Primary Tasks

A prerequisite to the effective UI design is an understanding of the users and their tasks - the tasks for which that interface will actually be used. In [16] we focused on an understanding of the fundamental tasks users may engage in while exploring Web-based 3D virtual environments. We constructed a “3D Web Taskonomy”, where we divided these tasks into hypermedia tasks, such as following hyperlinks, and virtual environment tasks, such as 3D navigation. We also included a review of mouse-based 3D interaction techniques useful in the context of 3D Web.

2.2 Varying User Experience Levels

One of our major goals was to provide an interface that meets the needs of both novice and experienced users. We assumed that most users would be new to 3D interactions. We therefore needed to design our UI in favor of novices. This meant adhering to well-established UI conventions and making 3D navigation as simple as possible. On the other hand, there are 3D knowledgeable users who can find the limitations and constraints of a novice interface frustrating. As we did not want to diminish the 3D interaction experience in any way, we needed to provide these expert users with much more freedom with regards to the 3D task support.

2.3 Multimedia Principles

Multimedia presentation was studied extensively within psychology, aiming at extracting principles that guarantee an effective design and facilitate learning; the central claim of multimedia is that providing information in more than one medium of communication will lead to better understanding [21]. The theory based on *Dual channel* [26], *Limited capacity* [6], and *Active processing* [35] assumptions suggests that if active processing takes place in both visual and verbal cognitive subsystems, learning is improved; dual coding of information is more effective than single coding; it is also critical that both visual and verbal representation are actively constructed, together with the connections between them. Supporting multimedia theory, studies have shown that verbal and visual representations in combination are often more effective than either in isolation [24, 10, 22]. On the other hand, Nielsen [25] warns that unconstrained use of multimedia can result in UIs that confuse users and make it harder for them to understand the information. Therefore, we guided our design based on the basic principles for designing multimedia learning environments [21].

2.4 Existing Body of UI Design Work

In this section we survey the work that has been done in the area of UI design for information spaces where hypertext and 3D graphics data are simultaneously available and linked. Intermedia [36] and Hyper-G [1] were probably the first hypermedia systems that integrated 3D documents. These systems were window-oriented, which means that they used document clusters to form groups of related documents and all document types stayed separated in their own windows. As a result, they were not bound to a particular UI metaphor. In contrast, the Web is a multimedia document based hypermedia system. Its user interface is based on single doc-

uments (HTML web pages) consisting of several parts of information of different types. The documents are designed by a web designer, who is responsible for placement of texts and media elements, and the overall aesthetics of a site.

Currently, in order to view and interact with 3D graphics in a web browser, a special browser plug-in (e.g. VRML/X3D, Flash, Unity3D, Java3D) is required that allows the 3D scene and UI controls to be displayed within the web browser window. These plug-ins usually provide users with the means for navigation through a 3D scene: on the one hand, they implement only one 3D navigation technique - the technique that is best suited for a given task; on the other hand, VRML/X3D browsers offer multiple methods of interaction based on examine, fly, walk and fly-to. The first approach limits interaction for the sake of simplicity. The second offers more freedom in terms of viewpoint control. Our goal is to combine the most useful features of these approaches.

Another related work is the research on integrating perceptual and symbolic information in VEs, and the further work on Information-Rich VEs (IRVEs) [3, 4, 5, 28]. IRVEs combine the power of VEs and information visualization, augmenting VEs with additional abstract information such as text, numbers, or graphs. IRVE applications show promise in increasing the utility of the VE experience [4]. In one of IRVE experiments evaluating depth and association cues between objects and their labels, Polys et al. [28] showed that screen-space interfaces outperformed object-space layouts. Therefore, we decided to use solely screen-space techniques for displaying annotations in 3D.

Another closely related work was carried out by a group of researchers under the direction of Thomas Strothotte [13, 14, 15] on labeling and annotating 3D interactive illustrations. In a study that strongly affected our design, Sonnet et al. compared methods of associating text with its 3D model [32]; they evaluated the effects of text positioning, connectivity, and visual hints on comprehension under three conditions: (a) annotations are attached to objects using translucent shapes; (b) annotations are located within the objects’ shadows; (c) area showing the 3D model and text area are separated. The authors suggest that *setting a* works well for short labels, while for extensive texts, *setting c* seems to be applicable because a user can explore a scene without any occlusions from the text.

2.5 Usability Principles

Usability was another major driving factor for our design. According to [19, 25] the main principles of Web usability are: websites should explain themselves; people do not read pages - they scan them; do not waste people’s time; people are creatures of habit - use existing Web conventions; people tend to get “lost in space” - make it easy to go home, choose typography that communicates, and allow going back.

Currently, game industry leads the development of 3D interactive graphics and it is where many cutting edge interface ideas arise. Based on observation of interfaces of popular 3D games, works on game design [30] and design guidelines for virtual environments [18, 31], we summarized the main 3D design principles: *Text*: keep it readable and let users select for details on demand; *Navigation*: minimize the number of navigation steps, simplify movement (keep movements planar, use collision detection), allow teleportation; *Wayfinding*: provide overviews (maps) and history keeping.

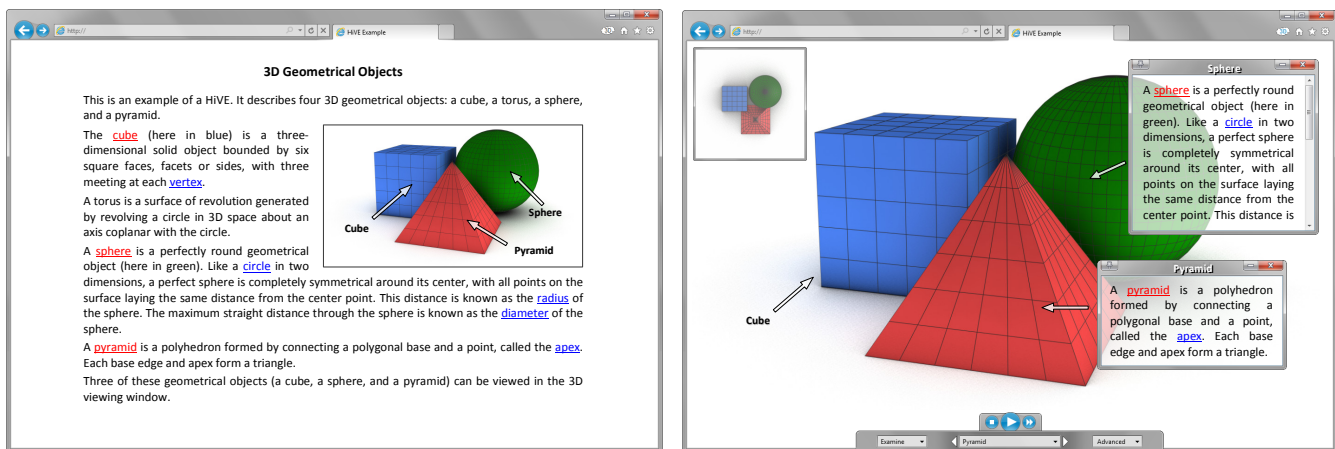


Figure 2: The same information presented in the hypertext and the 3D mode of the dual-mode user interface.

3. PROPOSED USER INTERFACE DESIGN

In the previous section we described the main driving factors in designing the interface for accessing integrated information spaces, where hypertext and 3D graphics data are simultaneously available and linked. Designing such user interface clearly presents some challenges as some factors contradict others. For this purpose we have developed a dual-mode user interface (DMUI) that has two modes between which the user can switch anytime (see Figure 2):

- (1) **Hypertext "don't-make-me-think" Mode** driven by simple hypertext-based interactions, where a 3D scene is embedded in hypertext;
- (2) **3D "take-me-to-the-Wonderland" Mode**, which immerses the hypertextual annotations into the 3D scene.

In the following we will describe in detail the two modes of the dual-mode user interface.

3.1 Hypertext "Don't-make-me-think" Mode

3D Web was (and still is) believed to have potential to be the next step in the Web's evolution, since it could benefit from graphics hardware and provide users with new and exciting experiences. Nevertheless, while various approaches have been proposed (most notably VRML/X3D), they have never seen much serious widespread use. One reason for the limited acceptance is the lack of 3D interaction techniques that are optimal for the hypertext-based Web interface.

Our approach to this problem is the hypertext mode of our interface (see Figure 2 (left)). This mode introduces a level of 3D-based interactivity and couples it with well adapted hypertext-based interactions. Our intention was to create, based on the Nielsen's [25] and Krug's [19] work on Web usability, a "don't-make-me-think" type of user interface. In the following we will describe the components of the hypertext mode of DMUI: hypertextual information and the embedded viewing window, where the 3D content appears.

3.1.1 Hypertextual Information

We define hypertextual information as an information set that can contain: textual information, non-textual information (e.g., static and animated images, audio, video), interactive information (e.g., flash interactive illustrations), nav-

igational means (e.g., hyperlinks). In the hypertext mode, hypertextual information is the primary information carrier. It is possible to read it without any interaction with a 3D scene - the information is not embedded into the 3D scene, but rather presented in a concise form familiar to the Internet users. Compared with standard hypertextual information that can be found e.g., on the Web, the hypertext mode of DMUI introduces two new user interface components/mechanisms: *3D-hyperlinks* and *hypersections*.

In our user interface, hyperlinks constitute not only a mechanism for navigation between hypertext documents, but also for navigation within 3D scenes. If a 3D scene contains a viewpoint node named `viewpoint1`, selecting a hyperlink connected to this viewpoint should smoothly animate the camera from its current position to the selected vantage point. By introducing 3D links, we aim to provide users with the ability to view 3D content from different points of view with a single mouse click. Visual cues are given as to where source anchors are located in a document. We use light blue highlighting as the default color for "hypertext links" and light red highlighting as the default color for "3D links" (see Figure 2 (left)). Both types of links can be embedded in hypersections.

Hypersections define sections of hypertextual information; they are analogous to HTML divisions that are often used to group block-elements to format them with styles. The difference is that hypersections are also designed to be:

- (1) *Annotations of related 3D objects*: Hypersections become annotations when the user switches to the 3D mode;
- (2) *Links between hypertextual and 3D information*: When a mouse cursor passes over a hypersection, the hypersection and the corresponding object in the 3D viewing window are automatically highlighted and the cursor changes its shape; the user is given visual cues as to what information is related to what object and where the related object is on the scene;
- (3) *Navigational UI components*: Pressing the middle mouse button over the hypersection animates the camera from its current position to the corresponding object.

We believe that 3D-hyperlinks and hypersections can greatly facilitate the interaction. We wanted to make possible for users with little knowledge of 3D interaction techniques to browse a 3D scene simply by making a single mouse click.

3.1.2 3D Viewing Window

The hypertext mode builds upon well established principles for including graphical resources in hypertext. For example, in HTML we use the `` tag to link to the image; the browser then displays the image where the tag occurs with the text flowing around the image (CSS is often used to define the appearance and layout of the page). Similarly, in the hypertext mode, the 3D viewing window is displayed by the browser "inline" with hypertextual information. This window renders 3D scenes through which the users can navigate and in which they can interact with objects.

Hypertext browsers often draw hyperlinks in a different color so that users can easily find them. Our interface provides three mechanisms to make certain that users could find selectable objects (objects in the scene linked to related hypertextual information): (1) they can be identified using labels, short textual descriptions connected to their referent objects with lines extending into the virtual scene (these labels are also links to more comprehensive explanations displayed in the 3D mode); (2) cyclic temporary highlighting of all selectable objects allows users to view all objects of a possible interest in the visible area of the 3D scene; (3) when a mouse cursor passes over a selectable object, the object, its label and the related hypersection are automatically highlighted.

While designing the hypertext mode of the dual-mode UI we have tried to accommodate as broad audience as possible by offering multiple ways to control the viewpoint, either by clicking selectable objects (easy for everybody), dragging the mouse across the 3D viewing window (more immersive, but also requiring some practice), or scrolling that gives people the option to see the whole scene in a guided tour.

We reserved a single left mouse button click in the 3D viewing window while the cursor is over a selectable object for *targeted movement navigation* [20, 33]. The click on an object of interest smoothly animates the camera from its current position to the selected object and optionally triggers a predefined for the given object camera movement (e.g. orbiting). Such approach was used to preserve the primary interpretation of clicking in the web browser window as following a hyperlink. The technique is easy to use, fast, and cognitive-friendly; it can also be easily integrated with other techniques [20]. On the other hand it has a major drawback: the target is always a selectable object.

The second possible interaction in the 3D viewing window is mouse dragging (moving the mouse while holding the left or right mouse button down) and is reserved for *general movement navigation*. This approach should, if possible, emulate real world behaviors and take into account information about the scene and the task at hand. For example, geographic VEs often employ a walking metaphor of camera motion where user positions are restricted to the 2D plane of the terrain; examine metaphor is often used to view different sides of objects and it is suitable for tasks where the user's goal is to view an object as though he or she were holding it. If the user's goal can be reliably determined, the moding between the navigation techniques should be automated.

There are some problems inherent in using general movement techniques. As they are designed to allow for unconstrained movement to any part of the VE, the user may move to unknown locations, look at things from awkward angles or miss seeing important features [11]. As a result, one cannot ensure that the user receives the intended message. Like

e.g., Galyean [12], we believe that empowering the author to bring some structure to the interaction experience can make VEs more suitable for the *new-to-3D* users. Therefore, our design balances the exploration methods with an ability to guide the user, while at the same time maintaining a sense of pacing or flow through the experience. We reserved scrolling for *specified trajectory movement navigation*. As users can navigate on a page by scrolling it, when the cursor hovers over a 3D scene, the mouse scroll wheel can also be used to navigate between the viewpoints defined for this scene.

If the user wants to have more freedom in terms of viewpoint control, he or she can switch to the 3D mode using a button located on a browser's tool bar (in the test application, this button is in the UI's bottom left corner). To avoid confusion, the state of the 3D environment (user's position, animations) is preserved when switching between UI modes.

3.2 3D "Take-me-to-the-Wonderland" Mode

Having a 3D graphics underlay invites interaction and having rich and immersive experiences. Yet, for sake of simplicity, the hypertext mode limits interaction with that layer. This can lead to a problem with supporting the feeling of immersion. What is immersion and why do we need it?

Immersion is often explained as *"the state of being absorbed or deeply involved"*. It is critical to Virtual Reality and can best be attained by visually immersing a user with HMD or CAVE, by using stereo displays and head tracking. However, immersion is also possible in desktop VEs, using desktop displays and common hardware for interaction (mouse and keyboard); as the user directly controls the interaction and focuses on it, he or she can be drawn into a 3D world [29].

The successful sensual immersion of the user in an imaginary 3D space is a very important part of the experience while interacting in a virtual environment. Achieving a close to real-life experience in a virtual world, creating a feeling of being there is crucial to give a potential virtual visitor the sensation of what the site is really like. Tan et al. [33] assert that the level of immersion that the user experiences greatly affect the navigation task and performance. The more immersed the user is, and the more easily the user can mentally integrate information acquired, the greater the chances of efficient navigation [29]. Certainly, much of a presence has to do with a quality of the presented material and the manner in which the user experiences it. Immersion can take place while we are playing a well designed video game, watching a quality movie, or even while reading of good novel, in spite of the lack of visual or perceptual immersion.

Comparing to the hypertext mode, the 3D mode of the dual-mode UI was designed to make users feel more present in an environment - more immersed. In this mode 3D graphics is the main information carrier. It provides users with much more freedom with regard to the 3D task support - it was designed to support unconstrained interactive navigation through a 3D scene. Furthermore, in this mode hypertextual data relating to an environment is embedded into that environment. The design of this, what we call "take-me-to-the-Wonderland" mode, was inspired by the work on IRVEs [5] and the work on annotating 3D illustrations [32]. In the following we will describe the components of the 3D mode of our interface: the viewing window, where the 3D content appears, hypertextual annotations, a dashboard designed to manage navigation, and a mini-map.

3.2.1 3D Viewing Window

In the 3D mode of the dual-mode UI, 3D scenes appear in the viewing window that spans the entire available screen space. A dashboard, a map and hypertextual information about an environment are rendered on top of the viewing window in *screen-space*, a 2D plane called HUD (head-up display) that overlays the 3D scene.

With regard to functionality, the viewing window in the 3D mode is similar to the one in the hypertext mode. By default, it offers the same navigation metaphors, the same behavior of selectable objects, etc. However, this mode provides a user with more than one "default" way of interacting with the 3D scene - the user can choose to use other methods of navigation (e.g., change from *walk* to *fly* or *examine*).

3.2.2 Hypertextual Information

One of the underlying premises of this research is that communicating information about 3D environments can be significantly improved by attaching annotations to the environment's objects. The 3D mode allows users to interactively recall and view the attached hypertextual information by clicking on labels connected to objects of interest during navigation. In response, the hypertextual information is presented in *screen-space* adjacent to associated objects in scrollable annotation boxes [13]. In fact, these hypertextual explanations are hypersections from the hypertext mode of our UI. Consequently, the annotation boxes may contain information from hypertext, through images and movies, to multimedia animations. They may also contain 3D hyperlinks for navigation within the 3D scene. Users may move and resize the annotation boxes, and toggle their visibility; they can also minimize them into labels. To better support human attention, better maintain the fluency of work, and to improve workspace visibility, annotations are rendered as semi-transparent user interface objects [17].

3.2.3 Dashboard

As we have already mentioned, for a given scene type and a task at hand, a designer should decide on the most intuitive mapping between input and interaction technique. However, very often there is insufficient input DOF for the task and user input must be moded. Therefore, the user has to be given explicit control of the different modes of navigation. A graphical dashboard (in Figure 2 presented at the bottom of the viewing window) provides ready access to the most important 3D interaction tools. By default, it should provide the methods of navigation based on examine, walk, fly and fly-to; it should also allow the user to switch between the viewpoints that are defined for the 3D scene.

3.2.4 Mini-Map

In addition to the difficulties of controlling the viewpoint, there is a problem of wayfinding, especially in large virtual worlds. This problem may manifest itself in a number of ways [9]: users may wander without direction when attempting to find a place for the first time, they may then have difficulty relocating recently visited places, they are also often unable to grasp the overall structure of the space ("*lost-in-cyberspace*" problem).

Maps proved to be an invaluable tool for acquiring and maintaining orientation and position in a real environment and according to [9], this is also the case in a virtual environment.

Influenced also by computer games, we decided to include a mini-map to the 3D mode of our interface (see Figure 2). It displays terrain, important locations and objects. It dynamically updates the current position of the user with respect to the surrounding environment.

4. COPERNICUS

In the previous section we described in detail the dual-mode user interface design for information spaces combining hypertext with interactive 3D graphics. To put our design into practice and evaluate it, we decided to build a testbed, a platform for the experimentation and for the assessment of both hypertext and 3D modes of our interface, a system that would allow to find a balance between 3D interaction techniques and well established hypertext interactions. We developed a wiki-type authoring environment called Copernicus. Its design was inspired by the popular MediaWiki (used to power e.g., Wikipedia); in addition to a classic hypertextual content, any page with an article can contain a 3D visualization of the place/object described in this article. In Copernicus, different types of information, from text, through images and video clips, to 3D graphics, can be easily collected, linked, and later made available as integrated information spaces in the hypertext based environment (the hypertext mode) or within the context of a virtual environment (the 3D mode). It is important to note that Copernicus was used to create the virtual museum of Warcraft for the user study described in this article.

Copernicus was implemented using .NET as XAML Browser Application (XBAP), so it can be deployed on the Web. Users with .NET framework can access Copernicus just like a Flash-enhanced web page using IE or Firefox on Windows. The project's source code is available under GPL license.

We had several opportunities to observe novices interacting with the dual-mode user interface using Copernicus. Most observations were made in primary and secondary schools (one of the objectives of DERI is to popularize the knowledge of science and promote engineering among young students) as well as at local science fairs. We also observed individual users at their personal work spaces (mostly at DERI and NUIG campus). These users were free to access and navigate any content they preferred; they then provide us with feedback on the positives and negatives of the system.



Figure 3: The youngest user of Copernicus.

The comments from the participants of this initial evaluation were extremely positive. The study has shown that due to only a slight modification of hypertext-based interface paradigm, the users had no problems interacting with Copernicus. The simplicity of the interaction techniques, essentially a single click in the hypertext mode and a drag action in the 3D mode, were immediately understood and usable by all our users.

5. EVALUATION

To characterize the presented user interface in terms of its efficiency and usability, we conducted a user study, where we compared our dual-mode interface to two other, currently most common user interfaces for integrated information spaces, where text and 3D graphics data are simultaneously available and linked. In the following we will describe our study in detail. First we discuss the evaluation setup used in the experiment. This discussion is followed by the description of the evaluated user interfaces and the procedure used in this evaluation. Finally we present and discuss the results of the experiment.

5.1 Participants

20 students, researchers and members of staff with normal or corrected-to-normal vision participated in the experiment. 5 of the participants were native English speakers. 5 of our participants were female. The participants ranged in age from 27 to 42, with 12 participants in the 27-30 range and 8 participants in 30-42 range. 8 participants were PhD students, 9 had higher education, and 3 were post-doctoral researchers. All of the participants were familiar with hypertext navigation; 3 of them had no or very little experience with 3D interaction, 13 had some experience navigating 3D scenes (playing 3D games sporadically), and 4 had considerable experience in 3D navigation from playing 3D computer games. Subjects were given gifts worth 15€/20\$ for their participation. Additionally, an iPod was awarded to one of the 5 best performing participants.

5.2 Apparatus

The experiment was conducted on the Intel Core 2 Extreme laptop computer equipped with 4GB of memory, GeForce 9800M GTX graphics card, connected to a 24-inch widescreen LCD display running at 1920x1200 resolution. Input devices were a standard 3-button optical mouse and a keyboard. The computer operating system used was Microsoft's Windows 7. The test application used for the evaluation was developed based on the Copernicus source code; the content for evaluation (see next section) was authored using our 3D wiki as well (visit <http://copernicus.deri.ie> to learn more).

5.3 Stimuli

For the purpose of this experiment we prepared a virtual museum featuring heroes, races, creatures, and weapons from the fantasy setting of the World of Warcraft (WoW) game; this choice was made to prevent the influence of previously gathered knowledge. The museum was divided into four exhibitions, one for the training session (heroes) and three for the main experiment (races, creatures and weapons). Each exhibition conveyed integrated hypertextual and 3D visual information. The virtual worlds created for the study were simple single-floor museum-like environments, populated with 3D objects and images/painting (see Figure 4). Below follows a description of each exhibition:

Heroes of Warcraft - an exhibition of four notable characters from the Warcraft game. The article consisted of about 480 words and 4 images. The 3D scene consisted of three rooms populated by 5 objects and 5 paintings.

Races of Warcraft - an exhibition of the twelve playable races of the Alliance (i.e. Humans, Night Elves, Dwarves, Gnomes, Draenei, and Worgen) and the Horde (Orcs, Trolls,

Tauren, Forsaken, Blood Elves, and Goblins) factions from WoW. The article consisted of about 1350 words and 12 images representing each race. The 3D scene consisted of four rooms populated by 12 race objects and 12 paintings.

Creatures of Warcraft - an exhibition of common creatures that can be found in the World of Warcraft, such as bears, saber cats, drakes, and wolves. The article consisted of about 920 words and 3 images; each creature was characterized by strength, agility, and intellect values. The 3D scene consisted of nine rooms populated by 14 objects and 13 paintings.

Weapons of Warcraft - an exhibition of weapons (such as swords and axes) from the Warcraft universe. The article consisted of about 1060 words; each of the 9 weapons was characterized by damage, bonuses (e.g. to strength, agility, etc.), and a price. The 3D scene consisted of one room; all 9 objects were positioned in the center of the room.



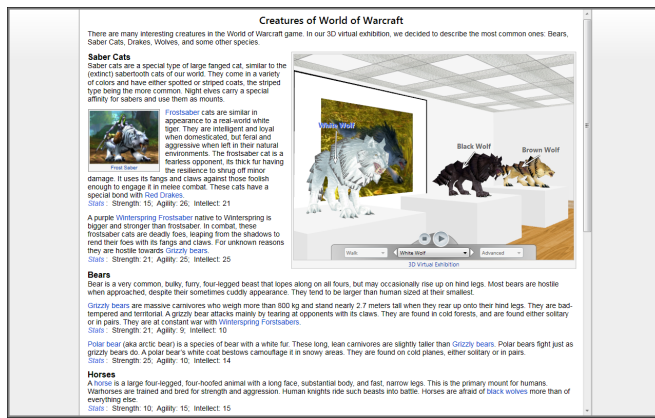
Figure 4: The 3D scenes used in the evaluation.

According to the classification of virtual worlds [8], all our environments are dense (relatively large number of objects and cues in the space) and static (the positions and values of the objects do not change over time). Moreover, the exhibition of weapons is a small world (a world in which all or most of the world can be seen from a single viewpoint), while all other environments are large (there is no vantage point from which the entire world can be seen in detail).

5.4 User Interfaces

As we have already mentioned, the hypertext "don't-make-me-think" mode of the dual-mode user interface was inspired by the state-of-the-art practice of embedding 3D scenes as part of an HTML page. The design of what we call 3D "take-me-to-the-Wonderland" mode was inspired by the work on IRVEs [5] and the work on annotating 3D illustrations [32]. To characterize the dual-mode user interface in terms of its efficiency and usability, we decided to compare it to these two inspirations that are currently user interfaces of choice for integrated information spaces, where text and 3D graphics data were simultaneously available and linked.

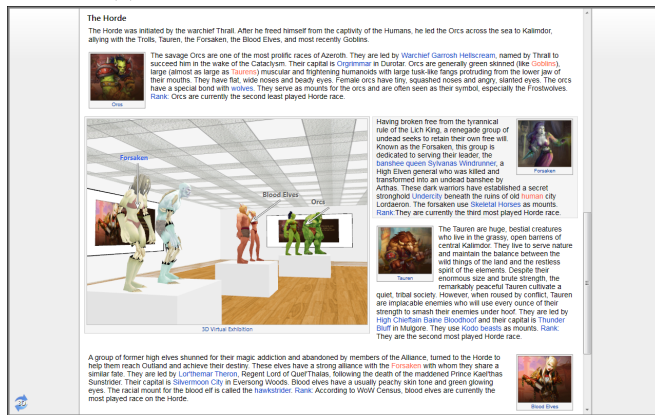
Hypertext UI - this interface was created by modifying the hypertext mode of the dual-mode UI. Features like an ability to switch to 3D mode, hypersections, and 3D hyperlinks were disabled. On the other hand, the dashboard UI component was added to the 3D viewing window (see Figure 5a).



(a) Creatures of Warcraft in the hypertext UI



(b) Weapons of Warcraft in the 3D UI



(c) Races of Warcraft in the hypertext mode of the dual-mode UI



(d) Races of Warcraft in the 3D mode of the dual-mode UI

Figure 5: The exhibitions presented in the evaluated user interfaces.

3D UI - this interface was created by modifying the 3D mode of the dual-mode user interface. Features like an ability to switch to the hypertext mode and 3D hyperlinks were disabled (see Figure 5b).

Dual-Mode UI - this interface integrates Hypertext UI and 3D UI into one modal interface. It allows users, while in the hypertext mode, to read about the collections and easily navigate through the rooms of the virtual museum using 3D hyperlinks and hypersections (see Figure 5c). The same UI, while in the 3D mode, also allows users to experience the 3D scenes that span the entire available screen space of the browser window; just like in the 3D UI, the user can walk through the rooms of the museum and click on object's labels to read more comprehensive explanations (see Figure 5d).

The user interface designs evaluated in this study differed in the method used to integrate the textual information with the objects in the 3D scene. On the other hand, the interfaces allowed for the same interactive exploration of 3D scenes; there were no differences in the techniques that enabled a user to navigate the 3D scenes. Movement was confined to 'walk' mode; guided tour navigation (scrolling over 3D scene) was disabled; collision detection was used to prevent users moving through objects and walls.

5.5 Tasks

Different possible measures could be used to determine the effectiveness and usability of the evaluated interfaces. In choosing tasks for the study, we looked for ones that are

both valid (resemble a 'real' act of browsing 3D content on the Web) and that are recognized for being able to detect significant differences. We decided to adopt tasks that were introduced by Chen et al. [7] and were later successfully used by Polys et al. [27, 28] to evaluate IRVEs. Thus, the participants performed 4 types of tasks, representing various conditions a user is likely to experience on a 3D Web site:

- (1) Search for textual information and then search for visual information (S:H-3D). Task 1 requires the users to first search for text information, and then to find the corresponding visual information in the 3D scene. An example task is: *Find the Horde race that uses Hawkstriders for mounts. What other races are to the left and right of this race?*
- (2) Search for visual information followed by textual information (S:3D-H). Task 2 is conceptually reversed, in that the users are required to find the visual information on the 3D scene first, and then to answer questions about the related text information. An example task is: *Find the sword which hilt/handle has a yellow dragon eye and ends with dragon claws. What is the price of this weapon?*
- (3) Compare text information and derive visual information (C:H-3D) (find visual attributes of items with a given text criteria). An example task is: *Find the strongest creature in the museum. What is the color of the creature's eyes?*
- (4) Compare visual information and derive textual information (C:3D-H) (search for textual attributes of items with a given visual criteria). An example task is: *There are two races with tails. What are their mounts?*

5.6 Measurements

Like in [7, 27], the study measured relative effectiveness of our user interfaces by both examining time taken to answer each question and correctness of answers. In addition, we developed a questionnaire to measure participants' subjective impressions of the user interfaces. The questionnaire contained continuous Likert scales regarding ease of use, learnability, efficiency, aesthetics, presentation and access to text, 3D navigation, wayfinding, immersion, and overall preference. Subjects were also welcome to add any comments relevant to their experiences. The test application also recorded the usage of different UI components (e.g., a number of selections in a viewpoint menu, the use of hypersections, 3D hyperlinks, etc.)

5.7 Procedure

Each test session started with an introduction to the test application. Its interface was divided into two parts: the window, where the user was presented with tasks and the browser window (1280x800), where the user could interact with the prepared exhibitions through the user interfaces evaluated in this study. The introduction was followed by a training session (4 practice tasks for each interface) to allow the subject to get familiarized with the test application, the interfaces, and the test procedure. The users were educated and guided on how to use the walk and go-to navigation metaphors, and the viewpoint menu for control in a virtual world; they were also introduced to the concepts of hypersection and 3D hyperlink. After the subjects indicated that they were satisfied, we proceeded with the actual trials.

The tasks in the main part of the evaluation were similar to the ones from the training session: for each exhibition-UI combination we asked 4 questions related to the content of the exhibitions (presentation of variables was counterbalanced by means of Latin square design). For each question there was a choice of 4 answers from which the user had to choose 1 and only 1 answer. The subjects were asked to complete the assigned tasks "as accurately and as fast as possible". They were also told that it was more important to solve the tasks correctly rather than to be quick. They were allowed to take a break between each set of questions. The participants were video recorded during the tasks and notes were taken about their actions and comments. After being presented with all 36 tasks (3 UI modes * 3 exhibitions * 4 tasks), the users were given the questionnaire and asked to directly compare the evaluated user interfaces. Each evaluation session lasted approximately 120 minutes - here it is important to stress the fact that for most of the participants the experiment was not tiring and seemed much shorter (actually, some participants expected more questions).

5.8 Results

We collected a total of 720 time and accuracy measurements (20 subjects * 3 UI modes * 3 exhibitions * 4 tasks), and 660 measurements of subjective impressions (20 subjects * 11 questionnaire parts * 3 UI Modes). We analyzed our results with analysis of variance (ANOVA). With ANOVA we modeled our experiment as a repeated-measures 3x3x4 design (UI x Environment x Task). Bonferroni procedure was used for evaluating the pairwise comparisons.

* Visit <http://copernicus.derri.ie/www2012.htm> to view experimental results and recordings from the test sessions.

5.8.1 Objective Results

Times for completion of each task were normalized on the overall average completion time. Normalization was used to remove any effects of base reading speed and 3D navigation experience among participants. As overall accuracy was very high (0.985%), we decided to simply double the times of wrong answers.

Analysis of the task completion time revealed significant main effects of all variables and their interactions ($p < 0.003$). Most importantly, it found significant main effects of UI ($F(2, 38) = 44.32, p = .00000$), interaction between UI and environment type ($F(4, 76) = 4.49, p = .0026$), and interaction between UI and task type ($F(6, 114) = 25.66, p = 0.0000$).

Post-hoc comparisons of means revealed that the dual-mode UI condition resulted in the best overall task performance ($p < 0.0001$), while the hypertext UI condition was marginally worse than the 3D UI ($p < 0.041$). To be more precise, executing tasks using the dual-mode UI was about 43% faster than using the hypertext UI (99s vs. 141s), and about 31% better than using the 3D UI (99s vs. 129s), while executing tasks using the 3D UI was about 9% faster than using the hypertext UI (129s vs. 141s).

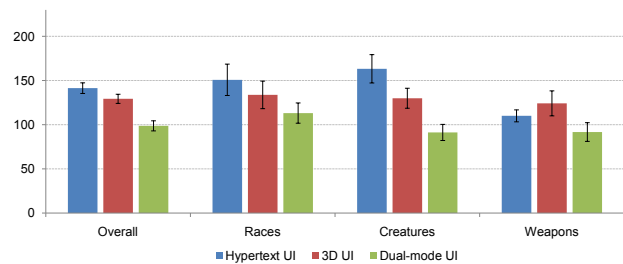


Figure 6: Overall results.

Comparisons of means for each exhibition revealed that the dual-mode UI was significantly better than the hypertext UI ($p < 0.015$) for Races and Creatures. It was also better than the 3D UI ($p < 0.011$) for Creatures. Figure 6 illustrates the overall results of our experiment and the results for each exhibition with respect to task completion time (error bars denote 0.95 confidence intervals). An interesting finding, visible in Figure 6, is that the hypertext UI was worse than the 3D UI in large environments, and it was better in a small one (Weapons). We believe this is because the small environment did not require much 3D navigation and users could not get lost in 3D space.

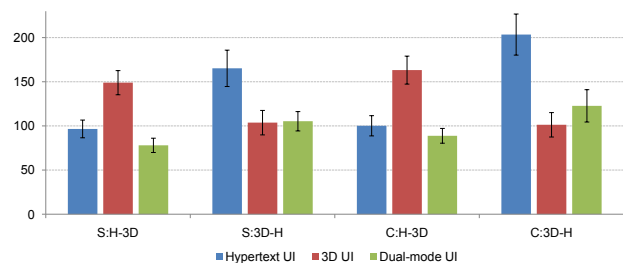


Figure 7: Interaction between UI and task type.

As we have already mentioned, we also found a significant main effect of interaction between UI and task type ($F(6, 114) = 25.663, p = 0.0000$) on task completion time. Not surprisingly, since the task types differed significantly (see Sec-

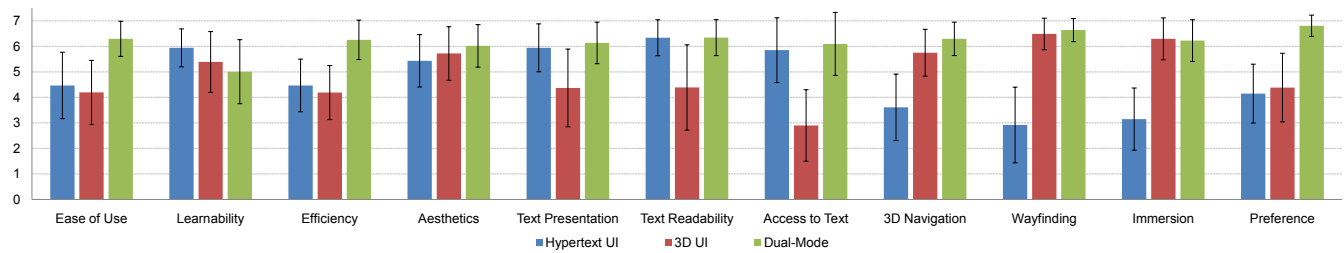


Figure 8: Subjective results from the questionnaire.

tion 5.5). Post-hoc comparisons of means revealed that executing tasks using the dual-mode and hypertext UIs was significantly faster than using the 3D UI for S:H-3D and C:H-3D tasks ($p < 0.0001$); using the dual-mode and 3D UIs was significantly faster than using the hypertext UI for S:3D-H and C:3D-H tasks ($p < 0.00006$) - see Figure 7.

5.8.2 Subjective Results

The average subject ratings with respect to ease of use (difficult/easy), learnability (difficult/easy), efficiency (completing the tasks fast and accurately was: difficult/easy), aesthetics (non-aesthetic/aesthetic), presentation of text (confusing/clear), readability of text (hard/easy to read), search and access to text (difficult/easy), 3D navigation (complicated/simple), wayfinding (complicated/simple), immersion (not involved/really immersed), and overall preference are illustrated in Figure 8, together with standard deviations.

Analysis of the ratings revealed significant main effects of UIs on all scores. The dual-mode UI was perceived easier to use and more efficient than the hypertext and 3D UIs ($p < 0.0001$). It was also perceived as more aesthetic than the hypertext UI ($p < 0.003$). On the other hand, the hypertext UI was perceived as easier to learn than the 3D UI ($p < 0.045$) and the dual-mode UI ($p < 0.0003$). Subjects perceived presentation, readability and access to text in the 3D UI as worse than in the dual-mode and hypertext UIs ($p < 0.0001$). In contrast, 3D navigation, wayfinding, and immersion in the hypertext UI were ranked lower than in the dual-mode and 3D UIs ($p < 0.0001$). Finally, the dual-mode UI was evidently preferred over the alternatives ($p < 0.0001$). These findings clearly support our analysis of task performance.

5.8.3 Discussion

The results from this competitive user study suggest users performed better with the dual-mode user interface over alternatives, i.e. the hypertext and 3D UIs, on tasks, which we believe are representative of a variety of 3D Web applications. The performance with the dual-mode UI was better because, except for switching costs (visible in C:3D-H), each mode of the interface could be used optimally for each task. Hypertext mode was employed more often for both H-3D types of tasks, while 3D mode was a choice for 3D-H tasks. The subjective comments also showed a preference for the dual-mode interface. Moreover, the evaluation results can help to understand better the relationship between the hypertext and 3D UIs. Like most controlled user-based studies, this one had some limitations that restrict the generality of our findings: although we tested three different virtual environments, we still managed to test only a small sample of possible uses of 3D content.

A viewpoint menu turned out to be very important navigation tool. Interestingly, some users preferred alphabetic

order of viewpoints, some preferred order based on the distance between the viewpoints; one subject noted that he would like to have categories in the menu. On the other hand, it is not clear whether it would be useful for environments with a large number of points of interest. A map also proved to be very important wayfinding aid to the majority of users. We noticed that few users clicked on the map - they expected instant teleportation to the selected rooms.

3 of the 4 participants with prior considerable experience in 3D navigation stated that they at times would have liked keyboard-based control of the camera, in addition to the mouse-only control we provided. Some participants asked about search support (Ctrl-F), both for text and 3D. One user did not like the grey highlighting of 3D objects in the 3D scene: *"such highlighting makes colors and details less visible"*. A few users criticized the label and annotation layout used in the experiment as we implemented a very simple layout scheme that places the label/annotation box on the left corner of a box that encloses the geometry's bounds (bounding box). With regard to the dual-mode UI, one suggestion was to move the button for switching modes and position it in the corner of the 3D viewer (in the hypertext mode).

6. CONCLUSIONS AND FUTURE WORK

During the First Hypertext Conference in 1987, Andries van Dam gave a keynote speech and listed user-controlled 3D graphics as one of the key issues we should be looking at while researching and developing hypertext systems:

"If a picture is worth a thousand words, a dynamic picture of time-varying objects is worth a thousand static ones. We need dynamics at the nodes, not just static pictures and text."

Andries van Dam [34]

24 years have passed and Andy's vision is still just a vision. The Web, today's largest and most important hypertext-based online information infrastructure, does not support 3D content and, although various approaches have been proposed (most notably VRML/X3D and now WebGL), there is still no clear design methodology for user interfaces that integrate hypertext and interactive 3D graphics.

We have presented a novel strategy for accessing 3D content on the Web. We have introduced a user interface that has two modes between which a user can switch anytime: the driven by simple hypertext-based interactions hypertext *"don't-make-me-think"* mode, where a 3D scene is embedded in hypertext and the more immersive 3D *"take-me-to-the-Wonderland"* mode, which immerses the hypertextual annotations into the 3D scene. Results from the competitive user study suggest users performed better with dual-mode user interface over alternatives.

There are a number of directions in which we are interested in taking this work. Firstly, our UI design is not yet grounded in any formal standard, as we have focused on the HCI perspective of the problem. As the participants of the Declarative 3D for the Web Architecture W3C CG, we want to explore the possibilities of applying our design methodology in the HTML5 specification, since it does not yet define how the integration of 3D imagery into HTML actually should look like. Another future avenue that we intend to explore is collaborative web browsing. Currently, browsing the Web is mostly an individual experience. People visiting the same web page at the same time are not aware of each other. DMUI supports this personal nature of browsing. On the other hand, research on Collaborative Virtual Environments [2] has shown that their entertainment and social nature can provide virtual communities with richer content and greater interactivity and greatly support collaborative work. We envision the Web (or rather part of it) evolving into a virtual space in which people, while being in the 3D mode of the dual-mode UI, can interact and communicate with each other through their avatars.

* This work has been supported by SFI under Grant No. SFI/08/CE/I1380 and by EI under Grant No. PC/2008/353. Visit <http://copernicus.deri.ie> to learn more about DMUI.

7. REFERENCES

- [1] K. Andrews. Soaring through hyperspace: A snapshot of hyper-g and its harmony client. In *Eurographics Symposium on Multimedia/Hypermedia*, 1994.
- [2] S. Benford, C. Greenhalgh, T. Rodden, and J. Pycock. Collaborative virtual environments. *CACM*, 2001.
- [3] J. D. Bolter, L. F. Hodges, T. Meyer, and A. Nichols. Integrating perceptual and symbolic information in vr. *IEEE Comput. Graph. Appl.*, 1995.
- [4] D. A. Bowman, L. F. Hodges, and J. Bolter. The virtual venue: User-computer interaction in information-rich virtual environments. *Presence*, 1998.
- [5] D. A. Bowman, C. North, J. Chen, N. F. Polys, P. S. Pyla, and U. Yilmaz. Information-rich virtual environments: theory, tools, and research agenda. In *VRST'03*. ACM, 2003.
- [6] P. Chandler and J. Sweller. Cognitive load theory and the format of instruction. *Cognit. and Instruct.*, 1991.
- [7] J. Chen, P. S. Pyla, and D. A. Bowman. Testbed evaluation of navigation and text display techniques in an information-rich virtual environment. In *Virtual Reality'04*. IEEE, 2004.
- [8] R. P. Darken and J. L. Sibert. A toolset for navigation in virtual environments. In *UIST'93*. ACM, 1993.
- [9] R. P. Darken and J. L. Sibert. Wayfinding strategies and behaviors in large virtual worlds. In *CHI'96*. ACM, 1996.
- [10] P. Faraday and A. Sutcliffe. Designing effective multimedia presentations. In *CHI'97*. ACM, 1997.
- [11] G. Fitzmaurice, J. Matejka, I. Mordatch, A. Khan, and G. Kurtenbach. Safe 3d navigation. In *I3D'08*. ACM, 2008.
- [12] T. A. Galyean. Guided navigation of virtual environments. In *SI3D'95*. ACM, 1995.
- [13] T. Gotzelmann, K. Hartmann, and T. Strothotte. Agent-based annotation of interactive 3d visualizations. In *Smart Graphics'06*, 2006.
- [14] T. Gotzelmann, P.-P. Vazquez, K. Hartmann, A. Nurnberger, and T. Strothotte. Correlating text and images: Concept and evaluation. In *Smart Graphics'07*, 2007.
- [15] K. Hartmann, T. Gotzelmann, K. Ali, and T. Strothotte. Metrics for functional and aesthetic label layouts. In *Smart Graphics'05*, 2005.
- [16] J. Jankowski. A taskonomy of 3d web use. In *Web3D'11*. ACM, 2011.
- [17] J. Jankowski, K. Samp, I. Irzynska, M. Jozwicz, and S. Decker. Integrating text with video and 3d graphics: The effects of text drawing styles on text readability. In *CHI'10*. ACM, 2010.
- [18] K. Kaur. *Designing Virtual Environments for Usability*. PhD thesis, City University London, 1998.
- [19] S. Krug. *Don't Make Me Think: A Common Sense Approach to the Web (2nd Edition)*. 2005.
- [20] J. D. Mackinlay, S. K. Card, and G. G. Robertson. Rapid controlled movement through a virtual 3d workspace. *SIGGRAPH*, 1990.
- [21] R. E. Mayer. *Multimedia Learning (2nd ed.)*. 2009.
- [22] R. E. Mayer and R. Moreno. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *J. Educ. Psychol.*, 1998.
- [23] M. Mohageg, R. Myers, C. Marrin, J. Kent, D. Mott, and P. Isaacs. A user interface for accessing 3d content on the world wide web. In *CHI'96*. ACM, 1996.
- [24] S. Y. Mousavi, R. Low, and J. Sweller. Reducing cognitive load by mixing auditory and visual presentation modes. *J. Educ. Psychol.*, 1995.
- [25] J. Nielsen. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, 2000.
- [26] A. Paivio. *Mental representations. A dual coding approach*. Oxford University Press, 1986.
- [27] N. F. Polys, D. A. Bowman, and C. North. The role of depth and gestalt cues in information-rich virtual environments. *IJHCS*, 2011.
- [28] N. F. Polys, S. Kim, and D. A. Bowman. Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments. In *VRST'05*. ACM, 2005.
- [29] G. Robertson, M. Czerwinski, and M. van Dantzich. Immersion in desktop virtual reality. In *UIST'97*. ACM, 1997.
- [30] R. Rouse. *Game Design - Theory and Practice*. 2000.
- [31] B. Shneiderman. Why not make interfaces better than 3d reality? *IEEE Comput. Graph. Appl.*, 2003.
- [32] H. Sonnet, M. S. T. Carpendale, and T. Strothotte. Integration of 3d data and text: The effects of text positioning, connectivity, and visual hints on comprehension. In *INTERACT'05*, 2005.
- [33] D. S. Tan, G. G. Robertson, and M. Czerwinski. Exploring 3d navigation: combining speed-coupled flying with orbiting. In *CHI'01*. ACM, 2001.
- [34] A. van Dam. Hypertext '87: keynote. *CACM*, 1988.
- [35] M. C. Wittrock. Generative processes of comprehension. *Educational Psychologist*, 1989.
- [36] N. Yankelovich, B. J. Haan, N. K. Meyrowitz, and S. M. Drucker. Intermedia: The concept and the construction of a seamless information environment. *Computer*, 1988.