Published in the Software Focus Journal, 2(2):52-59, John Wiley & Sons, 2001.

Life on the Web

IGOR S. PANDZIC
Department of Electrical Engineering
Linköping University, SE-581 83 Linköping
igor@isy.liu.se

Abstract

With an increasing number of companies and research centers producing Virtual Characters aimed specifically for the Internet, the Web may soon become very much alive with all sorts of interesting and useful creatures guiding us, giving information or simply trying to sell us something. This article tries to predict the possible applications for Virtual Characters in the context of the Web, analyzes the technical requirements to enable these applications and provides a comparative review of currently available technologies with respect to the outlined requirements.

Keywords: Virtual Characters, Virtual Humans, Facial Animation, MPEG-4, FBA, VRML, H-ANIM, Text-to-Speech

Introduction

The technology of modeling and animating Virtual Humans has been a fascinating research topic for more than two decades, shifting in interest over the years from off-line Computer Animation to Real Time applications and, most recently, to Web applications. The impact of this vast research base has been felt in many applications – most striking examples are 3D Animation, computer games, film and special effects, all of which use various Virtual Humans technologies extensively.

With the recent boom in the Internet services, new applications for Virtual Humans technologies can be identified. We believe in particular that talking Virtual Characters [1][2][3][4][5][6][7][37] can be useful when integrated in Web sites to provide novel services as virtual hosts, salespersons, newscasters and other. In this article we attempt to identify classes of applications and give examples of possible individual applications that we believe could emerge in this field. We base our analysis on research work as well as emerging commercial attempts that are currently multiplying on the Internet.

After presenting the applications, we analyze the technical requirements for bringing such applications to life. This leads to a set of criteria which we then use to analyze the emerging technologies for bringing Virtual Characters to the Web. To support this analysis, we first briefly discuss the underlying techniques to bring the 3D graphics support to the Web, which is usually a necessity for Virtual Characters simulation. We then present the state of the art in Web-based Virtual Characters by identifying five main technical approaches to the problem. We analyze the advantages of each approach and discuss concrete examples. We classify all found examples of Web-based Virtual Characters in a comparative table with a number of distinguishing criteria drawn from our discussion on requirements.

In the final section we summarize the given analysis and attempt to draw conclusions as to technologies and applications most likely to emerge in this exciting new area.

Published in the Software Focus Journal, 2(2):52-59, John Wiley & Sons, 2001.

Applications

The technology to simulate humans on the Web may seem narrowly specialized. However, one must remark that the contents of most produced audio-visual material (films, TV, cartoons etc.) are dominated by human characters. Speech, gestures and facial expressions are main communication channels for humans, and bringing them to the Web is a potentially powerful publishing tool that can be usefully deployed on many Web sites, from small personal home pages to huge commercial sites.

We believe that placing a human-like Virtual Character on a Web site can bring the following benefits:

- Give a personality to the Web site.
- Enable to talk to each person visiting the site; people like to be talked to.
- Make visitors remember main messages better.
- A talking person can be more persuasive than written text.

Additionally, as outlined in [8], the Virtual Character may in certain situations alleviate the (sometimes unavoidable) Internet waiting times by "entertaining" the user. The same study has also shown that the end users show a preference for a service enhanced with a Virtual Character to a plain text service.

If correctly implemented, Virtual Characters have low bandwidth requirements and high interaction capacity, making them a natural replacement for video streaming in "talking head" scenarios.

We believe that we are just beginning to recognize possible practical applications, and what is presented here may well be just the tip of the iceberg. We can broadly classify the emerging application areas in the following categories: entertainment, personal communications, navigation aid, broadcasting, commerce and education.

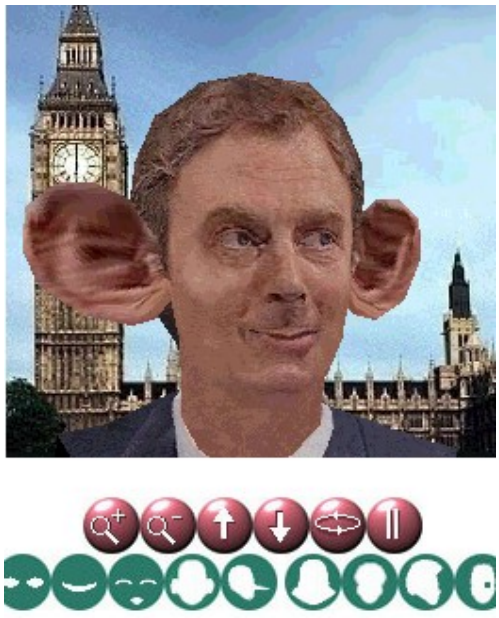

Figure 1: Animated caricature of a politician [9]; viewer can trigger funny actions (© BioVirtual Ltd.)

Published in the Software Focus Journal, 2(2):52-59, John Wiley & Sons, 2001.

Entertainment

Jokes and amusing contents seem to be among the first targets of many creative technologies and Web-based Virtual Characters are no exception. Animated caricatures of famous politicians (e.g. Figure 1) have already made their appearance, as well as other amusing creatures. Virtual Characters are suitable for delivering this kind of content because they generally do not require very high bandwidth (compared to video) and at the same time they can offer interactivity, i.e. instead of delivering a linear story, the user can trigger various reactions from the character.

Personal communications

Invitations, birthday wishes, jokes, all kinds of personal messages can be delivered either on personal Web pages or through email by appealing, talking Virtual Characters. Electronic greeting cards with 3D Virtual Characters have already been offered by at least one company [10].

Navigation aid

In the simplest form, a Virtual Character can be used simply to welcome visitors to a Web site by appearing on the home page and delivering a short welcome address to each visitor. More interestingly, a virtual guide can accompany the visitor navigating around numerous pages of a big Web site, always appearing in a corner of the screen and providing guidelines, tips, or even a full guided tour where the Virtual Guide shows various Web pages in particular order and comments them while the visitor sits back and watches the tour.

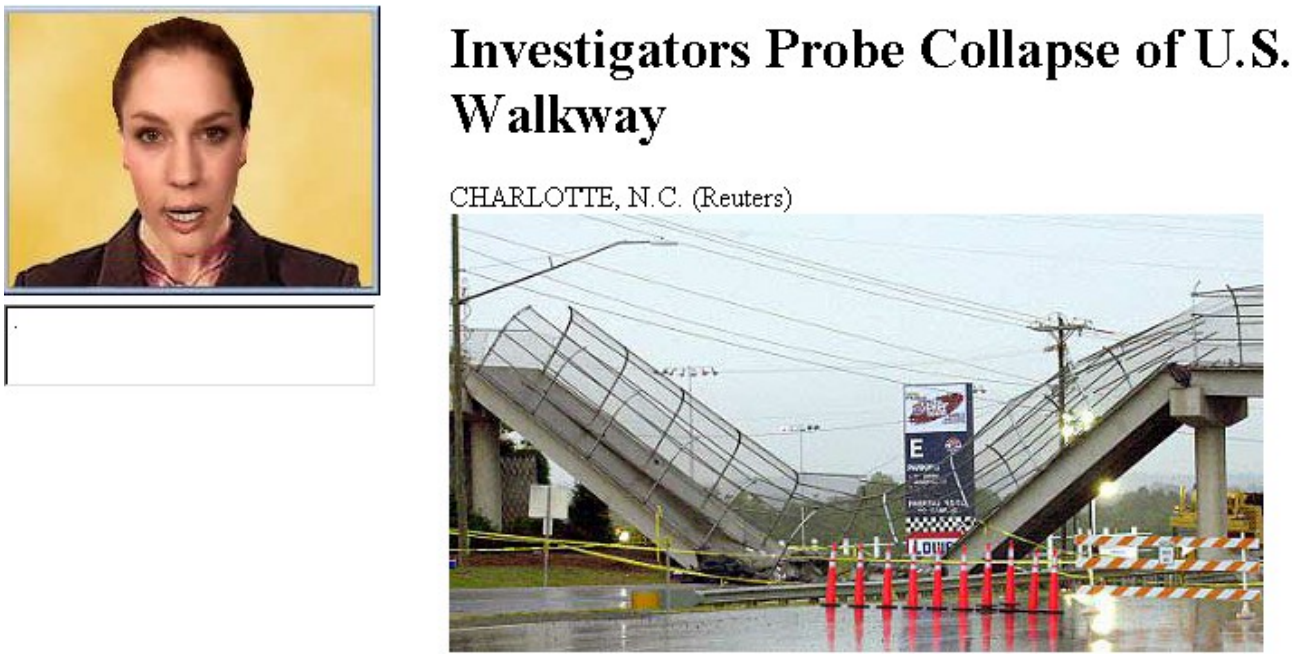

Figure 2: A virtual newscaster; images on the right appear as she talks (© W Interactive SARL)

Broadcasting

Virtual newscasters will deliver daily or hourly news updates. Unlike the TV newscaster, the one on the Web can be personalized and remember that a particular user is interested in sports, stock market and weather and likes to hear the related news in this particular order.

This may spawn an interesting new service integrating advantages of both TV and the Internet. The TV news are broadcasted to millions of people with no possibility of personalization. We have to

watch them at designated times, and can only broadly choose contents by switching between specialized channels. As a reward for the lack of choice we get rich, high quality, easily and quickly digestible contents.

On the Internet, we can easily subscribe to news services and personalize their appearance, allowing us to see the news we are interested in, in the order we like, and any time we choose. However, the news is mostly text with possibly a few images.

A Virtual Newscaster can combine the advantages of both approaches by presenting personalized news, at the users request, and in the audio-visual form by reading the news and simultaneously presenting relevant images (see Figure 2).

Figure 3: A text based conversational character (© eGain Communications Corp.)

Commerce

Some companies specializing in Customer Relationship Management (CRM) applications already offer “intelligent” conversational characters that a company can use as the front-line customer support on their web site. Currently most of these characters are purely text based. The customer types in questions in plain language and gets the answer in form of text, sometimes also displaying an appropriate web page. The characters are programmed with an extensive database of answers to all kinds of questions that customers may ask, as well as conversational rules enabling them to perform a semblance of natural conversation. While relatively simple, these characters can answer many “standard” questions in a satisfying way, in many cases avoiding the need for human intervention.

Currently it is usual for such characters to display images of a person (see Figure 3) as a means of giving it an identity. An animated, talking virtual character is the logical next step for such applications, giving them even more personality and making them more user-friendly. Since the characters’ answers are mostly pre-programmed in form of text, using text-to-speech and lip synchronization techniques to bring them to life is just a small step away.

Similar approach can be employed for sales. Instead of just showing an image of the product and main information, a virtual salesperson can simultaneously provide a “sales pitch” and answer any additional question the potential buyer may have. Alternatively, the customer can ask for a tour of the product, obtaining a rich audio-visual experience with the virtual sales representative explaining the qualities of the product while at the same time showing appealing images, as the example in Figure 4 shows.

Education

Even though we did not find existing examples of using Virtual Characters for educational purposes on the Web, we did find a certain interest in the education community for using this technology in online learning, e.g. staff training, online degrees, distance education.

Published in the Software Focus Journal, 2(2):52-59, John Wiley & Sons, 2001.

Figure 4: Virtual sales representative presenting a car model (© W Interactive SARL)

In the following sections we discuss the technologies needed to deploy all these possible applications by first analyzing the technical requirements defined by the application scope, then presenting and analyzing the currently available technologies.

Requirements

Based on the proposed applications, in this section we analyze the technical requirements on the implementations of Virtual Characters for the web, and their possible implications. We have identified the following requirements: visual quality, easy installation, fast access, interactivity and Web integration.

Visual quality

It is obvious that Virtual Characters featured in any of the mentioned applications must have a visual appeal. They must look nice, and animations should look natural. This does not necessarily mean photo-realistic models, or even very high-resolution models. A fairly simple cartoon character, if cleverly designed, can carry loads of appeal. The real implication of this requirement is that any system hoping for success must provide convenient means for visual artists to design models and animations, preferably using the tools that they are already used to.

Easy installation

In many of the mentioned applications the virtual character is not in itself the main attraction of the proposed service, but rather an extra bonus, an improvement. The users who never used such a novel application might be reticent to install software on their computers only to experience an unknown improvement of the service. This creates a barrier. To make the user jump over the barrier, one can either provide strong incentive in form of attractive marketing, or lower the barrier by making it easy to install the virtual character support. Ideally, no installation should be necessary, i.e. the character should appear on the Web page immediately using a standard browser.

Published in the Software Focus Journal, 2(2):52-59, John Wiley & Sons, 2001.

Fast access

Although broadband-for-everyone is being promised by many, the reality is that most users do not have fast Internet access. Therefore the new applications involving Virtual Characters should not require high additional bandwidth. This implies several things. First, the virtual human models should not be too complex. Second, they need to be compressed for download. Third, both audio and animation data should not only be compressed, but also streamed, rather than downloaded and played. This allows for a faster response time.

Interactivity

When asked why Virtual Characters are better than video streaming, the easy answer is: they save bandwidth. But that is just the beginning. The main advantage is that they can interact with the user, i.e. provide immediate responses with respect to users' actions (mouse clicks, typing, in the future possibly speech, expressions and gestures). These interactions are unique for each individual user. This kind of interaction is impossible with video streaming, and this is the real advantage of Virtual Characters.

To achieve interactivity, several issues are important.

First, there needs to be an easy, preferably automatic way to generate contents. This is because no interesting interaction can be achieved without enough variety of contents. In simple words, a virtual character that can just say "yes" and "no" is not very interesting. If generating contents is expensive, designing the whole application will be very expensive. It is preferable to have the possibility to generate contents automatically. In the ideal case, the speech and behavior of the virtual character is completely generated on-the-fly, putting no limits to the variety of interaction. Unfortunately, this is in most cases opposing the requirement on the quality (best content is still created with a lot of manual work or expensive tracking equipment and talent), and a compromise must be found.

Second, it is desirable to have a fairly sophisticated decision-making mechanism, possibly based on AI algorithms, to deliver meaningful interactions.

Third, there must be some kind of user input. It can be within the virtual character application itself, or it can come through the rest of the Web page through standard Web page elements. This brings us to the next requirement, the Web integration.

Web integration

The virtual character will in most cases not appear alone on a Web page. There will be other contents: text, graphics, forms, buttons etc. It is important for the Virtual Character to be well integrated with all this, e.g. to be able to react to users' actions in the Web page, and to control and change the contents of the page. This implies that an open interface should exist to make logical links between the Web page elements and the virtual character, in both directions.

Having presented the potential applications and technical requirements, in the next sections we will review the current efforts in implementing this kind of systems. We will first briefly analyze the underlying technology of 3D graphics on the Web, in order to support the next section which reviews virtual human implementations on the Web.

3D on the Web

Although solutions for Virtual Characters using 2D technology do exist, for best results and full interactivity 3D graphics are required. We therefore briefly review the current options for delivering 3D on the Web.

Delivering full 3D content is currently not supported in major Web browsers. The standards that offer this functionality, like VRML [11] and MPEG-4 [38] are not yet widely deployed on average users' desktops. Therefore current efforts to deliver 3D content broadly align in two groups.

The plug-in approach requires the user to download a plug-in for their browser in order to view the 3D content. The most popular plug-ins provide general support for VRML viewing, like Blaxxun Contact [12] or Cosmo Player [13]. Other approaches are more proprietary and application specific, and usually associated with tools for content production, like Pulse 3D [14] Cult 3D [15] or Mendel 3D [16].

The second approach to delivering 3D on the Web is to bundle the rendering system in a Java applet. This requires the renderer to be implemented in an extremely efficient way, but very successful examples exist today, like Blaxxun 3D [12] and Shout 3D [17]. These implementations typically contain a 3D rendering engine and VRML engine in an applet of approx. 150 K. Shout3D offers a separate performance-enhancing OpenGL plug-in, allowing immediate 3D experience without the plug-in, and a boost in performance if plug-in is later installed.

Virtual Characters on the Web

In this section we present the state of the art in Virtual Characters on the Web. While we do not claim to provide an exhaustive list of all researchers and companies offering such solutions, we did attempt to cover as many solutions as possible, and certainly to capture the broad categories of solutions. We concentrate exclusively on technologies for delivering Virtual Characters on Web pages and do not discuss closely related technologies such as human modeling, animation, lip sync etc. In the review of individual products/applications we included only those that can be actually seen and tested on the Web (the URLs are included in the reference list).

While reviewing various approaches we try to focus on the criteria outlined in the section on Requirements, and at the end of this section provide a comparative table (Table 1) according to those criteria.

The current efforts in Virtual Characters on the Web can broadly be categorized in the following classes: text based bots, 2D animation, video streaming, 3D plug-ins and 3D without plug-in.

Text based bots

These applications, often called chatterbots, use artificial intelligence algorithms ranging from simple keyword search to more complex pattern matching or neural networks algorithms in order to perform a conversation with a human. The users are usually typing their input to the dialogue in a text field or prompt, and the answers come back as plain text. The earliest and probably most famous example of this technology is Eliza [18], the computer program capable of engaging a human in a conversation.

Currently companies like eGain Communications [19], Artificial Life [20] and Native Minds [21] offer bot-based products aimed at providing customer service on the Web. These products are purely text based and have no speech or animation capability. Usually an image of a friendly-looking person is shown next to the text input/output field, and sometimes images are changed with each new response in the dialog. The bot is programmed with a database corresponding to a relatively narrow subject field (e.g. product information) making it realistic to obtain very reasonable responses, especially to the questions that are frequently asked.

Although they offer no animation, we include chatterbots in this review because we believe that this type of application can successfully be merged with a graphical simulation of a human, as well as speech capability, in order to provide a full conversation experience.

2D animation

In this category we find implementations that provide animation by using a set of pre-rendered images. The images correspond to key postures of a face: visemes, expressions and movements. A viseme is the visual effect of pronouncing a phoneme and it represents a particular mouth shape. As mouth shapes for certain groups of phonemes (e.g. p, b, m) are highly similar, there are less visemes then phonemes. It is therefore possible to obtain a crude speech animation with a limited set of viseme images, typically between 8 and 15. To this are added a few images for expressions (e.g. smiling, surprise) and movement (e.g. nod). Therefore the total volume of data required for a simple but effective animation stays low, which makes it suitable for the Internet applications today, when broadband Internet access for the masses is still some time ahead. Additionally, fast manipulation of images is possible in current Web browsers using standard tools like Java or Flash. Low bitrate and standard tools mean that this approach is perfectly realistic and these applications easily accessible by virtually all Web users. The tradeoff is in accepting a lower quality and variety of animation then it would be possible with a full 3D solution (e.g. the face can not turn or smile while talking).

Examples of this approach include W Interactive [22] [37], with fully interactive, personalized Java-based Virtual Characters, and Virtual Personalities [23] offering an animation-enhanced chatterbot apparently based on Flash.

Video streaming

This approach consists in rendering a video sequence containing a virtual character off-line, then using off-the-shelf video streaming technology (e.g. RealVideo) to deliver the video sequence to the user. It inherits the disadvantages of the video streaming on the Internet of today, namely the small image size and low quality, unless using a broadband connection. Another drawback is the lack of interactivity. The user can only play the video or stop it. The potential of Virtual Characters to interact, deliver answers to specific questions, or provide customized contents is completely lost.

The example of this approach is Ananova [24], “the first virtual newscaster”, used as a gimmick in a brilliant marketing campaign to promote an otherwise conventional news site. The virtual newscaster comes in form of a link in the corner of the page that opens a video streaming window.

3D plug-ins

This is currently the most wide-spread approach to delivering Virtual Characters on the web. In some cases, a general-purpose 3D plug-in is used to deliver the characters. In other cases, the plug-in is designed specifically for the virtual character animation. We will review these two approaches separately.

General-purpose plug-ins

Although none of the currently available 3D plug-ins is truly widespread among Web users, these plug-ins offer a wider functionality and it is therefore more likely that they will be widely accepted and become de-facto standards.

As mentioned in the section about 3D on the Web, several plug-ins exist for displaying VRML contents in browsers. This enables us to exploit the functions of VRML and use them to display and

animate Virtual Characters. Babski and Thalmann [25] present various techniques for doing this, including sensors, interpolators, scripts and Java applets using the External Application Interface. However, same authors point to several problems in using current implementations of VRML plug-ins. They do not support all VRML functions in exactly the same way, and it is therefore difficult to design an application that will run in all configurations (usually creators recommend a particular plug-in). Also, the plug-in implementations are sometimes not very stable, and may demand a lot of resources.

A specific example from the work of Babski is the MPEG-4 Body Animation player [26]. It features a H-ANIM compatible body in the VRML browser, controlled through the External Application Interface (EAI) by a Java applet that interprets MPEG-4 FBA-compatible body animation sequences.

Another example of a virtual character in a VRML world with Java control through EAI is the Frank project [27], a low-resolution virtual human that can be commanded to walk and perform simple gestures.

A 3D plug-in that has rapidly gained popularity is Pulse3D. The Pulse3D production system allows easy export of animated objects from the widely used 3D design software 3DS Max. These objects and animations can then be played in Web pages using the Pulse3D plug-in. They are compressed in a proprietary format for relatively fast delivery, and the rendering capabilities of the plug-in allow for very good graphical quality. The use of a familiar tool, interactive production system and appealing results have attracted the computer animation community and Pulse3D has become quite popular.

Several independent productions [28] [29] [9] [30] illustrate the use of Pulse3D. In terms of visual quality, these are certainly among the finest animated Virtual Characters on the Web today. It is obvious that they have been designed by animation artists with great care for detail. However, it is also obvious that producing these demonstrations is relatively expensive in terms of artists' time and in some cases also in terms of tracking equipment and talent used. Most of the examples [28] [29] [30] feature animation synchronized to a pre-recorded real speech, and augmented with gestures and expressions to make it look appealing. The only example attempting a practical application is Cynthia Knows [28], a search engine augmented with the virtual character that gives comments and instructions about the search. Given the download time preceding each comment from Cynthia, and the fact that she takes most of the screen, making it difficult to see search results, it is hard to believe in the practicality of this particular solution. The problem of expensive production of animation also shows – there are only a few lines that Cynthia can pronounce, making her attractive appearance quite useless in practical terms. The other mentioned examples are entertaining, which is a value in itself, and they do not seem to pretend any use beyond this.

Plug-ins for Virtual Characters

These implementations require the user to download a particular plug-in (in one case [31] two plug-ins are required!) that is used only to deliver their virtual character contents.

On the positive side, this can enable a superior quality of animation. Furthermore, a custom plug-in can contain advanced audio and animation streaming capabilities, as well as interaction capabilities. Integration in the Web page can be enhanced by providing an open interface and giving Web designers full control over the virtual character from the rest of the page. However, not all of these opportunities seem to be exploited by the reviewed implementations.

The negative side of the custom plug-in is obvious: the user is required to download and install it just to access this particular content. In the best case, this is done automatically and the user just has to click on the Accept button. Even in this case, the users may not feel confident installing a virtually unknown piece of software on their computers. In the worst case, the installation takes time, effort and in the end may not work. In any case the user has to make an active decision, however minimal, to install the plug-in and therefore the incentive to do so must exist. This is not very convenient for introducing new services.

Haptek Inc. [10] offers the Virtual Friend, an entertaining application for chatting with a virtual person. The quality of animation in the standalone application is quite high. Unfortunately, we did not manage to install the plug-in and test the examples on the Web, nor the greeting card service that the company offers.

The First Agency for Virtual Models and Actors [31] aims to transfer its experience from live shows and TV to the Web. Two plug-ins are required, one for the animation and one for text-to-speech. The text-to-speech capability should offer interesting interaction possibilities as the speech can be generated locally in real time.

Famous3D [32] offers content-creation tools from either text or video. In case of video, animation is extracted by video analysis. In case of text, a process named text-to-animation is used, though from the description it is not quite clear whether this also involves text-to-speech, or is the audio produced and synchronized by some other means. The implementation uses a video clip paradigm. The file including the model, sound and animation is downloaded, then the full animation is played without any interaction possibilities. The quality of both models and animations is very good.

Mendel3D Avatar [16] claims to provide automatic lip sync from audio, but this does not seem to work very well in the presented examples which invariably featured jerky animation and interruptions in sound track. The scene and audio track are downloaded completely, then played as a video clip.

3D without plug-in.

Having pointed out the drawback of the plug-in approach, i.e. the plug-in itself, it is attractive to look for a solution that would work on most users' desktops immediately and without any installation. Currently such solutions are possible using Java applets, which are supported on the vast majority of desktops today.

This solution also has its downsides. First, an applet providing full 3D capability is necessarily fairly big, on the order of 150K, which is acceptable but not very convenient. Second, the performance of a Java applet is still lower than that of a well-implemented plug-in, limiting the complexity (and therefore visual quality) of the Virtual Characters that can be used.

On the positive side, beside almost universal portability, the applet programmer can implement streaming capabilities, as well as interaction support and seamless integration with the rest of the Web page.

Red Ted offers Javahead [33], an applet that displays a talking face. Javahead uses a proprietary streaming audio format and automatic lip sync with audio. This makes for the easy creation of contents based on natural speech recording, but the lip sync quality is quite poor. The faces rendered are fairly flat masks that rotate in the window. The motion is also automatic and looks unnatural. Except for playing an animation there is no other interaction possibility.

The author of this article has implemented a full 3D MPEG-4 [38] compatible facial animation system [34] based on the Shout3D [17] technology. This applet plays MPEG-4 FBA [4][6]

bitstreams that are streamed from the server in real time. The audio track is streamed as well. Currently provided facial models are fairly simple, but the possibility of creating the models and animation rules in popular commercial 3D software packages like 3DS Max leaves it open for more attractive models. As the demonstration page [34] shows, the applet is fully controllable from the Web page by Javascript, making all interactions possible. The virtual character can also bring up Web content in other frames in the Web page, synchronised with its talk. These features enable natural integration in a Web site. Limitations include the size of the applet, which is considerable at 200K and must be optimized; as well as un-optimized and sometimes unstable audio delivery.

The Interface project

This project [35], funded within the European Commission's Information Society Technologies programme, is unique in the fact that it unites a consortium of industrial and academic partners specializing in various domains relevant to the production of interactive Virtual Characters. We therefore felt it is justified to devote a separate subsection to it. The project includes specialists in computer graphics, Virtual Humans simulation, image processing, speech analysis and synthesis, as well as artificial intelligence, all working together to make man-machine interaction as natural as possible, based on everyday human communication by speech, facial expressions and body gestures.

The results will include tools for speech and video analysis for both low-level geometrical parameters and high-level emotional parameters; synthesis tools for automatically synchronized emotional speech and facial animation; and dialog management tools based on an open-source natural language artificial intelligence tool [36] with the addition of handling emotions.

The project plan includes fully integrated demonstrations on the Web, as well as commercial exploitation.

Comparison of implementations

In this subsection we try to provide a concise comparison of all implementations found currently on the Web and already mentioned in the previous subsections. Table 1 compares the implementations along the criteria discussed in the Requirements section of this article. Many of the criteria (e.g. visual quality) are evaluated subjectively, without undertaking precise measurements. These comparisons should therefore not be taken as an absolute measure, but are supposed to give an overall impression of the state of the art.

The presentation method corresponds to one of the main technical approaches outlined in previous subsections.

The visual quality of the virtual character (static) and its animation is judged subjectively by comparing all implementations.

The next criterion is ease of installation, evaluating how difficult it is for the user to get the virtual character up and running. Implementation where the installation is not needed are obviously the best here. Those depending on fairly widespread plug-ins (VRML, Pulse3D) are judged as easy installation.

Startup time is a measure of how fast the Web page will fully load and the character start talking, or performing.

In the "Streaming" column we summarize the data streaming capabilities of the implementations.

Speech generation and lip synchronization measure how easy it is to generate contents. Implementations use either text-to-speech or real speech recording, with manual or automatic lip synchronization (based on video or audio analysis, or on facial tracking techniques).

Published in the Software Focus Journal, 2(2):52-59, John Wiley & Sons, 2001.

Name	**Presentation method**	**Visual quality (static/anim.)**	**Installation**	**Startup time**	**Streaming**	**Speech generation/sync**	**Decision making**	**User input**	**Web integration**
eGain Assistant [19]	text bot	NA	not needed	very fast	NA	NA	AI	text	fair
Alife [20]	text bot	NA	not needed	very fast	NA	NA	AI	text	fair
VRep [21]	text bot	NA	not needed	very fast	NA	NA	AI	text	fair
Saphire [23]	Text bot, Flash 2D	good/poor	not needed	slow	none	real voice; probably manual sync	AI	text	fair
WebFace [22]	applet 2D	good/poor	not needed	fast	none	text-to-speech, auto sync	none	buttons, web page	good [3]
Ananova [24]	RealVideo plug-in	high/ medium[1]	easy	slow	video	text-to-speech, probably semi-auto sync	none	none	none
Virtual Friend [10]	custom plug-in	NA[2]	did not work	NA[2]	NA[2]	NA[2]	NA[2]	NA[2]	NA[2]
Nodna [31]	custom plug-in	medium/ medium	complicated	medium	None[4]	Text-to-speech, auto sync	none	web page	good [3]
famous3D [32]	custom plug-in	medium/ high	medium	medium	none	real voice, video analysis	none	none	none
Mendel3D Avatar [16]	custom plug-in	high/poor	medium	medium	none	real voice, probably auto sync	none	none	none
Ask Sal [29]	Pulse3D plug-in	high/high	easy	medium	none	real voice, man. Sync and/or face tracking	primitive	buttons, text	none
CynthiaKnows [28]	Pulse3D plug-in	high/high	easy	medium	none	real voice, man. Sync and/or face tracking	none	text	fair
Phony Blair [9]	Pulse3D plug-in	medium/ medium	easy	medium	none	real voice, man. Sync and/or face tracking	none	buttons	none
Sherer Digital [30]	Pulse3D plug-in	high/high	easy	medium	none	manual animation	none	none	none
Frank [27]	VRML plug-in	low/low	easy	slow	none	no speech	none	buttons	none
Babski [26]	VRML plug-in	medium/ medium	easy	slow	none	no speech	none	buttons	none
Javahead [33]	applet 3D	medium/ low	not needed	fast	audio	real voice, auto sync	none	none	none
Pandzic [34]	applet 3D	medium/ medium	not needed	medium	FBA, audio	MPEG-4 FBA[5]	none	web page	good [3]

[1] Providing enough bandwidth is available; otherwise quality may be poor

[2] Although this implementation seems interesting, and stand-alone application gives good results, the plug-in could not be correctly installed

[3] JavaScript controls the virtual character from the Web page; the virtual character can display Web pages in other frames

[4] Text-to-speech is installed locally, which makes streaming of audio and animation unnecessary

[5] Plays any MPEG-4 FBA bitstream regardless of its source; examples include TTS, facial tracking, video analysis and manual production.

Table 1: Comparison of different implementations of Virtual Characters on the Web with respect to the requirements discussed in the Requirements section of this paper

"Decision Making" refers to whether implementations include a decision-making mechanism, such that would allow more complicated interactions (conversations) than just straightforward reactions to mouse clicks.

Next we summarize the user input capabilities of each implementation. These capabilities enable interaction with the Virtual Character.

Finally, we compare how implementations are integrated with Web pages, i.e. is it possible for the virtual character to control the contents of the Web page and vice versa.

Conclusions

We have attempted to predict the possible applications for Virtual Characters on the Web based on current examples and research. We have analyzed the technical requirements for these applications and their implications. Based on these criteria, we have reviewed a large (though not exhaustive) body of current implementations of Virtual Characters specifically aimed at the Web. Finally, we have summarized this review in a table comparing all reviewed implementations along the defined criteria. So what can we conclude from all this?

- First, no current implementation satisfies all criteria.
- Second, solutions exist to satisfy each of the given criteria. The difficulty lies in combining them in one coherent system. Of course, we do not suggest that there is no room for improvement along each individual problem!
- Third, applications with Virtual Characters on the Web are possible today if one is willing to compromise between various parameters (e.g. fast, portable implementation but medium quality).

We believe that we are seeing a beginning of an emerging new class of interactive Web applications and services that will benefit from Virtual Humans technologies. We also believe that current technologies are very close to enabling such applications and that high quality animated conversational characters on the Web are within reach. On the commercial front, we are bound to see a plethora of competing implementations, a situation that may begin to settle down little by little once a standard solution for delivering 3D to the Web finally emerges. On a longer-term research front, there are numerous issues to tackle. One of the more interesting is automatically including natural-looking emotional cues and gestures in the Virtual Character responses, and synchronising these with equivalent TTS-synthesized auditory cues in order to automatically produce a natural, believable behavior of the Virtual Character – a task that will keep many researchers busy for a fairly long time.

Acknowledgments

This research is partly supported by the VISIT programme of the Swedish Foundation for Strategic Research (SSF) and the European Commission's IST project Interface [35].

References

[1] M.M.Cohen and D.W.Massaro, "Modeling Coarticulation in Synthetic Visual Speech." In M.Thalmann & D.Thalmann (Eds.) Computer Animation' 93. Tokyo: Springer-Verlag.

[2] Cosatto E., Graf H.P., "Sample-Based Synthesis of Photo-Realistic Talking Heads", Proc. Computer Animation '98, Philadelphia, USA, pp. 103-110.

[3] P. Eisert, S. Chaudhuri and B. Girod, "Speech Driven Synthesis of Talking Head Sequences," 3D Image Analysis and Synthesis, pp. 51-56, Erlangen, November 1997.

[4] Facial Deformations for MPEG-4, M. Escher, I.S. Pandzic, N. Magnenat-Thalmann, Computer Animation 98, Philadelphia, USA, pp. 138-145, IEEE Computer Society Press, 1998.

[HANIM] H-ANIM Humanoid Animation Working Group, Specification for a Standard Humanoid Version1.1, ece.uwaterloo.ca/~h-anim/spec1.1

[5] Kalra P., Mangili A., Magnenat-Thalmann N., Thalmann D., Simulation of Facial Muscle Actions based on Rational Free Form Deformation", Proceedings Eurographics 92, pp. 65-69
[6] Tekalp M.A., Ostermann J., "Face and 2-D Mesh Animation in MPEG-4", Image Communication Journal, Tutorial Issue on MPEG-4 Standard, Elsevier, 2000.
[7] "Computer Facial Animation", F.I. Parke, K. Waters, A K Peters Ltd. 1996., ISBN 1-56881-014-8
[8] "Synthetic Faces: What are they good for?" Igor S. Pandzic, Joern Ostermann, David Millen, The Visual Computer, 1999
[9] Tony Blair caricature using Pulse3D www.phonyblair.com
[10] Virtual Friend, Haptek Inc., www.haptek.com
[11] VRML, ISO/IEC 14772-1:1999, www.web3d.org/fs_specifications.htm
[12] blaxxun interactive, www.blaxxun.com
[13] Cosmo Software, Computer Associates International, Inc., www.cai.com/cosmo
[14] Pulse 3D, www.pulse3d.com
[15] Cycore Cult 3D, www.cult3d.com
[16] Mendel3D, Mendel 3D Avatar, www.mendelbox.com
[17] Shout 3D, Eyematic Interfaces Incorporated, http://www.shout3d.com/
[18] Weizenbaum, J., "ELIZA - A computer program for the study of natural language communication between man and machine", Communications of the ACM 9(1):36-45, 1966.
[19] eGain Assistant, eGain Communications Corp., www.egain.com
[20] ALife, Artificial Life, Inc., www.artificial-life.com
[21] vRep, NativeMinds, Inc., www.nativeminds.com
[22] WebFace, W Interactive SARL, www.winteractive.fr
[23] Saphire, Virtual Personalities Inc., www.verbot.com
[24] Ananova, Ananova Ltd., www.anannova.com
[25] "3D Graphics Define Virtual Humans on the Web", C. Babski, D. Thalmann, Software Focus 1(1), pp. 6-14, 2000.
[26] MPEG-4 BA applet, Christian Babski, LIG-EPFL, ligwww.epfl.ch/~babski
[27] Frank, Milan Polytechnic, www.lucia.it/vrml/projects/frank/eng_doc.html
[28] www.cynthiaknows.com
[29] Ask Sal, Excite Inc., arcadia.excitextreme.com/asksal
[30] Abraham Lincoln caricature using Pulse3D www.shererdigital.com
[31] Nodna, Nodna AG,www.nodna.com, demonstration: www.popkomm.de/popclash/pophead
[32] famous3D, famous3D Pty. Ltd., www.famous3d.com
[33] Javahead, Red Ted, www.redted.com
[34] "A Web-Based MPEG-4 Facial Animation System", I.S. Pandzic, Proc. ICAV 3D 2001 (to be published), demonstration at www.icg.isy.liu.se/~igor/MpegWeb
[35] The Interface project, IST-1999-10036, www.ist-interface.org
[36] A.L.I.C.E. natural language A.I. parser and chat robot, www.alicebot.org
[37] "From Photographs to Interactive Virtual Characters on the Web", Igor S. Pandzic, Gael Sannier, Proc. Scanning 2000, Paris, France
[38] ISO/IEC 14496 - MPEG-4 International Standard, Moving Picture Experts Group, www.cselt.it/mpeg