Cultural Preservation and Sharing using Multiplayer Virtual Worlds

Bart Thomee Michael S. Lew
LIACS Media Lab
Leiden University
{bthomee, mlew}@liacs.nl

ABSTRACT

We have developed technology that can both help preserve cultural assets and assist in sharing culture through the world. The focus of the technology has been a massive multiplayer framework which supports both cultural asset preservation and natural communication between the users through realistic 3D environmental shared audio. In this paper we present an evaluation of the audio quality and of behavioral changes of people in interpersonal interaction as they engage in this massive multiplayer virtual world. Additionally, we introduce the cultural preservation tools we have created.

Keywords: virtual reality, immersive environments, real-time audio communication, human interaction, cultural preservation and sharing

1 INTRODUCTION

One of the most significant scientific problems in the 21st century is cultural preservation and sharing as reflected for example in the major international and national scientific research frameworks throughout Europe and the world [1-5]. This year's budget of the European Union for the topic of 'Digital Libraries and Content' alone is already over 200 million euros [6].

In previous research and projects, the main emphasis was placed on the preservation of textual libraries and collections. Currently the preservation of other cultural items, such as historical sites, monuments and art, is found to be equally as important. Over the course of years historical buildings are taken down, sculptures destroyed or paintings stolen and thus digital preservation is not only worthwhile for ourselves, but also for our offspring and future generations. From a different perspective, the ability to easily share the experience of walking around historical locations and cultural objects is likely to increase the appreciation of such assets and can enhance understanding and respect for our own and other cultures.

The overall goal of our research is to develop technology which can both help preserve cultural assets and assist in sharing culture through the world. The focus of the technology has been a massive multiplayer framework which supports both cultural asset preservation and natural communication between the users through realistic 3D environmental shared audio. Currently we know of no other systems which support thousands of clients throughout the world where audio is properly handled in the 3D environment (i.e. other audio frameworks typically ignore the 3D environment or are not meant for very large multiplayer worlds [7-12]).

Besides giving an impression of the 3D editing tools we have designed for creation of cultural assets, we also evaluate the subjective behavioral changes of people in interpersonal interaction as they engage in our virtual world and try to answer questions such as "do people become more or less social/tolerant when having virtual conversations with known and unknown persons?" and "how natural do the virtual conversations feel to people?"

2 MASSIVE MULTIPLAYER FRAMEWORK

The optimal way to experience or share a particular culture is to go out and engage in conversations with others about the culture or to actually visit the native locations where the culture exhibits itself. Often, doing so is quite time-consuming, expensive and can even be dangerous. The digital world provides the ability to overcome these hurdles and mimic the real experience from the comfort of one's home.

We have created a virtual reality framework that integrates audio, video and 3D structure and as such allows players in the virtual environments to talk with each other as in real life. The environments can be designed to reflect any location, and in our case would pertain to sites of cultural interest such as museums (e.g. Guggenheim), monuments (e.g. Taj Mahal) and crowded plazas (e.g. Times Square).

2.1 3D Engine

For the visual and structural aspect of the framework, we used the 3D engine from the game Quake III: Arena. We revised its source code to correctly handle thousands of clients and added the 3D audio component as described in Section 2.2. The reason we chose the Quake engine is because the clients require the least network traffic; there are open source 3D editing tools for the 3D worlds available; and it natively has the option to split the video signal into binocular components for a more immersive 3D experience using either shutter glasses or a head mounted display.

2.2 Positional and Structural Audio

The main feature of the framework is a novel audio algorithm that deals with handling partial structural occlusion. The structural audio problem occurs when the 3D structure interacts with the audio signal; examples include simply going around a corner or walking into a room and closing the door. In both cases, the 3D structure affects the audio – typically lowering the amplitude but potentially also causing audio reflections and refractions. In some ways it is more complex than the hidden surfaces problem in that in most cases hidden surfaces are simply not drawn by the renderer. For structural audio, we can not simply cut off the audio. We must have a natural drop off due to the interference with the 3D world.

To determine the attenuation factor that should be applied to a sound, the algorithm employs Cauchy's probability distribution to weight a grid that is placed with its center at the listener's location, pointing at the origin of the sound. The 'audibility' of each point on the grid its determined by tracing the visibility between itself and the sound, while accounting for an obstruction of the direct line of sight, see Figure 1. The attenuation factor is formed by adding all weighted audibility values together, resulting in smooth sound transitions when moving around objects and corners while talking to other players.

The audio algorithm utilizes the Cauchy distribution because it (a) has been shown in other areas to be more realistic to real world distributions [13] and (b) in the future will allow us to adaptively adjust its parameters, e.g. simulate different environments or modifying sound perception through the use of in-game items. In previous research, sophisticated methods [10] have been described for real-time audio modeling in distributed virtual environments. However, these methods typically had the assumption of non-moving audio sources. In contrast, our framework supports moving audio sources in real-time and only requires low computational complexity.

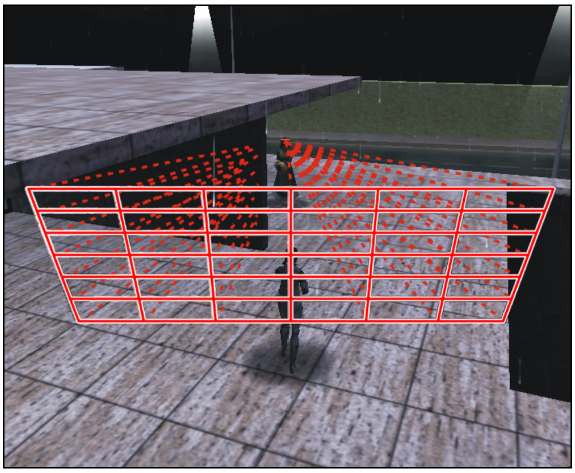

Figure 1. Tracing audibility

3 CULTURAL ASSET CREATION

One of our goals was to allow non-experts to create and share cultural assets. We designed a new interface toward enhanced usability for asset creation based on the GtkRadiant [15] editor. This tool was used by professional 3D model creators to design the virtual worlds of the original game. We enhanced it to make it more intuitive, and its effectiveness, facilitating the creation of cultural assets. See Figure 2 for a screenshot of the tool interface.

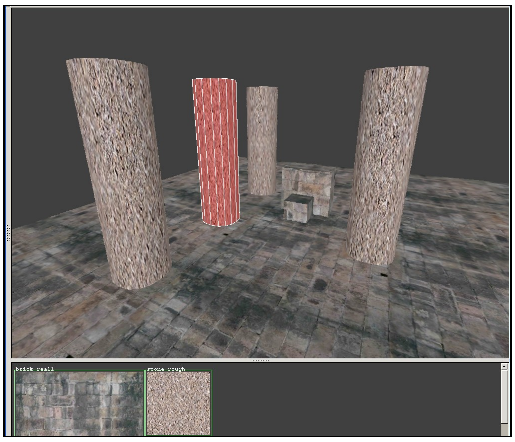

Figure 2. 3D cultural preservation editing tool. The red column is selected for manipulation.

Our primary objective at the moment is recreating cultural points of interest and at present we are working on a replica of Times Square, New York, as it is a landmark well-known across the globe and will serve as an example of the possibilities of our tool.

4 EXPERIMENTS AND EVALUATIONS

4.1 Audio Quality

We conducted several experiments to quantitatively measure the quality of the audio over the network with a focus on increasing numbers of players to determine preliminary user satisfaction.

4.1.1 Data Collection

The client systems involved in the experiments had Pentium 4 processors in the 1,8-3,0 GHz range with 512-1024 MB RAM and contained generic video and sound cards. The client network connections were a mixture of LAN and ADSL connected to a 3.4Ghz P4 server.

In a time span of approximately six hours, 100 clients joined the server. Initially, the groups were kept small, but with increasing amounts of clients, groups grew larger. Note that this also indicates that 'more connected clients' corresponds with 'more nearby clients' and vice versa. This reasoning is used when illustrating the experimental results. Latency could be determined only with a resolution of 10ms.

During these preliminary experiments, the selected 3D environment was a reconstruction from plans of Mies van der Rohe's German pavilion for the 1923 Barcelona Exhibition.

4.1.2 Evaluation

In Figure 3 we have plotted the measured latency against the number of clients that a client had within hearing range. ITU-T Recommendation G.114 [14] suggests that the mouth-to-ear delay, or latency, should not exceed 400ms and preferably be lower than 150ms. We can conclude that even though the latency increases with a growing number of clients, it remains within acceptable limits. In comparisons with typical mobile phones between students, the latency was found to be subjectively lower and the audio quality better. The curve of the graph might indicate that in overly populated areas the latency can grow unacceptably large; in the future we will perform experiments with more clients.

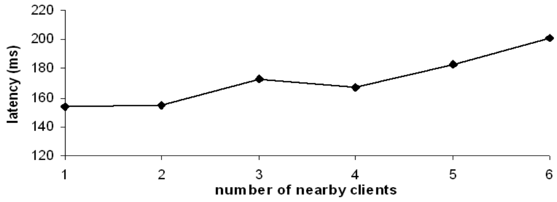

Figure 3. latency vs. number of nearby clients (client-side)

Figure 4 shows the packet loss against the number of nearby clients. Generally, the acceptability of an audio signal experiencing packet loss depends on the codec with which the signal is encoded, but the figure indicates that the amount of packet loss is small enough to have little to no effect on the audio quality. Packet loss at the server was negligible.

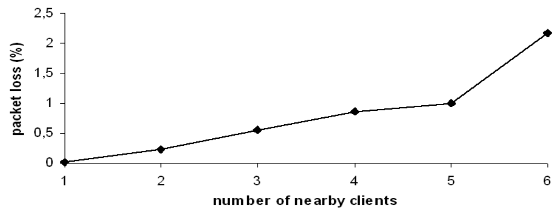

Figure 4. packet loss vs. number of nearby clients (client-side)

In a separate smaller scale session, we have also tested with intercontinental ADSL connections (typically 1.5Mbps downlink, 500kbps uplink) and found that the packet loss increases based on the distance. For the clients connecting from the USA to The Netherlands, there was approximately 1.3 percent packet loss.

The total amount of CPU usage by the server is graphed in Figure 5 against the number of connected clients. We notice that the CPU usage increases linearly with a growing amount of clients. Only a slight increase in memory usage by the server can be noticed in Figure 6, which is due to the allocation of new client objects in the audio server. Both these figures suggest that many more clients can be supported easily, provided a more powerful server to handle the expected increase in CPU usage.

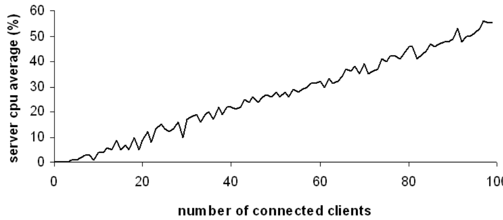

Figure 5. CPU usage vs. number of connected clients (server-side)

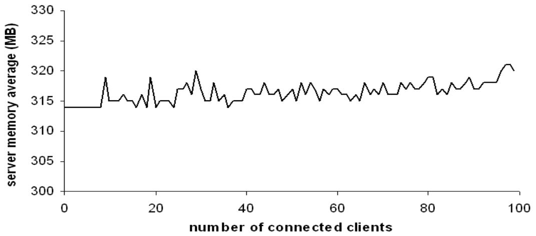

Figure 6. memory usage vs. number of connected clients (server-side)

In a third session of experiments, involving 10 students, the focus was on determining the subjective quality of the audio. Note that the following results are merely an indication of the quality; future work will involve user experiments on a much larger scale to obtain statistically valid data.

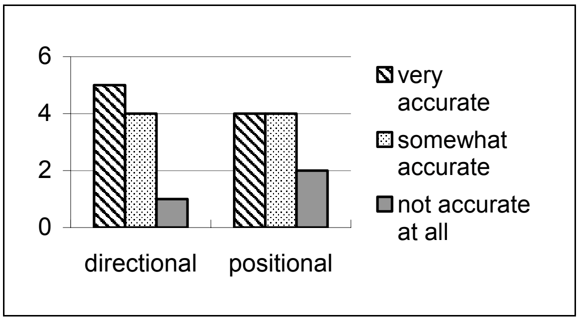

Figure 7. Accuracy of directional and positional audio

As can be seen in Figure 7, the users are inclined to indicate that the perception of directional audio (i.e. sounds coming from the left, right, front and rear), and positional audio (i.e. sounds from far away and close by) is accurate in both cases. Some users noted that in the case of directional audio the distinction between left and right was more obvious than that between front and rear, and in the case of positional audio that the sounds carried a little too far compared with reality.

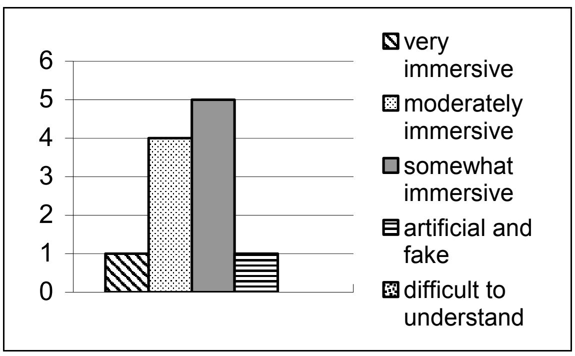

Figure 8. Sense of immersion with 3D audio

In general, the users felt that the 3D audio gave them a reasonable sense of immersion/reality, see Figure 8.

4.2 User Behavior in Different Worlds

During the third set of experiments, we also monitored users as they engaged in i) an online multiplayer game where conversing was only possible using text-based chat and ii) an online multiplayer game with our 3D environmental shared audio system. Their experiences were compared between the two situations and were also compared with having conversations in real life.

According to the users, the experience of chatting using text was not really comparable with real life, while chatting using 3D audio felt much more natural, see Figure 9.

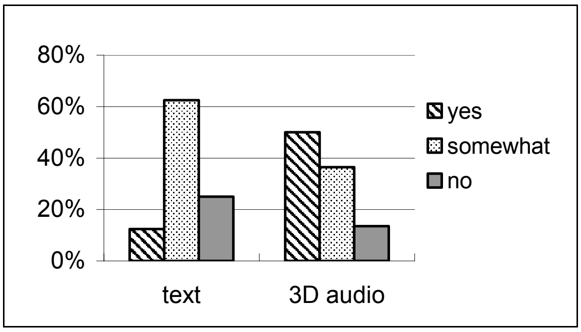

Figure 9. Naturalness of conversations, compared with reality

The inability to communicate non-verbally in the online game had a larger negative effect on the conversation when chatting using text; with 3D audio this influence was less and the users were better able to understand each other. The users generally agreed that the use of styles such as irony, sarcasm and humor was better conveyed to others through 3D audio than through text.

The majority of the users felt that their perception of the character of the person(s) they talked to was altered because of the communication medium and also felt that the perception of their own character was altered. Compared with real life, the users thought that text-chat changed the perception of characters in both ways to less friendly, less tolerant, less social and more aggressive. In contrast, communicating using 3D audio changed the perception of characters to more friendly, more tolerant and more social, but also still more aggressive than in real life. See figures 10-13 for the results.

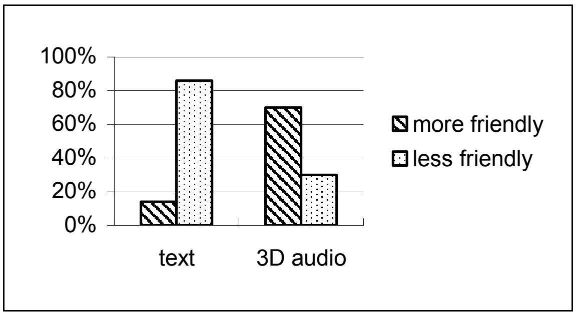

Figure 10. Altered perception of characters: friendliness

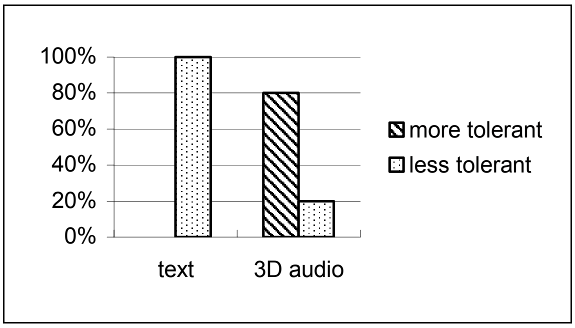

Figure 11. Altered perception of characters: tolerance

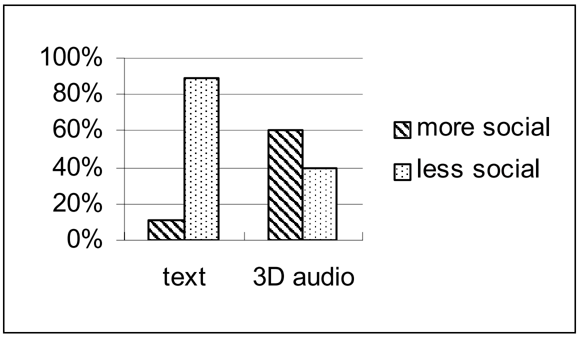

Figure 12. Altered perception of characters: sociality

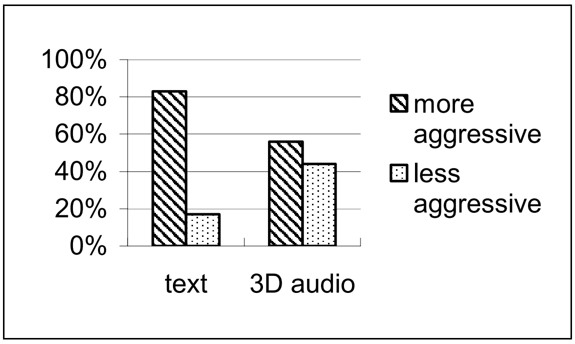

Figure 13. Altered perception of characters: aggressiveness

During the experiment we let users talk to people they were already acquainted with and with people they did not know yet, to see to what extent their behavior would change, e.g. would they be more open when meeting a new person or would they rather be more closed than in real life. However, the experiments were too small scale to draw any significant conclusions and we leave this to future work to determine on a larger user base.

One of the main drawbacks of using text to chat is that typing a message is slow and takes away the focus of the game – when typing, the in-game character cannot do anything. When talking to multiple people simultaneously however, this drawback is turned into an advantage as messages of other users can be read at an easy pace. With audio-based chat, conversations have the same speed as conversations in real life and this posed a problem for the majority of users. The main reason was that the voices captured by the microphones had a somewhat 'metallic' sound and caused confusion in determining who said what; the reasonable accuracy of positional and directional accuracy could not prevent this. Despite this shortcoming, the users generally felt that chatting with multiple people simultaneously using audio was still much more realistic than chatting using text.

5 DISCUSSION AND CONCLUSIONS

In this paper, we presented technology that can both help preserve cultural assets and assist in sharing culture through the world. Our main focus has been a virtual reality environment which supports both cultural asset preservation and natural communication between the users through realistic 3D environmental shared audio. The novel Cauchy-based audio algorithm provides immersive sound attenuation through its knowledge of the structure of the virtual world and only requires minimal CPU effort for real-time modeling. Through the use of our 3D editing tools, cultural assets can be digitized and thus preserved. Moreover, people in the virtual worlds can view these objects and discuss them in a natural fashion.

Our user experiments have shown that the 3D audio quality was accurate and immersive; compared with traditional text-chat our 3D audio provided the possibility of having realistic conversations in the virtual world. In addition, the users found that they became friendlier, more tolerant, more social and less aggressive through the use of audio communication, in comparison with using text.

REFERENCES

[1] European Framework Program 6, “Access to and preservation of cultural and scientific resources”, http://cordis.europa.eu/ist/so/culture/home.html

[2] European Framework Program 6, “Learning and Cultural Heritage”, http://cordis.europa.eu/ist/telearn-digicult/index.html

[3] European Framework Program 7, “Digital libraries and technology-enhanced learning”, ftp://ftp.cordis.europa.eu/pub/ist/docs/telearn/plenary_challenge4-1.pdf

[4] National Endowment for the Humanities, “National Digital Newspaper Program”, http://www.neh.gov/projects/ndnp.html

[5] Virtual Museum Canada, http://www.virtualmuseum.ca/

[6] European Framework Program 7, “ICT in FP7”, http://ec.europa.eu/information_society/research/documents/fp7-ict-4poverview.pdf

[7] P. Boustead, F. Safaei, M. Dowlatshahi, “DICE: internet delivery of immersive voice communication for crowded virtual spaces”, *in IEEE Virtual Reality Conference 2005*, pp. 35-41, 2005.

[8] C. Greenhalgh, S. Benford, “MASSIVE: A collaborative virtual environment for teleconferencing”, in *ACM Transactions on Computer-Human Interaction*, pp. 239-26, 1995.

[9] J. Robinson, J. Stewart, I. Labbe, “MVIP - audio enabled multicast VNet”, in *Proc. of Symposium on Virtual Reality Modeling Language*, pp. 103-109, 2000.

[10] T. Funkhouser, P. Min, I. Carlbom, “Real-time acoustic modeling for distributed virtual environments”, in *Proc. of the 26th annual conference on Computer Graphics and Interactive Techniques*, pp. 365-374, 1999.

[11] M. Naef, O. Staadt, M. Gross, “Spatialized audio rendering for immersive virtual environments”, in *Proc. of the ACM Symposium on Virtual Reality Software and Technology*, pp. 65-72, 2002.

[12] C. Nguyen, F. Safaei, D. Platt, “On the provision of immersive audio communication to massively multi-player online games”, in *Proc. of the IEEE ISCC 2004*, pp. 1000-1005, 2004.

[13] N. Sebe, M. Lew, N. Huismans, “Toward Improved Ranking Metrics”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1132-1143, 2000.

[14] ITU-T Recommendation G.114, “One-Way Transmission Time”, http://www.itu.int/rec/T-REC-G.114, May 2003.

[15] GtkRadiant, http://www.qeradiant.com