Eurographics Symposium on Virtual Environments (2004)
S. Coquillart, M. Göbel (Editors)

Integrating Real-time Binaural Acoustics into VR Applications

I. Assenmacher†, T. Kuhlen†
T. Lentz‡, M. Vorländer‡
Center for Computing and Communication RWTH Aachen University†
Institute of Technical Acoustics RWTH Aachen University‡

Abstract

Common research in the field of Virtual Reality (VR) considers acoustic stimulation as a highly important necessity for enhanced immersion into virtual scenes. However, most common VR toolkits do only marginally support the integration of sound for the application programmer. Furthermore, the quality of stimulation that is provided usually ranges from system sounds (e.g. beeps while selecting a menu) to simple 3D panning. In the latter case, these approaches do only allow the user to correctly detect sounds that are at quite a distance from his current position.

Binaural synthesis is an interesting way to allow the spatial auditory representation by using few loudspeakers or headphones. This paper describes a system that combines the efforts of creating a binaural representation for the listener who is interacting in a common visual VR application in real-time, thus allowing the research on interaction between visual and auditory human perception systems. It will describe the theoretical background to establishing a binaural representation of a sound and the necessary hardware set-up for this. Afterwards, the infrastructure and software interface which will allow the connection of the audio renderer to a visual VR toolkit is discussed.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Three-Dimensional Graphics and Realism]: Virtual Reality, 3D Audio, VR-Toolkit, Binaural synthesis, Cross-talk cancellation

1. Introduction

Virtual Reality (VR) environments aim to immerse the user in a computer generated world. In order to achieve this, in theory, all human sensory-systems have to be stimulated in a natural fashion. In practice, most modern VR systems show a well established set of methods to do this for the human visual system. As the visual system is considered to be an important source of information in human perception, most efforts were put into this field of science. Haptics are a rather well practised stimulation method, too. There exist different commercial devices like the Phantom haptic device or the haptic workstation, among other force feedback devices. Acoustic stimulation,too, is considered to be a very important issue for the natural perception of simulated worlds [FMC99],[NSG02].

The utilisation of the human visual system in common VR applications is the presentation of a computer generated representation for each of the two eyes which enables stereoscopic, more natural views. The obvious analogy, that humans have two ears as well as two eyes, is a rare research topic in the field of VR. This is a major drawback in current VR systems, as auditory feedback from the environment enhances the liveliness and, as a result, the user's immersion into the virtual scenery. To enhance the immersion into a virtual scene, an audio system that performs spatial sound is required. But adding "convincing" spatial sound to a visual virtual scenery is not trivial. Position, direction, level and distance of each sound source must be synthesised and reproduced according to the visual stimuli. Many commercial solutions for *animated* scenery exist in the field of entertainment. The popularity of home cinema systems increased along with the launch of digital recording formats, which allow storing multi-channel audio information. The goal is to immerse the listener into the scenes of a film with a multi-speaker configuration. Speech, sound effects, and background music, among other sound types, are distributed to the different channels to create the surrounding experi-

© The Eurographics Association 2004.

delivered by EUROGRAPHICS DIGITAL LIBRARY
www.eg.org diglib.eg.org

ence. These systems achieve good performance, but are an insufficient solution for more sophisticated, interactive applications.

Immersing a listener in an auditory sound scene, or, in other words, three dimensional sound spatialization, is not an easy field in audio engineering. Many solutions and implementations have been proposed through the years, based mainly on two theoretical approaches. The first approach is it to surround the listener with many loudspeakers (not only four to six as used by 5.1 systems). By doing this, a similar sound field as in a real acoustical environment is reproduced. The second, which will be the focus of this work, is the binaural audio approach. This consists of reproducing exactly the same signal at the eardrums of the listener as a real listening situation would produce. The name of the method already suggests that the generation of two signals is required, one for each ear.

Unfortunately, common VR systems do only provide basic support for auditory stimulation during VR simulations. A very common case of use seems to be the possibility to provide feedback on simple interaction schemes like selection feedback or system sounds in general[Bie00], [CPGB94]. Needless to say that these sounds remain ”flat” in terms of their spatial representation. In addition to system sounds, almost all approaches make use of the fact that VR systems track the user's head position and allow simple spatial sound conversions on top of the possibilities of sound libraries like OpenAL[Ope] or the possibilities of the WorldTool-Kit[Sen], Syzygy[SG02] or AudioJuggler, as a part of VRJuggler[Bie00]. However, few VR systems provide enhanced algorithms and approaches to near head auditory stimulation. Recent efforts of the authors provide a comprehensive lightweight plug-in architecture for enabling sophisticated methods that combine visual and auditory stimulation in modern VR applications.

2. Binaural hearing

Due to the fact that humans hear with two ears, a direction can be assigned to sound events[Bla97]. As for visual stimuli, the brain compares pictures from both eyes to determine the objects' placing in a scene and with this information, it creates a three-dimensional cognitive representation that humans perceive as a three dimensional image. In straight analogy, stimuli that are present at the eardrums will be compared by the brain to determine the nature and the direction of a sound event. Depending on the horizontal angle of incidence, different time delays and levels between both ears consequently arise. In addition, frequency characteristics dependent on the angle of incidence are influenced by the interference between the direct signal and the reflections of head, shoulder, auricle and other parts of the human body. The interaction of these three factors permits humans to assign a direction to acoustic events[Møl92]. The characteristics of the sound pressure at the eardrum can be described in the time domain by the Head-Related Impulse Response (HRIR) and in the frequency domain by the Head-Related Transfer Function (HRTF). These transfer functions can be measured individually with small in-ear microphones or with an artificial head. Figure 1 shows a measurement of the ITA artificial head under 45° relating to the frontal direction in the horizontal plane. The interaural time difference ITD can be assessed in the time domain plot (HRIR). The interaural level difference (ILD) is shown in the frequency domain plot and clarifies the frequency dependent level increase of the ear turned towards the sound source and the decrease at the ear that is turned away from the sound source.

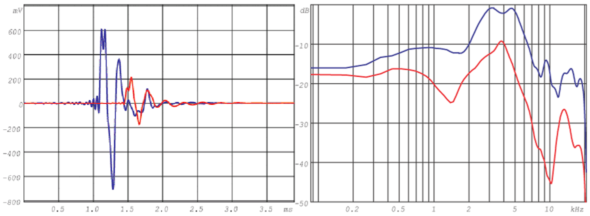

Figure 1: *HRIR and HRTF of a sound source under 45°. It was measured using the ITA artificial head.*

3. Sound reproduction

As mentioned in the introduction, there are mainly two different approaches to reproducing a sound event with true spatial relation, i.e. wave field synthesis and binaural approaches.

3.1. Wave field synthesis

The basic theory of the *wave field synthesis* is the *Huygens principle*. To reproduce an entire real sound field, an array of loudspeakers (sometimes more than 100) is placed in the same position as a microphone array was placed at the time of recording the sound event. Figure 2 shows the principle of recording and reproduction[The03].

The main drawback of this system, beyond the high effort of processing power, is the size of the loudspeaker array. Furthermore, mostly 2D solutions have been presented so far. The placement of those arrays in video projection VR systems like a Holobench or a Powerwall is only just possible, but in display systems with four to six surfaces like a CAVE it is nearly impossible.

3.2. Binaural (transaural) approach

A *binaural signal* represents the sound field at the input of the auditory channels. The major problem is the reproduction of a signal at these positions in such a way, that the listener perceives it as natural. It is convenient and sufficient to reproduce that sound field only at two points, the ears,

© The Eurographics Association 2004.

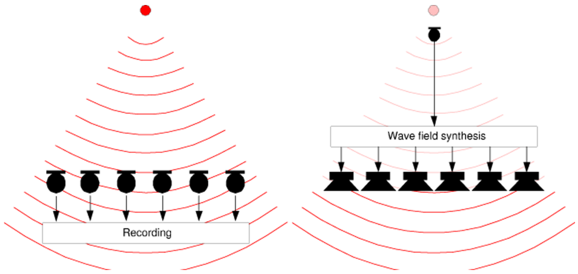

Figure 2: *The left figure depicts the recording set-up of a sound field and a microphone array for wave field synthesis. For replay, which is shown on the right hand side, the recorded signals can be reproduced directly using an array of loudspeakers. This can alternatively be simulated using a wave field synthesis filter.*

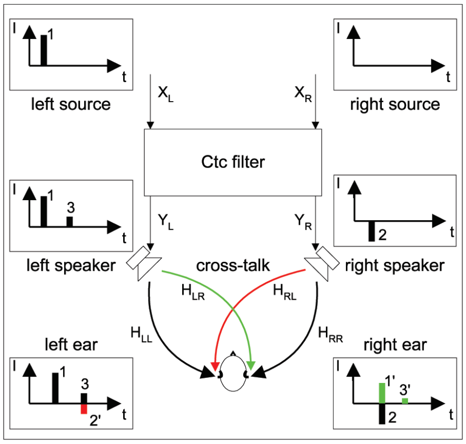

Figure 3: *The principle of cross-talk cancellation. The desired signal for each ear is shown as input to the Ctc filter. The filter adds the inverse cross-talk signal, resulting in a proper reproduction of the desired signal at the listener's left and right ear.*

and not over the complete space. To realise this requirement two technical solutions are appropriate, *headphones* and the *dynamic cross-talk cancellation*.

Headphones

From a technical point of view, the presentation of binaural signals by headphones is the easiest way[Beg94], as acoustic separation between both channels is perfectly solved. But equalization of headphones and reproducibility when headphones are removed and replaced is not a trivial task. However, most results obtained are unsatisfying in the subjective sense of listeners. Furthermore, when the ears are covered by headphones, the impression of a source located at a certain point and distance to the listener often does not match the impression of a real sound field. Another point is that the high effort presenting 3D video in a CAVE instead of using a technically considerably simpler Head Mounted Display and the wearing of headphones do not fit together. For this reasons a reproduction by loudspeaker can be considered to create fewer problems concerning the naturalness of the presented sound and the acceptability of the environment.

Dynamic cross-talk cancellation

The well-known problem of loudspeaker reproduction is the cross-talk between the channels that destroys the three-dimensional cues of the binaural signal. The requirement for a correct binaural presentation is, that the right channel of the signal is audible only in the right ear and the left one is audible only in the left ear[Sch93]. This problem can be solved by a cross-talk cancellation filter which is based on the transfer functions from each loudspeaker to each ear. These transfer functions are shown in Figure 3 and labelled H_{LL}, H_{LR}, H_{RL} and H_{RR}. One of the first approaches in this field was made by Bauer[Bau63]. Cross-talk cancellation for a moving listener requires a system being able to provide a valid filter-set for each position[Gar97]. In the approach presented here, filters are calculated online using an HRTF database and a head tracking system to detect the listeners current position. The software is implemented on a PC architecture realising a cross-talk cancellation structure which is able to change its parameters in real-time[LS02]. To calculate the correct filters, a head tracking system transfers the listener's position and orientation to one of the cluster PCs which sends the information to the video rendering PCs and the audio PC. Depending on this information, HRTFs will be chosen from the database. After the calculation of a new filter-set, the cross-talk cancelling structure will be updated. In this way, the current binaural audio signal is filtered with the correct cross-talk cancelling filter for the listener's present position. Using a 1GHz processor, the latency of this system is below 20ms, which is short enough for real-time applications. The fact that the cross-talk cancellation is valid only for one listener is not a disadvantage in those interactive position-dependant 3D video systems, because the optical image is also generated and valid for one user. The input compatibility which makes it possible to use the headphones for reproduction during the development of a sequence to control the audio component at a normal PC workplace and use the cross-talk cancellation system in the CAVE is an additional plus. In both cases the same binaural signals are used as input. This is does not happen in the wave field synthesis.

4. Binaural synthesis

Filtering a sound source with an appropriate HRTF pair in order to obtain a synthetic binaural signal is called *Binaural Synthesis*. The synthesised signal contains the direction information of the source which is provided in the HRTFs. In other words, the binaural synthesis transforms a signal without position information into a signal representing the sound

© The Eurographics Association 2004.

pressure at the ears of a listener from a virtual source located at the point defined by the HRTF. Reproducing the resulting signal over a dynamic cross-talk cancellation system should enable us to create the illusion of sound emanating from any direction in 3D space. Superposition allows us to extend the process in order to generate N different sources at different positions. The resulting signal is again binaural. In a video virtual reality system the position of any object which can be rendered with a sound, too, is given by the scene graph in a world coordinate system. To chose the right HRTF from the database, only the relative position to the user has to be calculated. When the listener moves his head, he changes the relative position, which is provided by changing the used HRTF, so that he is able to locate the virtual source every time at a fixed position in the world coordinate system. Furthermore, moving sources are also possible.

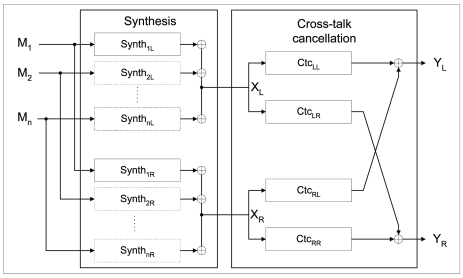

Figure 4: *Combined binaural synthesis and cross-talk cancellation filter. Individual spatial information for N mono sources will be assigned by the binaural synthesis. The resulting binaural signal will be passed to the loudspeakers over the cross-talk cancellation.*

Influence of reflections

The localisation results in an anechoic chamber, an ideal acoustical environment, are very good, but when using the cross-talk cancellation and the binaural synthesis together with a video VR-system the influence of reflections has to be examined to ensure the applicability. All screens for video projection cause also interfering reflections as well and may reduce the ability to localise the virtual sources at the correct position. A preliminary hearing-test was performed to investigate the difference of the localisation comparing an ideal environment to an environment with reflecting walls around the test person. The results are very good despite of the reflections[Len03]. Due to the fact that reflections arrive at the ears later than the direct signals, humans are still able to detect the right direction of the source. These tests show that cross-talk cancellation together with binaural synthesis is an appropriate technique for sound reproduction in virtual reality systems. A more detailed description of the influence of reflection influence is still a topic of further research.

5. Coupling video VR with audio VR, a list of requirements

The following passage will describe the general soft- and hardware systems and settings which are needed for and provided by the presented work. The complete system is embedded into a vivid environment of VR infrastructure as CAVE, Holobench and Powerwall display technology. It is working well on standalone PCs (for downscaled applications) and fits perfectly into PC cluster architectures. As a transport medium, standard 100MBit Ethernet was used.

Conventional Virtual Reality systems have a strong focus on presenting visual stimuli by means of a traversed scene graph that captures all the necessary information for the detailed graphical layout of the virtual scene. In order to enable stereoscopic views and immersion through user-centered projection, the head motion is measured with tracking devices such as the Flock of Birds, Intersense's supersonic tracking system or optoelectronic systems like ART-track. Apart from the existing differences in the underlying concept of each system and their measuring quality, all these systems deliver samples of the position and the orientation of measure points in a set-up specific rate.

The data rate that the tracking hardware usually delivers is well suited to or synchronized with the repaint rate of the video devices used and application needs. Recall that a VR application needs a video refresh rate of at least 20-30 Hz in order to be considered interactive. A typical VR application does not need to catch every sample that is provided by the measuring hardware to immerse the user. However, human acoustic perception is more sensible to signal distortion or disruption. Thus, utilising binaural acoustics raises the requirement of delivering the tracker samples at the best available rate, which is usually determined by the tracking hardware used. Consequently, that the tracking information needs to be available to different system components at different rates (e.g. the video signal is produced on the last incoming tracking data with a lower frequency than the computation of the audio signal). In addition, the computational power that is needed for the binaural acoustics and the graphical rendering almost dictates that the processing for each of those two is done on different host computers which are interconnected over a network. Recall that the aim of this efforts is to produce high quality VR audio in conjunction with visual stimulation and not only the presentation of flat or pseudo 3D sounds for indicating system responses.

In terms of software requirements, it is clear that samples from the tracking hardware have to be processed as fast as possible in addition to the fact that the information has to be routed through the VR application, as the tracking data is used to update the visual presentation and the auditory information. This raises the need for an infrastructure that enables asynchronous processing and flexible handling of the data that is being generated by the tracking hardware. The integration of a software architecture that can fulfil the above

© The Eurographics Association 2004.

requirements will enable application programmers to incorporate near-body 3D acoustics into their VR applications and existing infrastructure. The described architecture is implemented completely in the VR system ViSTA [vRKG*00] which is developed by the Virtual Reality Center Aachen (VRCA) in C++. ViSTA covers all aspects of VR technology. It is scalable in order to integrate all kinds of display systems, ranging from high-end visualization displays like CAVE systems down to standard desktop monitors, and offers a variety of 3D input devices. It is implemented as a comprehensive toolkit, so that VR applications for a specific purpose can be developed rapidly.

6. Technical design

The following section will present the technical design of the coupling of the two systems from the ITA and ViSTA. It will focus on the static system architecture and then describe the communication layer in detail.

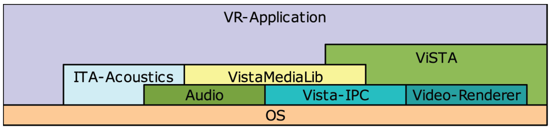

Figure 5: *The system is implemented on top of well defined interfaces to the ITA-acoustics package and the ViSTA VR-toolkit. The communication is performed by means of the VistaInterProcComm module, a platform independent library that defines inter-process-communication and synchronization patterns. Audio integration is part of the ViSTA-Media Library (Vml), which is a new module of ViSTA. It serves as a basis for the integration of general multi-media aspects in VR applications.*

6.1. Design of the audio- and video-system coupling

The most important aspect of the interface design was the possibility to write sound-enhanced VR applications in a rather simple manner, using the information that is naturally provided by any typical VR system. Additionally, the usual sound in VR applications that is used to reflect certain special system states which require attention should be possible. The coupling between the visual VR system ViSTA and the auditory VR system of the ITA was designed to be loose, so that the systems retain their scalability and flexibility.

Due to the computational power that is needed by the binaural synthesis, the system in general is installed as a client-server networked topology, where the audio host is a constantly running server to which exactly one client can connect in order to use it. Note that any client can connect to more than one server. As soon as the client does not need the service anymore, the server regains its initial state and waits for another client to provide sound services to.

The interface of the sound system presents a thin layered abstraction to sound processing as seen from an applications programmer's point of view. It is possible to register sounds and important parameters symbolically within an *init phase*. After that, the application marks the ending of the init phase and begins to process on the *changes* of the virtual scene. During user interaction and navigation the application can start, stop and modify sounds freely. A typical VR application encounters a finite number of distinct sounds that are updated in position and direction during the execution. Figure 6 depicts this protocol in more detail.

Note that applications only need to define a mapping between sources of change in the virtual world and the sound identification that is known to the sound subsystem. This mapping is realised by the generator abstraction, which is later explained in more detail.

The transformations of all objects are considered to be highly versatile during run-time, which means that sources as well as the listeners can change position and orientation during run-time at any step. In case of a change, the updated information about sound sources is given to the audio subsystem on demand of the application. However, the system will enforce such a notification at any change of the head position of the listener.

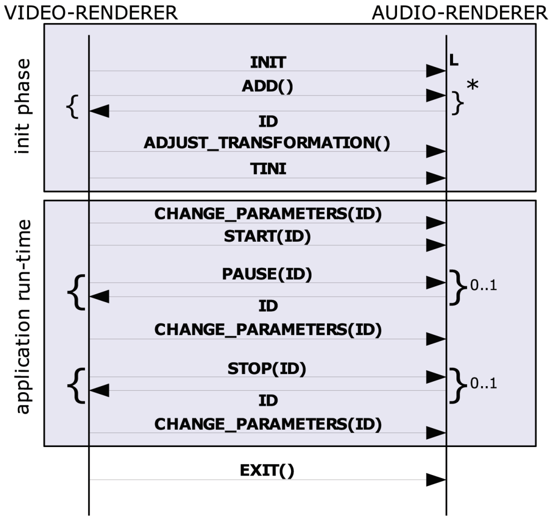

Figure 6: *The figure shows the communication protocol between the video-rendering unit and the audio-rendering server. Basically, any VR application must distinguish between the init-phase, the run-time phase and the exit-phase, which will perform cleanup. During the run-time phase, it is basically allowed to change sound parameters only when the sound is not played (e.g. paused or stopped), which is not true to change of spatial information, as this can change at any time.*

© The Eurographics Association 2004.

6.2. Software abstractions to generate spatial data

The ViSTA application provides the application programmer with an interface that is very general and is implemented by the binaural renderer of the ITA. It does allow the registration of a listener, sound sources and sound data. Listeners and sound sources do have a position and an orientation in the virtual world. In addition to that, the listener has a position and orientation in the real world. The latter is continuously updated by the attached tracking hardware. On the software level, a *generator* serves as technical base class to define a template algorithm for producing a virtual world position and orientation, which can then be processed by the audio computation. This enables a flexible way of associating various data sources in virtual environments, e.g. hardware, logical devices, nodes in a scene graph or even functions that generate data from scripts, as inputs. A generator serves as a source of a spatial data that is specific to a single VR world entity, e.g. there can be generators for specific sub trees in a scene graph that "react" to the change of the scene graph and propagate this change to the audio subsystem. The advantage of indirectly binding generators to model entities of the VR application is that any change in position or orientation of the entity is automatically transferred into the audio subsystem, there is no need for the application programmer to explicitly call an update method or to keep track of all sound entities in proprietary datastructures. The current implementations in ViSTA enable the choice between data updating strategies that are synchronous and concurrent to frame refreshing. This is defined by template algorithms that are technically defined through instances of *updaters*.

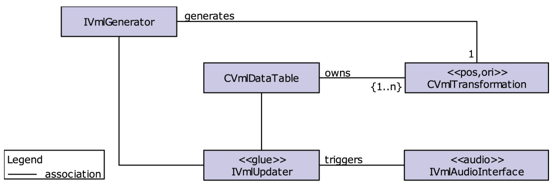

Figure 7: *Generators provide the application programmer with an interface to map application events to sound transformations. Updaters define a template algorithm that selects when and how updates will be processed. The data table is a collection of spatial data (transformations) for all registered sound entities in the VR application.*

One sensitive point worth mentioning is the description of certain attributes that will have an impact on the rendering of a given sound. In addition, the overall behavior of sounds has to be specified. Primitives for sound behaviors are defined in looping and stopping. One basic idea was to allow the implementation to be flexible enough to introduce new sound attributes during research without breaking existing code. Attributes are therefore set by using symbols that can be ignored or respected by the underlying implementation of the interface. The application is enabled to change sound parameters that will have impact on the overall perception of sounds by the interacting user during the usual update or application cycle in between frame refreshing.

6.3. Transparent network protocols

There are clearly two parties that are involved in network communication, one party for the audio rendering, and one for the video rendering. Physically, those two can be distinguished as the machines that contain either a video (visual rendering) or the sound-card (audio rendering). The machine that provides the audio rendering is called the *server* machine in a client-server layout that is described in the following passage. A static class layout that describes the implementation is given in figure 8.

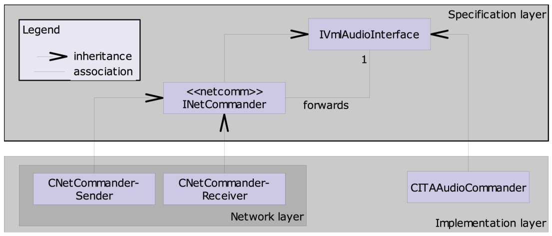

Figure 8: *The protocol that enables the networked architecture is transparently implemented by forwarders for the audio-server (receiver of spatial updates) and for the video-client (sender of spatial updates). The application programmer only programs the audio-interface and has not to worry about the network communication.*

As said before, the ViSTA implementation provides a transparent network layer that enables sound rendering on a remote host in real-time.

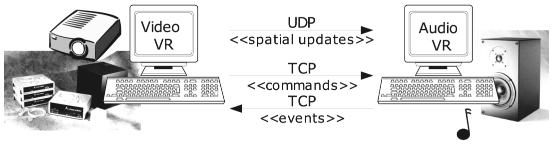

Figure 9: *The set-up for the described system. Video and audio stimuli are rendered on different hosts. Spatial input data is processed in different rates, as the auditory rendering needs the information typically at the highest possible rate. Physically, state change is sent over a UDP connection, which allows fast transfer. Transmission errors can be omitted, assuming a continuous universe.*

All net communication is hidden from the application programmer in a separate layer. As stated above, there is no difference to the application layout as seen by the VR programmer whether there is a network set-up involved or not. The decision about this is made by dynamic binding during

© The Eurographics Association 2004.

run-time. The transparency aspect is realised by a specific implementation that hides the complete network communication. The communication scheme defines a bidirectional protocol where all methods and arguments are encoded using a low level serialization technique. Data transfer respects byte order and can thus be used in heterogeneous system environments, e.g. an IRIX host for presenting the VR scene while doing sound rendering on a Win32 client.

Client and server communication is modelled for both ways, in terms of commands and spatial updates from the client to the server as well as notification and error passing from the server to the client, by means of different and independent channels, which will be presented in the following passage (see figure 9 as well).

We distinguish between three channels of communication. The first one to be described is called the *command channel*, which is used to control the audio server from the video rendering machine. This channel is used to register, modify, delete and play sounds during the run-time of the VR application. Clients give commands with well defined parameters and receive answers of known type, a simple RPC mechanism. For example a command like *StartSound* is serialized over this channel as a sequence of command token, the sound id and the loop parameter. This message is dispatched on the server side, executed and the result (in this case a boolean value indicating success or failure) is sent as an answer to the client (see figure 6 as well). This channel is, although inherently bidirectional, directed from the client video renderer to the audio server machine.

The second channel, the *event channel* is for notifications from the audio server to the video client. It will provide information about exceptions on the audio server side or, if enabled, continuous status information. The latter one consists of implementation-specific data, like the current position within a distinct sound in terms of milliseconds or samples. The client implementation will monitor the event channel asynchronously to its normal processing, as events, exceptions and notifications from the audio server can arise at any time. The command channel as well as the event channel are realised as connection-oriented streams.

A third and final channel consists of the *update channel*, which is a connectionless client-to-server data channel that continuously, at the highest possible rate, provides spatial data about the listener and various sound sources to the audio server. The data transferred is mainly a table of positions and orientations that is used on the audio server to calculate the current filter settings as described in section 4.

It is clear that complex auditory data can not be processed over a slow interconnection-bus like a network. The system assumes that all necessary audio data (large sound files in a specific format, e.g. WAV or RAW) for the VR application is present on the audio rendering host. Optionally, the implementation allows a transfer of the sound files during the previously described init phase from the video rendering host to the audio server. While the application is running, only steering information is sent over the network bus in order to synchronize the state of the audio server and the video rendering machine.

The implementation does not need any cumbersome infrastructure to be installed in the environment, as it uses TCP/IP for communication between the sound server and a video client. More complex solutions based on middleware technologies like CORBA can be considered for realising the communication between the two systems as well. But there are some arguments against such a solution. First of all, CORBA is a very general technology that, if used in a naive way, will not meet the real-time requirements that are needed for an application like the one described here. In addition, most middleware platforms do need infrastructure before any actions can take place. This infrastructure is sometimes hard to install and to administer and is not always available.

7. Conclusion

This paper presents a comprehensive, but straightforward approach to combining conventional VR applications with a high quality binaural synthesis approach that allows near head VR acoustics with modest hardware requirements. The interfaces presented are fully implemented in the visual VR toolkit ViSTA and the works of the ITA on the topic of VR audio. It is embeddable into existing ViSTA applications and will be used heavily in upcoming work in order to research on the interaction between the visual and the auditory perceptional human system.

Further topics of research will be on the cost of the communication between the loosely coupled technical systems, especially the reduction of network latency. Another topic will be the usage of the directivity of sound sources, e.g. an avatar that is talking to the user but not facing his direction and the effect of things like these on the overall immersion of the user.

References

[Bau63] BAUER B.: Stereophonic earphones and binaural loudspeakers. *Journal of the AES 9* (1963).

[Beg94] BEGAULT D. R.: *3-D Sound for Virtual Reality and Multimedia*. MA: Academic Press Professional, Cambridge, 1994.

[Bie00] BIERBAUM A.: *VR Juggler: A Virtual Platform for Virtual Reality Application Development*. Master's thesis, Iowa State University, 2000.

[Bla97] BLAUERT J.: *Spatial Hearing, Revised edition.* Cambridge, Massachusetts: The MIT Press, 1997.

© The Eurographics Association 2004.

[CPGB94] CONWAY M., PAUSCH R., GOSSWEILER R., BURNETTE T.: Alice: A rapid prototyping system for building virtual environments. In *Proceedings of ACM CHI'94 Conference on Human Factors in Computing Systems* (April 1994), vol. 2, pp. 295–296.

[FMC99] FUNKHOUSER T. A., MIN P., CARLBOM I.: Real-time acoustic modeling for distributed virtual environments. In *Siggraph 1999, Computer Graphics Proceedings* (Los Angeles, 1999), Rockwood A., (Ed.), Addison Wesley Longma, pp. 365–374.

[Gar97] GARDNER W.: *3-D audio using loudspeakers*. PhD thesis, Massachusetts Institute of Technology, 1997.

[Len03] LENTZ T.: Untersuchungen zum einfluss von reflexionen bei der Übersprechkompensation. In *Fortschritte der Akustik: DAGA 2003 in Aachen, Tagungsband* (2003), pp. 864–865.

[LS02] LENTZ T., SCHMITZ O.: Realisation of an adaptive cross-talk cancellation system for a moving listener. In *21st Audio Engineering Society Conference, St. Petersburg* (2002).

[Møl92] MØLLER H.: Fundamentals of binaural technology. *Applied Acoustics 36* (1992), 171–218.

[NSG02] NAEF M., STAADT O., GROSS M.: Spatialized audio rendering for immersive virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology, Hong Kong, China* (2002), pp. 65 – 72.

[Ope] OPENAL: "http://www.openal.org".

[Sch93] SCHMITZ A.: *Naturgetreue Wiedergabe kopfbezogener Schallaufnahmen über zwei Lautsprecher mit Hilfe eines Übersprechkompensators.* PhD thesis, Institut für Technische Akustik, RWTH Aachen, 1993.

[Sen] SENSE8: Worldtoolkit virtual reality support software.

[SG02] SCHAEFFER B., GOUDESEUNE S.: Syzygy: Native pc cluster vr. In *Proceedings IEEE Virtual Reality* (2002).

[The03] THEILE: Potential wavefield synthesis applications in the multichannel stereophonic world. In *24th AES International Conference on Multichannel Audio* (2003).

[vRKG*00] VAN REIMERSDAHL T., KUHLEN T., GERNDT A., HENRICHS J., BISCHOF C.: Vista: a multimodal, platform-independent vr-toolkit based on wtk, vtk and mpi. In *Fourth International Immersive Projection Technology Workshop (IPT2000), Ames, Iowa* (2000).

© The Eurographics Association 2004.