Surround Digital Audio in Architectural Animation and VR Environment

Cagri Rona, Dr. M. Saleh Uddin
University of Missouri, Columbia U.S.A
www.missouri.edu/~uddinm

To stimulate all five senses through a realistic representation, audio plays a significant role. Architectural representation in digital media primarily falls in the area of visualization. Recent developments in simulation of 3d animation, lighting effect, material options and texture quality demonstrate that technology has come to a satisfactory level for representing realistic environments. But, at the same time, designers may not have paid due attention in regards to simulation of quality audio in architectural representation, particularly in 3D animation.
"Surround Sound" or technically, the Dolby Digital and DTS technology has been used for entertainment purposes in film and movie production for a long time. As realistic materials and lighting increases the value of reality, sounds and sound effects too can add value to represented virtual reality architectural environments.
Surround Sound technology simply delivers different signals from a 360-degree range. The reason beneath this is to break the sense of looking towards a single direction. Ideally, the audio should come from the direction of the visuals on the scene. Synchronization of both visual and Dolby Digital audio can enhance the experiential quality of an audio-visual representation. At the same time, it can break the notion of audio coming from a single direction. This paper suggests and demonstrates the technique of combining surround sound in digital 3D architectural animation for the purpose of enhancing its visual and sensory quality.

Keywords: *Architecture, animations, 3D Surround Audio.*

Introduction

„Graphics alone cannot create a--sense of place--for the user, and do not usually make users feel as if they are present in the virtual environment. Sound is critically important in achieving such results with VR systems. It can be used to enhance immersion, to help orient the user in the virtual environment, and to allow for communication between user and the system and/or other users (Jens Herder, Michael Cohen and William L. Matens, 1997)."

Five basic human senses needed to be engaged to simulate a true experience of reality in animations and V.R. environments. Noteworthy develop-

ments in that effort would include stereographic eye-glasses, digital gloves and other hardware that are currently being used in many areas of virtual simulation. But, for architectural animation, simulation of realistic sound effects has not advanced to a satisfactory level where audio could truly enhance the experiential quality of 3D architectural representation. This paper investigates, implements, and simulates surround digital sound effects to match the moving elements in a case study architectural animation.Basics of Surround-sound

Audio and surround sound (Dolby Digital and DTS encoding)

„Surround Sound“ or technically, the Dolby Digital and DTS technology is being used for entertainment purposes in film and movie production for a long time. The technology is obviously being practiced for increasing the aspect of reality for our pleasure and already being appreciated by the spectators. As realistic materials and lighting increases the value of reality, sounds and sound effects too can add even more to the considered environments. Surround Sound technology delivers different sonic signals from 360 degree range synchronizing the direction of audio with the direction of visuals on a screen or monitor.

Understanding Basics of Surround-sound (Dolby Digital and DTS encoding)

Audio and Surround-Sound (Dolby Digital and DTS encoding)

„Surround Sound“ or technically, the Dolby Digital and DTS technology has been used for entertainment purposes in film and movie production for a long time. The technology is obviously being practiced for increasing the aspect of reality for our pleasure and appreciation by spectators. As realistic materials and lighting increases the value of reality, sounds and sound effects too can add even more to the considered environments. Surround Sound technology delivers different sonic signals from a 360-degree range synchronizing the direction of audio with the direction of visuals on a screen or monitor.

Foley, ambient and linear composite sounds

Although the Audio technology is the same for both „animations“ and „VR environments“, the implementation technology differs. Unlike animation, audio in VR environments is not linear. There must be a varying algorithm due to characteristics of the environment that vary in real time. The soundscape must be dynamic and respond to the changes in the environment by the user input, which creates the considered algorithm. The audio in VR environments can be categorized as Foley and Ambient sounds. Foley sounds are basically the variable sounds which is triggered by the user depending on the action taken such as footsteps, door sounds. In other words, they are the real-time sounds. Ambient Sounds are background sounds (often looping), to create a sense of atmospheric environments such as rain, wind, birds and also background music, which may be considered in an ambient sound category. Linear Composite Sounds are simply the background music. It can be considered as linear because it is, in most cases, the non-stop background music in which the volume may increase or decrease depending on the situation. At the same time, it can also be considered as composite because the background music can be exposed by any of the 5 speakers simultaneously and can be consequently changing the channels for different effects.

Sonic images

Sonic images are simply the virtual images created by sound and its dynamic values (amplitude, distance and direction) through human perception. By the calibration of the visual data and sonic data, the visual information is being supported by the sonic images. The term „sonic image“ is actually an abstract term, defined as the sound and sound

effects which are used in a physical animation environment to emphasize visual data such as materials. The wood or steel sounds created by footsteps will tell the observer that the flooring is either wooden or steel without actually seeing it. The image of the flooring material will occur in the observer's brain by just hearing the sound. This also means that we do not have to show every unnecessary detail in the animations, since the audio will create the „sonic image" of some of the details in the design.

Sound resolution or S.R.I (Sonic rendering index)

S.R.I can be considered as the terminology of creating sonic images. Similar to C.R.I (Color Rendering Index) of a light, determining how we observe a color of a material or the whole environment, Sonic Resolution can be named as S.R.I (Sonic Rendering Index) holding audio information provided in given visual information. As an example, wooden flooring is provided as the visual data, but the audio values are the data, which actually is rendering the given visual information through an animation. In this case, there are a number of values in audio data to be considered: Distance, amplitude, and direction.

Surround sound for architectural animation

Animations and V.R environments

„Unlike sight, the sense of hearing is often neglected in the implementation of a virtual world. Regardless of considerable evidence on its immersive potential, audio is often banished as the poor stepchild of virtual reality (Jarrel Pair)."

Audio is not only a self-player member of an animation or interactive media but also a helper for visual inputs. High quality audio can be used to compensate for low quality visuals. It can be used where the visual instruments do not suffice, or when specific sounds need emphasizing.

3D Surround-Sound technology can be implemented in both animations and VR environments. However, since linear animation uses composite audio, it is simpler to simulate surround audio using both Foley and Ambient sound. This paper implements simulation of surround audio only in an architectural animation.

Possible implementation techniques of the technology in architectural animations

Artificial Dramatization

Implementation of audio into architectural animations is not only a matter of calibration of physical and sonic data, but also a matter of psychological imaging of the brain reacting towards what and how it perceives sound. Human reactions on perceived sonic data in general, are human reactions, which can be mentioned as common assumptions.

The human brain tends to create a virtual scene by the input of differing sound types mentioned above. Since, we are dealing with combination of both visual and sonic data together, we are not expecting the observer to create his own imaginary environment by listening to a 3D audio.

Sense of acoustics and materials

Considering 3D sound in architectural environments for emphasizing and improving the sense of reality, it is important to render the materials with their sonic values (S.R.I or Sonic Rendering Index of a sound). Materials such as wood, metal, glass, stone and

High Sonic Magnitude:	Far Distance (Distance Perception)
Low Sonic Magnitude:	Close Distance (Distance Perception)
Rough Sonic:	Rough Materials and Construction
Smooth Sonic:	Smooth Materials and Construction

Table 1
Types of artificial dramatization

even grass can be rendered with their own sounds to enhance a given virtual environment together with the virtual details such as light, color and material textures.

Virtual sense of natural environment
In a given architectural environment (either an animation or an interactive environment), the existing natural environment can also be rendered with sonic values. The considered natural values may be rain, wind, fountain, etc. Taking rain as an example, considering a rainy site, the sound of rain and its direction can be emphasized as direction of the rain and the sound difference as hearing the rain from outside and inside of a building.
On the other hand, we often use trees in our animations that are not static; often leaves may be shaking due to a wind. The noise of trees caused by wind can be implemented to an animation by changing the direction of the sound source from the speakers parallel to the movement of the tree. The technique of creating such effect is mentioned in the „Case Study“ section of this paper.

Critical issues

Synchronization is a major issue for implementation of Dolby Digital Audio to a video or animation visuals. To achieve satisfactory results, enough time should be spent on synchronization. To avoid undesirable effects, following synchronization types and their application would be important to understand while working on a Dolby Digital Channel Encoder.

Calibration of sound and visuals
In an animation, the visual images and their sonic sources must match their respective positions along the camera movement. The sound source and the audio channels must be synchronized.

Object oriented synchronization
The synchronization between an object (a sound source) and the actual speaker source can be named as „Object Oriented Synchronization“. The objects creating sounds must be synchronized with a speaker or speakers surrounding the observer. The sound created by an object must follow the object through the 5 speakers while it moves on the screen. This must be carefully implemented by following the timeline layer on both the video and audio channels. In this paper, Adobe Premiere software is used as a video editing tool and Steinberg's Nuendo is used as the audio encoder for the case study.

Timeline synchronization
Synchronization between the complete animation and the complete audio can be called „Timeline Synchronization“. It is crucial to synchronize the complete product in terms of the audio and video data. If not, the audio track and the video track will not match. This is not desirable since the whole idea is a complete audio-visual combination and synchronization.

Implementation and synchronization of 3D Dolby Digital audio in a case study example

This sample work will focus on creation of Dolby Digital Surround Audio by integrating a previously created video animation file with appropriate audio files. During the example process, four main steps will be followed as listed in the chart below:

Working on the raw animation and choosing audio files
As a staring point, a simple, previously created/ rendered raw animation (without audio) is used. The visuals in the animation have following key features that needed to be highlighted through appropriate audio effects:

- A tree and its leaves moving along with the wind
- A fountain
- A person walking from the right side of the frame to the left

Step 1:	Working on the Raw Animation and Choosing Audio Files
Step 2:	Taking Notes of Action Sequences (Sequencing)
Step 3:	Creating Audio Files to be Encoded and Dolby Digital Encoding
Step 4:	Combining Audio and Video Files for the Final Product

Table 2
Steps to be followed.

Below is the exploded view of the layers of Visual Elements that needed be emphasized through audio effects.
According to the actions in the animated scene, necessary raw audio files (standard audio files such as wave and/or mp3) are to be prepared or obtained from other sources. Audio file to be used are:

- Wind sound (Ambient sound)
- Sound of leaves swinging along the wind (Ambient sound)
- Fountain sound (Ambient sound)
- Footsteps of a person walking on grass (Foley Sound)

The above audio files will be used in a later step. However, it is important to decide the sequence of order in the early stage to make the following „sequencing“ step easier.

Taking notes of action sequences

Once the raw animation file is ready, it is time to note down the important key frames and their sequences. Importing the video file into a digital video editing application can do this. For this example, Adobe Premiere is used as a digital video editing software. This is where the animation is laid out for marking. Now, the three action sequences needed to prepare for layering with the video sequence. These three actions are: 1) the leaves of the trees moving, 2) the water flow in the fountain, and 3) the person walking across the screen. By following the timeline, it is necessary to take notes of starting and ending time (seconds) of each action.
The continuous actions are recurring sounds, also called as ambient sounds. Ambient sounds do not have specific stopping and starting sequences other than the beginning and ending of the animation. They continuously loop. These ambient audios are the sound of fountain, wind, and leaves. As opposed to ambient sounds or recurring sounds, the third action, the person walking across the screen is a Foley sound. Thus, it has to be laid out precisely. As the figure below shows, it is necessary to take notes when (which second on timeline) the person enters, and exits the screen. This is to determine in which second the footsteps will start and end. Also the transfer of footsteps from one speaker channel (in this example, right speaker) to another (left speaker) will be determined by this step. It is not desirable to hear audios of footsteps before the actual person enters on the screen.

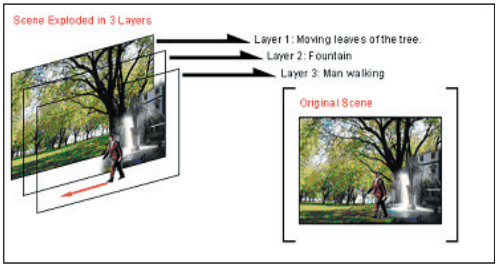

Figure 1
Exploded layers of an animation.

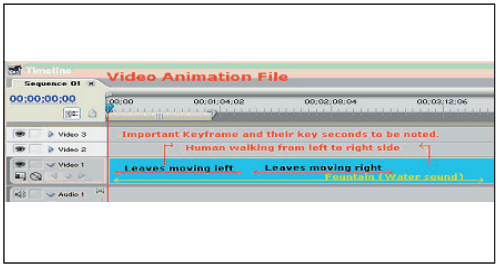

Figure 2
Video timeline in Adobe Premiere.

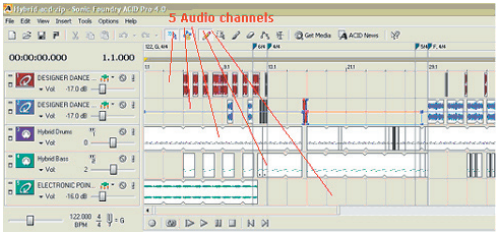

Figure 3
Placing Audio files on Dolby Digital channels (screenshot from Sonic Foundry, Acid4).

Creating audio files for Dolby Digital encoding

At this stage, the video file is left intact, and work needed to be performed in an audio channeling software. This step is to organize and locate the audio files that were prepared earlier (fountain, wind, leaves, and footsteps). This is the step where the previous „sequencing“ step is cross-referenced. The audio channeling software has several channels (usually from 2 to 7 channels) where individual sound tracks can be placed. The sounds are to be placed by following the timeline according to the actions happening in the video animation. For example, the sequence where the person enters the screen, the footstep audio file must be triggered.

The screenshot below illustrates „Sonic Foundry-ACID4.0“ an audio editing and mixing software that can accomplish layering of audio according to timelines (more sophisticated software such as „Steinberg Nuendo“ are available for such works).

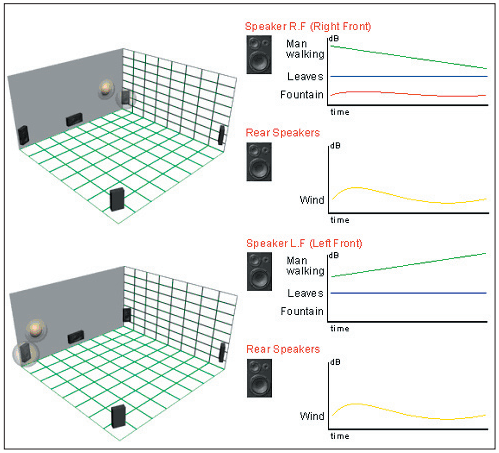

Figure 4
Assigning audio files to channels (speakers).

However, it is not only the sequence of actions that has to be resolved. It is also to be decided from which speaker (channel) the sound will be delivered. This step is rather simple. According to the direction of the sound sources, the ambient (background or non-moving) sounds have to be delivered from the desired front or rear speaker channels. The Foley sounds are the critical ones in the whole process of compilation. In this example, since the person is walking from right to left, the footsteps needed to be panned from right speaker to left. It is to be noted that the panning of sound from one direction (speaker) to another is not simply switching the sound from one speaker to another. More importantly, the sound should shift from one speaker to another by decreasing the sound volume (dB) in one speaker, and at the same time increasing the sound volume (dB) in the other speaker channel. The figure below explains assigned sound channels (speakers) and their volume in dB values.

After organizing the audio files and placing them on the timeline, the final step is to exporting the audio files using specialized software. The nature of this software is to provide (export) 5 different audio files for LF (Left Front Channel), C (center Channel), LFE (Woofer Channel), RF (Right Front Channel), LR (Left Rear Channel), RR (Right Rear Channel).

The process then is completed by combining and compressing those 5 exported audio files and encoding through a 3D Dolby Digital Sound encoder to create a single *.ac3 (Dolby Digital audio) file. This procedure does not involve any synchronizing and/or sequencing. The software interface is also quite simple. The only task to be executed is to locate the audio files in proper speaker channels for the final compression.

Combining audio and video files for the final product (output file)

At this point, we have one video animation file, and one single Dolby Digital Audio file. These two files

Figure 5
Dolby Digital encoding (screenshot from SurCode CD-DTS Plus).

can be combined in any digital video editing software (like Adobe Premiere) and can be exported in „mpeg2“ format to be recorded on either a DVD or CD-R.

Conclusion

„The relationship between auditory and visual perception can draw on synesthesia, mental imagery and creativity.“

Andrew D. Lyons

Surround digital audio technology is not new. But its use in architectural animation is. Traditionally, this technology has remained within the high-budgeted projects in Hollywood studios using professional equipment. This paper illustrates successful implementation of technology for architectural animation in a cost-effective way using a regular home PC.
If this technology satisfies and entertains us through movies, it can easily be implemented in architectural animation to enhance its auditory and visual perceptions.

References

Jens Herder, Michael Cohen and William L. Matens, Sound for Virtual Reality Environments, University of Aizu Forum for University-Industry Cooperation, November 1997, pg1

Jarrel Pair-Virtual Environments Group, Graphics, Visualization and Usability Center, College of Computing, Georgia Institute of Technology, pg3

Andrew D. Lyons, Abstractly Related and Spatially Simultaneous Auditory-Visual Objects, The Sydney Conservatorium of Music, The University of Sydney, Australia, pg10