BERGEN SOUND SERVER
SPATIALIZED SOUND EXTENSION
OVERVIEW

Severin S. Todt
University of Applied Sciences Wedel
Feldstrasse 143, 22880 Wedel, Germany
Email : to@fh-wedel.de
08/29/2002

Abstract

Bergen Sound Server is a sound server capable of replaying sound samples, tones or white noises. It is implemented within a client-server model. The server's activity is dedicated to replaying sounds, managing and processing sound data and to communicate with the client side. A sound's state is controlled by the client side by sending messages to the sound server. Currently only one channel audio reproduction is supported. No positional nor orientational information is taken into account for audio reproduction.
A Spatialized Sound Server Extension is written to enrich the Bergen Sound Server to include positional and orientational information within audio reproduction. It further enables the Bergen Sound Server to use different techniques for producing spatial sound characteristics.
With the integration of spatial sound features a more realistic reproduction of sound sources used in virtual reality applications is possible. Thus virtual reality application developers are enabled to position a sound source, thereby producing the audio characteristics according to the graphical counterpart.

Introduction

The inclusion of spatialized sound in virtual reality application can increase a virtual reality system's immersion as well it helps user's to quickly identify, localize and recognize objects appearing in a virtual reality scene.

A sound system that is able to reproduce spatialized sound characteristics is necessary to allow a feedback as described. To calculate the accurate sound characteristics the sound source's position and orientation as well as the listener's position and orientation have to be settable and determinable. Further information about the hardware set-up used for audio reproduction is taken into account for calculation of the sound characteristics.

In this paper an approach is described to obtain a spatialized sound system within virtual environments. The work is based on the already existing Bergen Sound Server system currently used at the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago (UIC). To set up a spatialized sound server approach the equipment available at EVL is used. Thus audio reproduction techniques are restricted to a loudspeaker set up using either one channel, two channel or four channel techniques.

An introduction into the data model used to manage the data in order to calculate spatialized sound characteristics is given as well as the algorithms to calculate the accurate sound are being introduced and explained.

Spatialized Sound Implementation

The implementation is based on the Bergen Sound Server written by Dave Pape at the EVL in November 2000. Bergen is a very simple, freely redistributable audio server and client library. The Bergen Sound Server is able to replay sound samples, tones and white noises using a one channel audio set-up on a SGI-IRIX or PC-Linux machine. The sound data is managed and processed on the server side of the implemented client-server model. Audio reproduction at the server side is controlled by the client side through adjusting the state of sound objects to be replayed. Client and server communicate over a UDP socket interface.

While the client is holding handles of each sound involved in the current sound application to identify them at the server side as well as the sound's key attributes, the server is holding references to the sound data and further attributes necessary for audio reproduction within a sound database (snerdDB). (See figure 1)

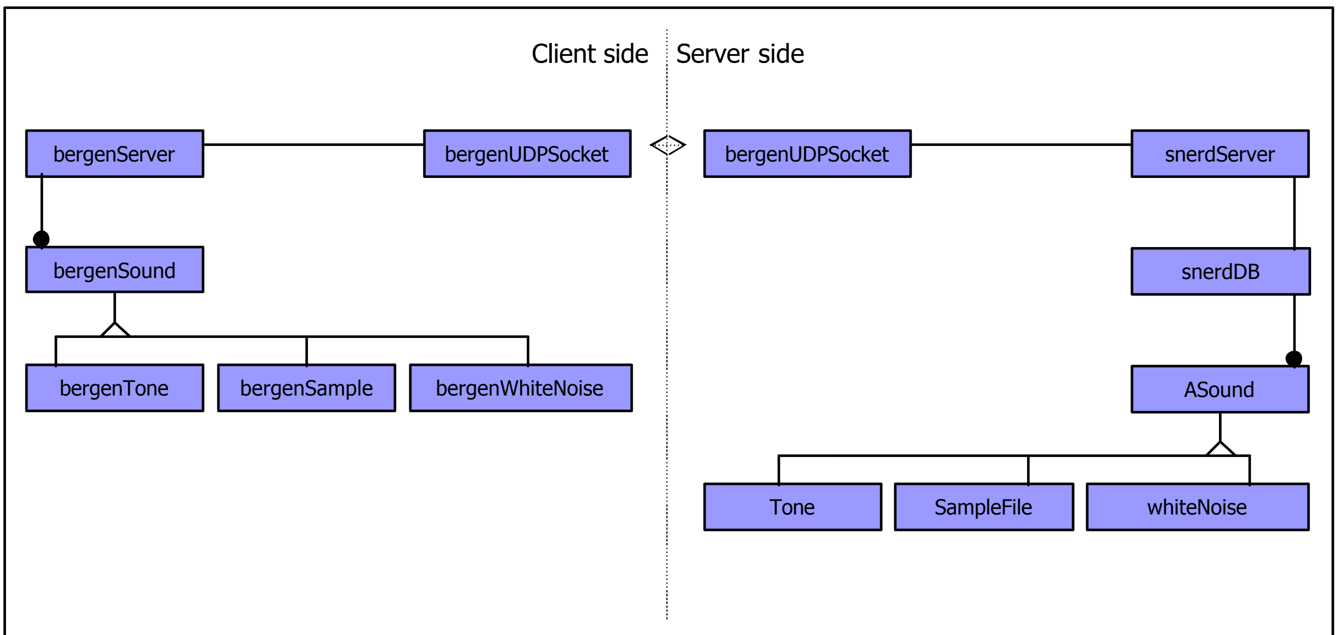

Figure 1 : Data Model Bergen Server

<u>Data Model</u>

For an accurate spatial sound reproduction the listener position and orientation as well as the sound sources' position and orientation must be stored and functionality to query and set these values must be offered.

Since the Snerd Sound Server may reproduce sound for a number of clients, a list of listener representations must be managed, one listener representation for each client. Each client holds a handle to a listener representation together with the most important listener attributes that are determined on the client side. The handle can then be used to identify a certain listener at the server side in order to query or set any value for it there.

For the sounds' positional and orientational information management the sound representations have to be enriched by attributes for position and orientation. Since the Snerd Sound Server is querying each of the Snerd DB's sounds to update their own buffers which are then used for final audio reproduction the sounds' update functionality must be adjusted to fit the current hardware set-up. There must be a different implementation available for each audio reproduction set-up supported by the Snerd Sound Server. At the moment only loudspeaker reproduction using one, two or four channels must be supported, since these are the only set-ups currently available at EVL.

The Snerd Sound Server has to ensure that only Sounds according to the current audio reproduction set-up are instanced. Due to varying approaches with each of the available audio reproduction set-ups there must be a Snerd Sound Server implementation for opening the audio device and writing data to it according to the chosen audio reproduction set-up.

For a spatialized sound implementation the data model is then enriched by a listener representation on the client and server side, sound representations for each reproduction set-up as well as a Snerd Sound Server implementation for each of the supported audio reproduction set-ups. (See figure 2)

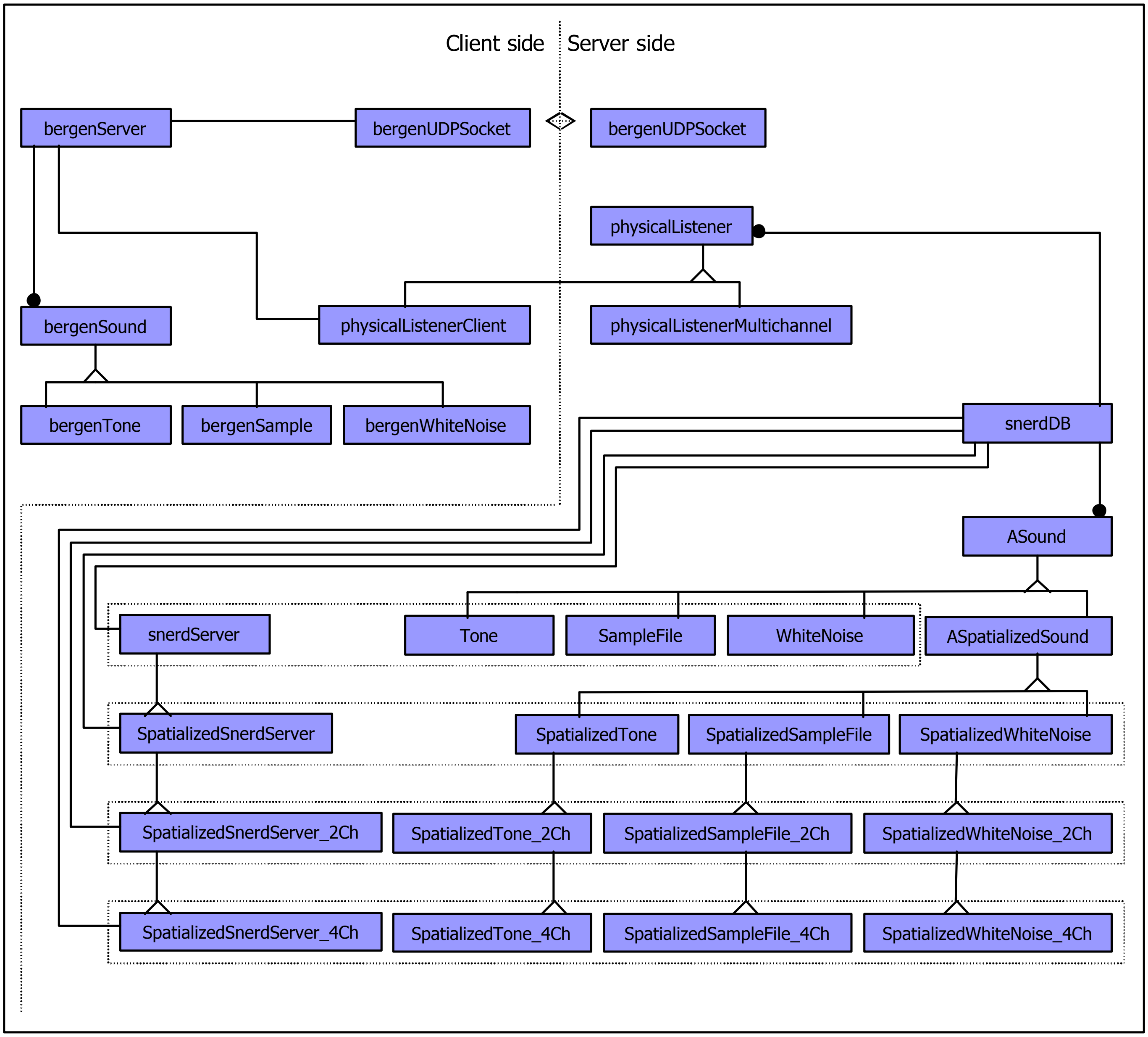

Figure 2: Spatialized Sound Enriched Data Model

Spatialization Techniques

The current Bergen Sound Server Spatial Sound Extension is restricted to the available hardware set-ups at EVL.

Different degrees of spatialization are reachable using one, two or even four channels for audio reproduction.

Spatialization is performed relative to the listener's position. Given an initial listener set-up within the audio reproduction system's hotspot, the listener position can be adjusted dynamically with listener movement.

One Channel Spatialization Set-Up

Using only one channel for audio reproduction, no directional information can be represented, since the application's virtual sound sources always seem to originate from the direction the one speaker is located at.
Beside the directional information, the distance of a virtual sound source can be simulated by adjusting the amplitude with distance from the listener.
The distance attenuation is performed using one of the possible attenuation laws.

The Inverse Square Law was proven to be the best model for distance attenuation. Regarding to the Inverse Square Law the amplitude is attenuated by -6dB with every doubling of distance. Given a reference distance at which a given gain is assumed and a rolloff factor which can be used to strengthen the attenuation the accurate amplitude is then determined by the following logarithmic formula:

Amplitude = GAIN – 20 * log10 (1 + ROLLOFF_FACTOR *
(distance – REFERENCE_DISTANCE)/REFERENCE_DISTANCE)

If a minimal or maximal amplitude is given the maximum of minimum amplitude and the calculated amplitude as well as the minimum of the maximum amplitude and the calculated amplitude is taken for distance attenuation.

A variation of the Inverse Square Law, also known as Clamped Inverse Square Law is taken into account a minimum distance and a maximum distance. Sound source's that a closer to the listener than the given minimum distance are attenuated taken into account the minimum not the actual sound's distance. Attenuation for sound sources further away than maximum distance is performed using the maximum distance instead of the sound source's actual distance.

A third technique for distance attenuation is a linear approach. With linear distance attenuation the amplitude is attenuated from 100% at a given reference distance to 0% at a given falloff distance.

The actual amplitude is then calculated by :

Amplitude = 1 - ((distance – REFERENCE_DISTANCE) / FALLOFF_DISTANCE)

The fourth distance attenuation model, also supported by the Spatialized Sound Extension is distance attenuation by factor.

Sound sources positioned further away than a given reference distance are attenuated with distance by a given falloff factor :

Amplitude = 1 - ((distance- REFERENCE_DISTANCE) * FALLOFF_FACTOR)

The distance attenuation model can be chosen for each sound source individually. Choosing non of the supported distance attenuation techniques results in an ambient sound source heard equally at each position in the application's virtual world.

Two Channel Spatialization Set-Up

Using a two channel audio set-up, directional information can be encoded in addition to the distance information.

Using stereophony techniques a sound source's origin can be simulated anywhere on the line between the two used loudspeakers or headphones. In connection with the distance attenuation, sound sources can positioned anywhere on a plane spanned by the vectors from the listener to the loudspeakers.

To spatialize a virtual sound source the speakers' gain is adjusted according to the speaker angle (a) and the sound source's angle (ß). (See figure 3)

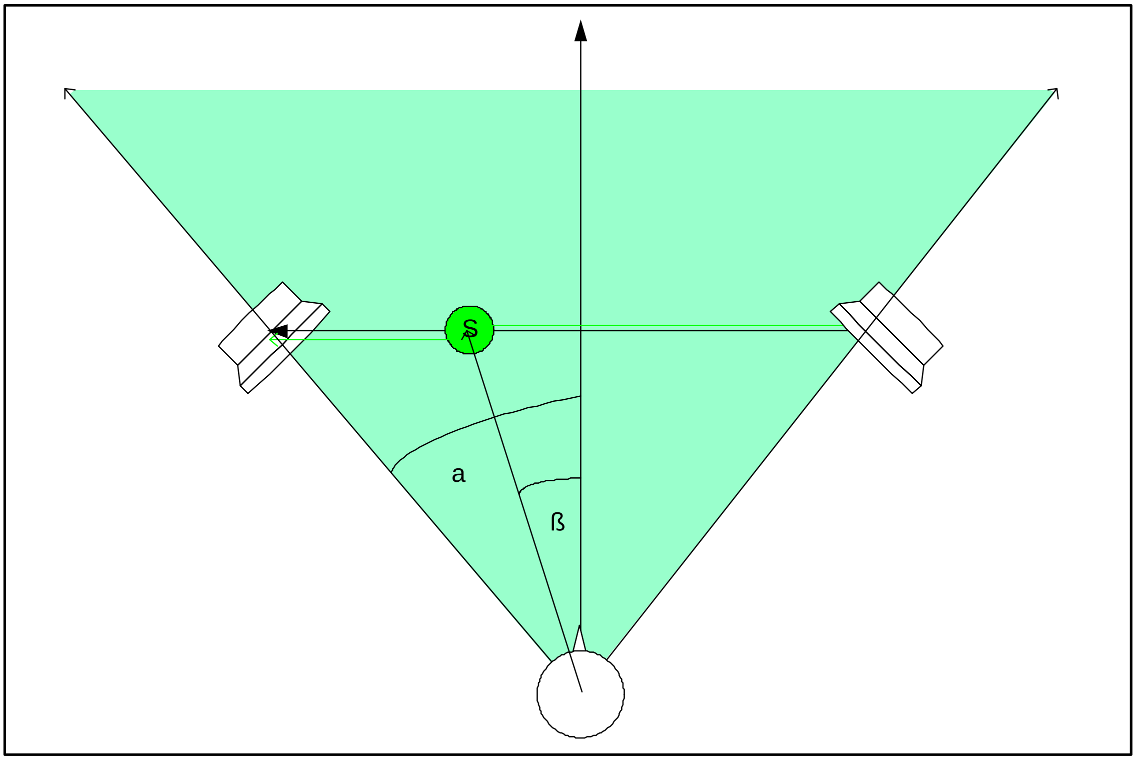

Figure 3 : Stereophony Sound Spatialization

The amplitude adjustment for each channel is performed by the Chowning panning law :

$Amplitude_n$ = sqrt ((a_m - ß) / (a_m - a_n))

$Amplitude_m$ = sqrt ((a_n - ß) / (a_n – a_m))

($Amplitude_n$ amplitude of speaker n, $Amplitude_m$ amplitude of speaker m, a_n = angle to loudspeaker n, a_m = angle to loudspeaker m)

To be capable of positioning a sound source anywhere on the listeners horizontal plane, sound sources with a sound source angle greater than the loudspeaker angle on the same side and less than 90 degrees are spatialized by totally attenuate the opposite loudspeaker while setting the amplitude to 100% for the speaker on the sound source's side.

Sound sources positioned beyond the listener are spatialized by mirroring their position at the X axis.

Techniques using Head Related Transfer Functions are not supported at the moment. Although these techniques promise to result in a precise spatialized sound impression when applied to a headphone set-up it is not an option for an integration in CAVE systems since it would result in the need of headphones to be worn by the application user shielding him from the real world.

Four Channel Spatialization Set-Up

Using four channels in addition to distance attenuation any sound source can be positioned on a plane spanned by the vectors between the listener and the loudspeakers by using stereophony techniques. (See figure 4)

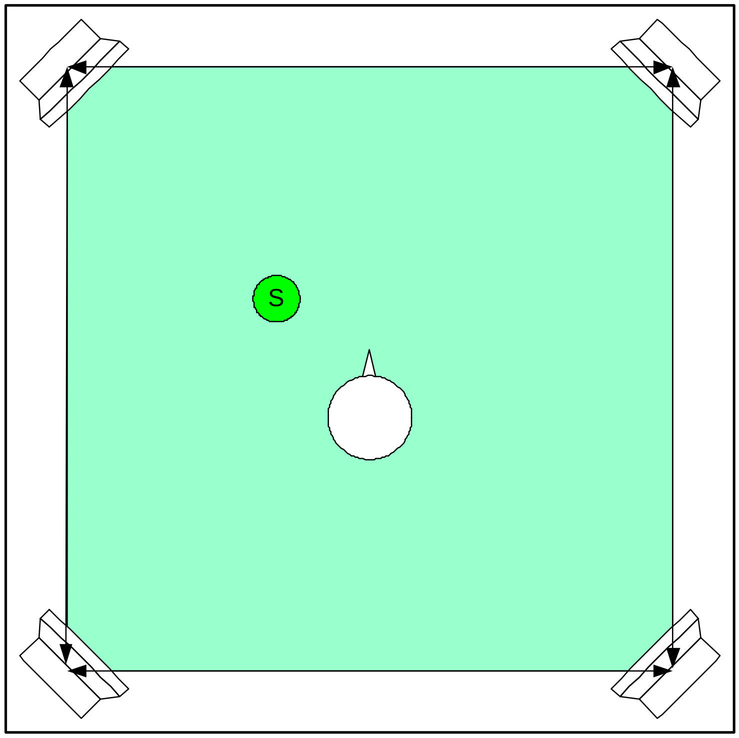

Figure 4 : Stereophony Sound Spatialization for four channels

For four channels the Ambisonic technique is used. Ambisonic is an amplitude panning method in which a sound signal is applied to all loudspeakers placed evenly around the listener with amplitude factors

$Amplitude_i = 1/N * (1+2\cos(a))$

where '$Amplitude_i$' is the gain of 'i':th speaker, N is the number of loudspeakers, and 'a' is the angle between loudspeaker and panning direction.

The sound signal therefore emanates from all loudspeakers which causes spatial artefacts.

Second-order Ambisonics applies the sound with amplitude factors

$Amplitude_i = 1/N * (1+2\cos(a) + 2\cos(2a))$

to a similar loudspeaker system. The sound is still applied to all of the loudspeakers, but the gains have prominently lower absolute values on the opposite side of a panning direction. This creates fewer artefacts. However, to get an optimal result, the loudspeakers should be in a symmetric layout, and increasing the number of them would not enhance the directional quality beyond a certain amount of loudspeakers.

Conclusion and Future Work

With the Bergen Sound Server Spatial Sound Extension a sound server has been developed capable of reproducing spatialized sound characteristics according to a sound source position and the current listener position in real-time.
For audio reproduction either one, two or four channel set-ups can be used, but the Spatial Sound Extension is designed to be open for further audio reproduction techniques like binaural audio reproduction based on HRTF.
Adding further spatialization techniques is one of the future goals within the Spatial Sound Extension.

Beside further output techniques additional input techniques may be implemented. At the moment no audio streams can be used for spatialization. With the integration of a streaming sound live audio data can be spatialized in real-time.
An additional audio capturing facility within the Spatialized Sound Server Extension may then enable the connection of two virtual reality applications using the Spatial Sound Server Extension in order to position the live audio streams originating from the opposite virtual reality application. Using avatars as the graphical representation of the opposite side's user, the audio stream arriving from the opposite side can then be spatialized to the avatar's position in order to easily identify the audio's origin thus simplify the interaction between the applications.

References

Pulkki, Ville; Spatial Sound Generation and Perception by Amplitude Panning Techniques; Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing; Espoo 2001

Durand R. Begault; 3-D Sound for Virtual Reality and Multimedia; Ames Research Center, Moffett Field, California; April 2000

Durand R. Begault; Preferred Sound Intensity Increase for Sensation of Half Distance; Ames Research Center, Moffett Field, California; April 2000