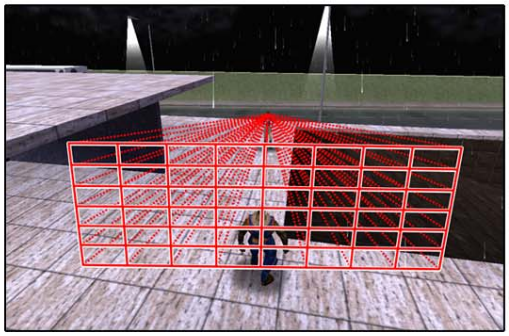
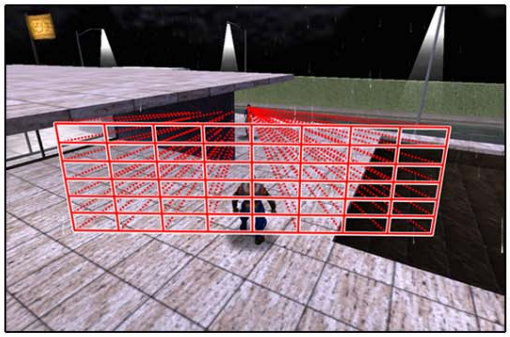
The Matrix: Integrating Audio, Visual and 3D Structure in a Real-Time Virtual World

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

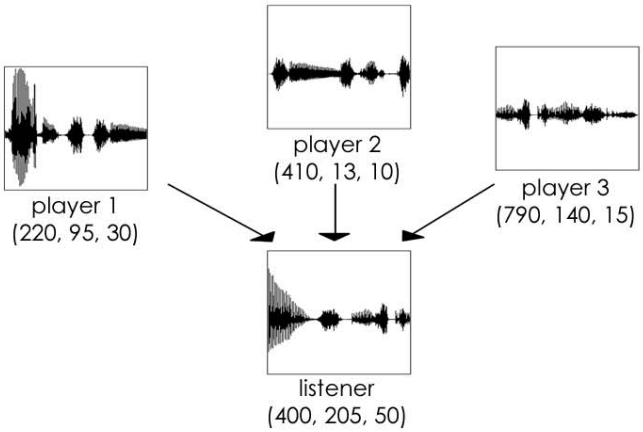




Overview

In the current work on multiplayer virtual reality, the research has focused largely on the visual aspect combined with text. Audio is often neglected and when present typically ignores the structure of the virtual world. In this work, we present a new system which integrates the audio, visual, and 3D structure of the virtual world. Specifically, our novel contribution is the creation of a system which models the effect of the 3D world structure upon the audio and visual aspects in a natural and intuitive manner for a massive multiplayer world: players in the virtual world can now talk with each other as in real life.

This massive multiplayer system with integrated realistic audio, visual and 3D structure is the first of its kind.

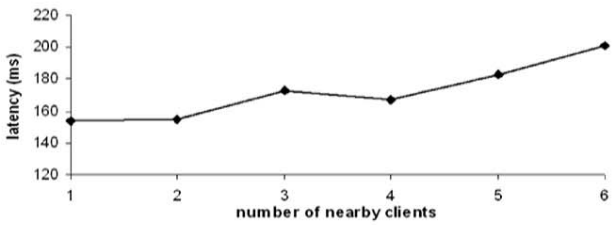
Mixing audio packets by applying distance, angle and structural attenuation

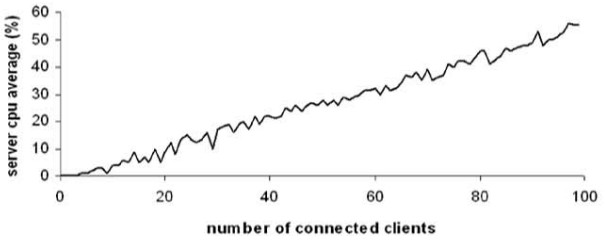


Design Goals

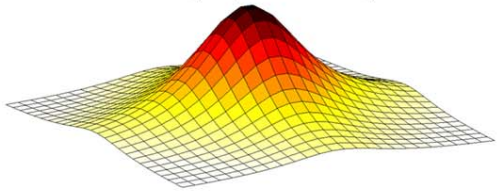
- Allow a large number of clients (>100) to connect.
- Require only moderate bandwidth (~128kbps).
- Work from behind routers and firewalls.
- Maximize portability (any virtual world can be used).

Experiments





Structural attenuation featuring a grid weighted based on Cauchy's probability distribution



Impression of the weighted Cauchy-based grid. Its values are calculated by filling in the 'radius' from the center grid point into an adapted version of Cauchy's probability density function:

$$g(v, w, \gamma) = c(r(v, w), \gamma) + \delta$$

where v and w indicate the axes along which the grid is oriented, with the origin at the center point of the grid, γ is the scale parameter (γ > 0), which specifies the half-width at half-maximum, and

$$r(v,w) = \begin{cases} |v| & \text{if } |v| \geq |w| \\ |w| & \text{otherwise} \end{cases} \qquad c(r,\gamma) = \begin{cases} \frac{1}{\pi\gamma} & \text{if } r = 0 \\ \frac{1}{4\pi r}\left(\frac{\gamma}{r^2+\gamma^2}\right) & \text{otherwise} \end{cases}$$

$$\delta = \frac{1 - \sum_{v=-\lfloor N/2 \rfloor}^{\lfloor N/2 \rfloor} \sum_{w=-\lfloor N/2 \rfloor}^{\lfloor N/2 \rfloor} c(r(v,w),\gamma)}{N^2}$$

The grid is placed with its center at the listener's location, pointing at the origin of the sound. The structural attenuation factor is determined by only including the weights of those grid points that can 'see' the speaker.

Conclusions

- Novel method for high-performance real-time estimation of interlinked audio/visual in immersive virtual environments.
- The structural audio technique results in smooth sound transitions when moving around objects and corners while talking to other players.
- Players can have conversations with many people at the same time, as the audio correctly appears to originate from the visual location of the players that are talking, with the correct volume level.
- Players are able to localize any sound source and direct visual attention to where the sound is coming from.

Universiteit Leiden
The Netherlands

liacs media lab
LEIDEN UNIVERSITY