A Hybrid Sound Model for 3D Audio Games with Real Walking

Iana Podkosova
TU Wien
Institute of Software Technology and Interactive Systems
yana.podkosova@tuwien.ac.at

Michael Urbanek
TU Wien
Institute for Technology Assessment & Design
michael.urbanek@tuwien.ac.at

Hannes Kaufmann
TU Wien
Institute of Software Technology and Interactive Systems
hannes.kaufmann@tuwien.ac.at

ABSTRACT

Spatialized audio is the only output that players receive in audio games. In order to provide a realistic view of the environment, it has to be of superior quality in terms of immersion and realism. Complex sound models can be used to generate realistic sound effects, including reflections and reverb. An implementation of a hybrid sound model based on the ODEON approach is introduced and adapted for real-time sound calculations. This model is evaluated and compared to a baseline model usually used in audio games in a user study in a virtual reality environment. The results show that the implemented hybrid model allows players to adjust to the game faster and provides them more support in avoiding virtual obstacles in simple room geometries than the baseline model.

CCS Concepts

•Human-centered computing → Sound-based input / output; •Computing methodologies → Virtual reality;

Keywords

Virtual reality, audio games

1. INTRODUCTION

Audio games are a class of computer games that have no visual output. They are designed for people who will not or cannot rely on visuals, e.g. visually impaired people. Sound is the main output in audio games. It has to be of high quality to provide an immersive and realistic experience. However, existing audio games only use simple sound propagation models [8, 11, 10, 5]. Such models use spatialized sound but lack more complex sound phenomena like reverberations. Most audio games are played in front of a computer screen, however, there are examples of audio games where users are able to move freely and interact with the environment [9]. The focus of the presented research is on this latter type of audio games.

In this paper, a hybrid sound propagation model is implemented in a game engine and evaluated in a 3D audio game. Users are able to walk in a virtual environment (VE) that is invisible to them, while their absolute positions are tracked. Depending on the user location, calculations for sound spatialization are performed in real-time. In addition to 3D sound, our implemented hybrid sound model allows simulation of hall and reverberation effects in real-time, depending on the VE structure and user position. The hybrid sound model is compared to a baseline model in a user study. Our motivation and hypothesis for this work is that a complex real-time hybrid model enhances task performance in 3D audio games when compared to a basic model.

2. RELATED WORK

The work presented in [8] was the first audio game in a 3D VE called AudioDoom. 3D sound was pre-calculated due to computational limits, complex sound effects were absent. The game was played on a desktop. The GRAB environment [11] combined haptics with auditive feedback. Taken footsteps produced echoes, which were however not geometry dependent. Terraformers [10], a commercial hybrid audio game, that can be played with visuals or without, provides the players with spatial 3D sound but does not use any complex sound models. In [5], blind users learned real world geometry by playing a desktop audio game. The authors showed that the users were able to successfully navigate in the real room remembering the VE after playing the game. Another audio game, called Demor provided the first outdoor game prototype which allowed real walking [9] by the use of a tracking module. However, the game did not use any complex sound model.

Complex sound models are often used in the domain of room acoustics simulation [2]. Wave-based approaches [3] calculate the output by modifying frequencies. Such approaches have high accuracy but the wave equation can only be solved for simple geometries. Statistical approaches [4] can be used to rather calculate noise levels than auralization. Therefore, they are not suitable for audio games. The image source method [1] is a ray-based approach which achieves good results in room acoustics. However, the calculation time in this method depends on the reflections' order and the amount of planes in the environment. The hybrid

CASA 2016, May 23-25, 2016, ,

© 2016 ACM. ISBN 978-1-4503-4745-7/16/05...$15.00

DOI: http://dx.doi.org/10.1145/2915926.2915948

sound model used in ODEON [6, 7] extends the image source method by introducing secondary sound sources. In this paper, the hybrid model is adapted to perform in real-time.

3. METHODS

3.1 Baseline Model

A baseline sound propagation model which is normally used in audio games was implemented in order to compare it to our implementation of the hybrid model. In the baseline model, the sound is spatialized. The larger the distance between a sound source and a listener is, the more the sound is attenuated. Spatialization also makes it possible to distinguish the direction the sound comes from. A low-pass filter is used on the original sound source if there is an obstacle in the direct path between the sound source and the listener. This simulates the effect of a sound source being behind a wall by filtering out high frequencies.

The audio game prototype is implemented in the Unity 3D game engine. Sound is calculated in the Audiokinetics' Wwise middleware solution, with the use of AstoundSound plugin for spatialization. Positions of sounds sources and the listener (a player) as well as distances between them are calculated in Unity3D and passed to the AstoundSound plugin for 3D sound calculations. The calculation takes between 1 and 5 ms.

3.2 Hybrid Model

Our implemented hybrid sound model adapts the ODEON approach for real-time use. The image source method is combined with an adaptation of the secondary sources approach, shown in Figure 1.

In the image source method, the actual sound source is recursively mirrored at every wall producing so-called image sources. The recursion is stopped after it has reached the maximum allowed number of iterations, which we refer to as the "maximum image source order". An audibility test is performed for every image source. A ray is shot from the receiver position towards an image source. If the image source has been produced by mirroring at the same wall that the ray intersects first, this image source is considered valid, and invalid otherwise. Valid image sources will emit sound and contribute to early reflections.

Secondary sound sources are used to account for late reflections and reverberation. In the proposed real-time implementation, rays originating from the sound source are shot in all directions (they are called secondary rays) and secondary sources are placed at the positions where they intersect the walls. Secondary rays get reflected by the walls following Snell's law. Their direction is changed by an added random vector to account for surface roughness. The next secondary source is then placed at the intersection with the next wall. This reflection process is terminated when a secondary ray length limit is reached, producing a set of secondary sources. In the ODEON model, the generation of secondary sources starts at the positions where the rays shot from the receiver to the valid image sources of maximum order intersect the walls and stops when the desired accuracy is achieved. In that case, the amount of calculated secondary sources cannot be spatialized at real-time framerates.

Parameters influencing the performance and the quality of produced sound are maximum image source order, the amount of rays and maximum ray length used for secondary

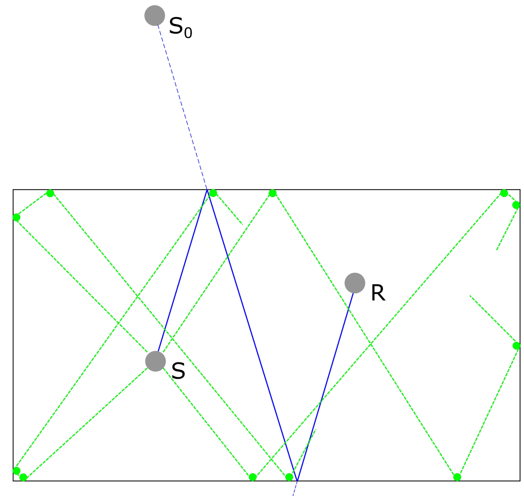

Figure 1: Calculation of the image (S_0) and secondary (green dots) sources for the hybrid model.

sound source calculation. Different values for each parameter were tested. The maximum image source order is set to three as suggested in [7]. The maximum amount of secondary rays used is 25, with the maximum ray length of 60 meters. This results in the simultaneous use of about 500 sound sources, including image and secondary sources, in our typical large room model. With the used software configuration, this number proves to be the maximum amount of sound sources that can be spatialized at the same time without considerably reducing the performance. The calculation of the whole model lasts about 160 ms, most of this being spent on the calculation of image and secondary sources that depend on the position of the primary sound source. This way, the recalculation is beyond real-time capabilities of the used setup if the primary source is moving. With stationary sound sources, the prototype runs at 65 fps.

4. EVALUATION

We conducted a user study to evaluate whether our implemented hybrid model can reduce the *difficulty* of tasks and contribute to the sense of *realism* and *immersion* in an audio VE.

4.1 Game Prototype & Hardware Setup

The game was tested in a 30x7m tracked open space. The participants wore an Oculus Rift HMD (DK2) and 7.1 surround sound headphones (Logitech G35). Players' positions were tracked with the use of a wide area optical tracking system. The prototype ran on a laptop with Intel Core i7 processor and Nvidia GeForce GTX 980M graphics card carried by participants in a backpack.

Before the start of the audio game, participants could walk in an introductory scene with visuals around a sound source and experience the sound from different positions. Walls were placed next to the sound source. After the participants had accommodated to the VE, they took the marked starting position in the VE and the experiment started.

In the main experiment, no visuals were shown. In the beginning, each participant was instructed to walk towards a virtual ringing phone, where they received further instruc-

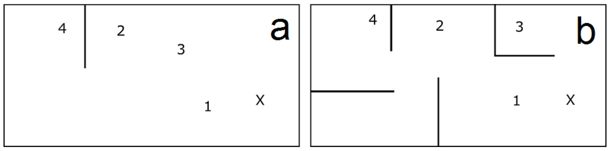

Figure 2: The simple (a) and complex (b) virtual rooms used in the game. The numbers indicate the sound source positions, (1) for the phone, (2) the 1st bomb, (3) the 2nd bomb and (4) for the 3rd. X is the player starting position.

tions. The task was to diffuse virtual bombs by walking into them in a virtual invisible room. Each bomb started ticking after the participant reached the location of the previous one. The game was successfully finished after the participant had found all three bombs.

Each participant played the game in two virtual rooms. The rooms are shown in Figure 2. The room with one wall is referred to as the *simple room* and the room with five walls as the *complex room*. Two rooms of different complexity were used to investigate the possible influence of geometry complexity on the sound models applicability.

4.2 Participants and Procedure

34 participants (20 male, 14 female) aged between 19 and 48 (median 28.5) took part in the study. The participants were split into two groups. The first group had the baseline sound model assigned, the second group the hybrid model, with 17 participants in each group. Game completion time and the number of times when a participant walked through a virtual wall (hit-wall rate) were used as objective measurements of the task difficulty. The hit-wall rate shows players' ability to recognize the presence of virtual obstacles.

The participants filled out a pre-test questionnaire that included questions about gender, age, experience of virtual reality and audio games. After the completion of the game in each room, the participants were asked to draw the virtual room twice. First, they had to draw the room and the sound sources in it on a map of the real environment. After that, they were given a plan of the room with sound sources marked on it but without inner virtual walls. The task was to assign the order of appearance in the game to the sound sources and draw the inner walls. In the end of the experiment, the participants were asked to fill out a post-test questionnaire that included questions aiming to assign participants' degree of immersion, adjustment speed, perceived realism of the sound and difficulty of finding sound sources. A short discussion took place in the end of the experiment.

5. RESULTS & DISCUSSION

5.1 Difficulty

The participants with the assigned hybrid model completed the game faster in the simple and the complex room (simple room baseline: mean completion time (MCT) = 202.6 sec., hybrid: MCT = 180.0 sec.; complex room baseline: MCT = 204.5 sec., hybrid model: MCT = 178.3 sec.) However, this difference has not been found to be statistically significant.

The participants were asked to rate how quickly they adjusted to the VE experience, on the scale from 1 (very slowly) to 7 (very quickly). According to the resulting scores, the participants adjusted to the hybrid sound model quicker than to the baseline model (simple room baseline: mean score (M) = 4.97, hybrid: M = 5.94, complex room baseline: M = 5.77, hybrid: M = 6.41). The difference is statistically significant ($p = 0.03$ in the T-test, $\alpha = 0.05$) for the simple room. The null hypothesis is valid with the probability close to marginal for the complex room ($p = 0.08$ in the T-test, $\alpha = 0.05$). Higher subjective adjustment scores are in accordance with shorter game completion times for the hybrid model.

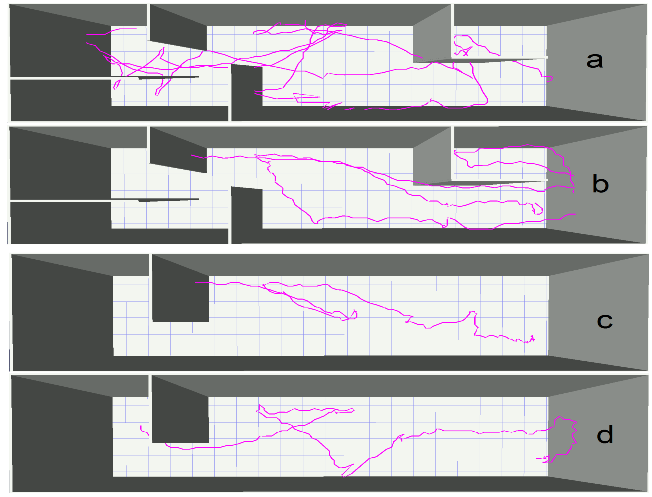

Figure 3: Example paths of participants in the complex (a,b) and simple (c,d) rooms. Paths on a and c correspond to a session using the baseline sound model, c and d - to a session using the hybrid model. The hit-wall rate is higher on a and c than on b and d. Paths on a and c and on b and d are walked by same participants.

The mean hit-wall rate in the baseline model was 1.06 and only 0.29 in the hybrid model in the simple room. This difference is statistically significant ($p = 0.00$ in Kruskal-Wallis test, $\alpha = 0.05$). In fact, 13 of out 17 participants found all sound sources in the simple room with the hybrid model without hitting a wall, while only one out of 17 participants with the assigned baseline model did not hit the wall. The examples of participants' paths can be seen in Figure 3, each representing either the baseline or the hybrid model used in the simple and complex room.

None of the participants were able to reconstruct the virtual room correctly without given borders. Only four participants reconstructed and ordered the sound sources of the simple room correctly. Eight could reconstruct parts of the complex room (none were able to recreate it entirely), while only six assigned the order of the sound sources correctly. Two of them assigned the order of the sound sources correctly in both rooms. The remaining 28 participants could not reconstruct the order in which they discovered the sound sources.

The results show that the reconstruction of a VE which is only audible but not visible is generally a difficult task. During the brief talks after the session, several participants reported that they had experienced a sudden change of loudness when they went through a wall but were not able to identify it as a wall during the test. This shows that the participants were not familiar with sound behavior and did not think about the possible obstacles in the VE. However, the hybrid model provides more precise guidance to the sound source in case of a simple VE with obstacles (simple room).

5.2 Realism and Immersion

Subjective scores for sound realism were high for both models (simple room baseline: M = 5.41, hybrid: M = 5.06; complex room baseline: M = 5.41, hybrid: M = 4.68; on the scale from 1 to 7, 1 is "not realistic at all" and 7 "very realistic"). However, the scores do not differ between the two sound models.

The participants were asked whether they had experienced sound reflections. 59% of the users who had the baseline model (without reflections) assigned stated that they had heard reflections. 24% had not heard any reflections, 17% were unsure. Corresponding numbers for the hybrid model (with reflections) are 35%, 29% and 36%. The confusion about sound reflections might be attributed to the fact that users are not trained in listening to specific sound effects. The simulated reflection might also be not perceived as a real one. A study with within-subject design might show the difference in the perceived sound realism between models.

Finally, the participants were asked about how aware they were of events occurring in the real world during the test and how involved they were in the VE experience. We introduced these questions to assess the participants' immersion. The answers indicate low awareness of real events (simple room baseline: M = 2.89, hybrid: M = 2.29; complex room baseline: M = 2.94, hybrid: M = 2.18; on the scale from 1 to 7, 1 is the least and 7 the highest degree of awareness) and high involvement (simple room baseline: M = 5.94, hybrid: M = 5.47; complex room baseline: M = 5.94, hybrid: M = 6.06; on the scale from 1 to 7, 1 is the least and 7 the highest degree of involvement). However, the mean scores are not different for both sound models. The tested prototype provides a highly immersive audio game experience independent of the used sound model.

6. CONCLUSION AND FUTURE WORK

A real-time implementation of a hybrid sound model that calculates VE geometry-dependent reflections and reverberation is presented in this paper. The implemented hybrid approach is compared to the baseline sound propagation model typically used in audio games. This comparison is conducted in a user study where the participants had to play a 3D audio game without visuals while walking in a large space. The results indicate that both tested models provide a highly immersive game experience, while the proposed hybrid approach provides players with more information about the VE. In the future we intend to intensify our work on the implementation of GPU-based advanced real-time sound models to compare the results of the implemented hybrid model to more accurate sound models.

7. ACKNOWLEDGMENTS

The authors would like to thank Jim and Julien Rüggeberg from Illusion Walk KG for funding the development of the tracking setup used in this work, as well as Wwise for providing a non-commercial license for the research.

8. REFERENCES

[1] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979.

[2] M. Cairoli, E. Moretti, and A. Gade. New opera house in astana: A recent opportunity to use a room acoustic scale model. *The Journal of the Acoustical Society of America*, 133(5):3581–3581, 2013.

[3] J. Jieyi, H. Kunpeng, and Z. Yuezhe. Calculating room acoustic parameters by finite element method. In *Consumer Electronics,Communications and Networks*, pages 4838–4841, 2011.

[4] R. H. Lyon. *Theory and application of statistical energy analysis*. Elsevier, 2014.

[5] L. B. Merabet, E. C. Connors, M. A. Halko, and J. Sánchez. Teaching the blind to find their way by playing video games. *PloS one*, 7(9):e44958, 2012.

[6] G. Naylor. Treatment of early and late reflections in a hybrid computer model for room acoustics. In *124th ASA Meeting*, 1992.

[7] J. H. Rindel. The use of computer modeling in room acoustics. *Journal of Vibroengineering*, 3(4):41–72, 2000.

[8] J. Sánchez and M. Lumbreras. Virtual environment interaction through 3d audio by blind children. *CyberPsychology & Behavior*, 2(2):101–111, 1999.

[9] E. Velleman, R. van Tol, S. Huiberts, and H. Verwey. *3d shooting games, multimodal games, sound games and more working examples of the future of games for the blind*. Springer, 2004.

[10] T. Westin. Game accessibility case study: Terraformers–a real-time 3d graphic game. In *Proceedings of ICDVRAT*, pages 95–100, 2004.

[11] J. Wood, M. Magennis, E. F. C. Arias, T. Gutierrez, H. Graupp, and M. Bergamasco. The design and evaluation of a computer game for the blind in the grab haptic audio virtual environment. *Proceedings of Eurohpatics*, 2003.