A Phong-based Concept for 3D-Audio Generation

Julia Fröhlich, Ipke Wachsmuth

Bielefeld University - Articial Intelligence Group
jfroehli@techfak.uni-bielefeld.de
ipke@techfak.uni-bielefeld.de

Abstract. Intelligent virtual objects gain more and more significance in the development of virtual worlds. Although this concept has high potential in generating all kinds of multimodal output, so far it is mostly used to enrich graphical properties. This paper proposes a framework, in which objects, enriched with information about their sound properties, are being processed to generate virtual sound sources. To create a sufficient surround sound experience not only single sounds but also environmental properties have to be considered. We introduce a concept, transferring features from the Phong lighting model to sound rendering.

Keywords: Artificial Intelligence, Virtual Reality, Intelligent Virtual Environments, Multimodal Information Presentation

1 Introduction

In order to improve user experiences and immersion within virtual environments auditory experience has long claimed to be of notable importance [1]. Still today, current virtual reality projects have a strong focus on realistic graphics rendering and user experience, but acoustics is rarely considered.

One approach to store further information in virtual worlds is to semantically enrich virtual objects. This concept has proven to be good and efficient to create smart objects [2]. But until now this has mostly been used to store additional knowledge about graphical representation. As an example intelligent objects were used to enable smart connections and parametric modifications [3].

In order to generate realistic spatial sound, many factors have to be considered. These include the position in the virtual space, distance to the user as well as an appropriate soundfile [4]. Defining all these factors for every virtual object is a complex and time-consuming task. To achieve a realistic auditory experience it is not sufficient to only generate multiple independent sound sources, yet it is necessary to create a virtual world where objects that interact with each other acoustically influence the environment.

2 Using Intelligent Virtual Objects for Audio Generation

Our framework enables the assignment of semantic information with regard to audio properties to virtual objects using so called metadata. Figure 1 shows the semantic enrichment of such objects by assigning descriptive values. The metadata are read in by an information processing step and are then compared to a database, which contains many sound files, that are semantically annotated. The idea behind this is to create objects that 'know' how they have to sound. This knowledge is not stored in a separate knowledge base, but is embedded directly inside the object itself.

If a database entry is found, which matches a certain audio file, a sound node is created inside the scenegraph with the corresponding object as its parent node. By this means, the prerequisites for the creation of a spatial auditory experience are complied with. The position and direction of the sound node in relation to the user can be calculated directly through the traversal of the scenegraph.

So far these steps only generate multiple independent sound nodes. To improve the auditory experience, a method to combine these sound nodes is needed, in order to create an acoustic 'atmosphere'.

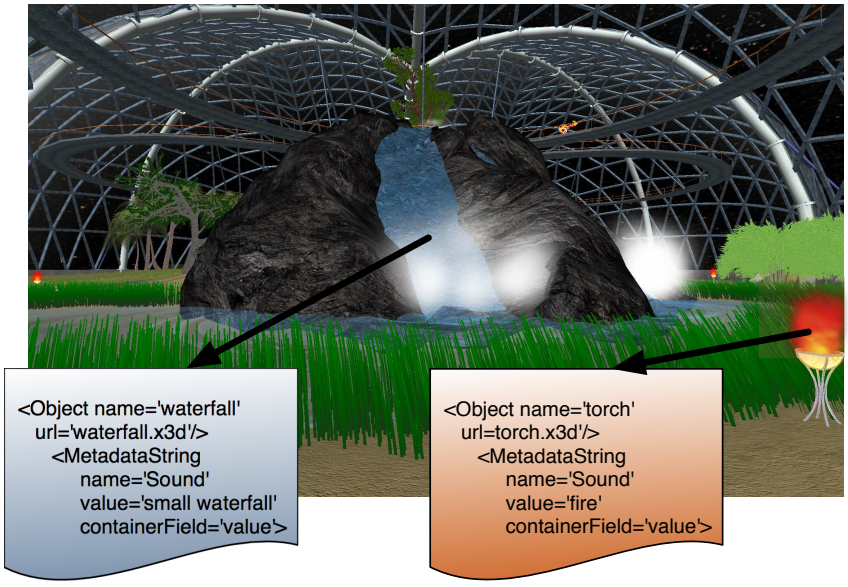

Fig. 1: Semantically enriched virtual world on the example of a waterfall and a torch.

3 Phong-based Sound Rendering

Present realistic, raytracing based methods, do not meet requirements like real-time calculation for virtual reality applications [5]. Therefore we propose a concept based on the Phong lighting model, transferring ideas from lighting to sound generation. The Phong reflection model has proven to be a good approximation to realistic lighting. Although not photorealistic, it provides a reasonable alternative while not using too much resources with respect to CPU/GPU power.

A similar approach for sound rendering in virtual environments is proposed here. Since realistic audio rendering is also computationally costly, a good approximation is needed. The division into three components: ambient, diffuse and specular as in the Phong model is used for 3D-Audio generation within the following concept. Although sound distributes in space in a different way than light, ambient reflection seems promising to transfer ideas from lighting to sound.

Ambient Sound. Mapping the ambient component from graphical lighting to auditory rendering is fairly straight forward. Similar to lighting, ambient sound (I_A) also represents a base level of output which is more or less constant over a larger region of a scene. Since ambient sound is already widely used by audio engineers (e.g. in the movie industry) ambient sounds are widely available. Adding ambient sound nodes to the scenegraph is done by defining major group nodes which describe self-contained areas of the virtual world. As long as the user is within the defined area, ambient sound will be played without direction and always at the same volume. Only one ambient sound can be played at once, therefore if the user is in a subarea of the world all not applying sounds are faded out.

In addition, this concept allows for the definition of *environmental properties* (k) which influence the audio rendering, to fit the environment, such as an outside scenario, a cave or a concert hall. These properties are stored at the same nodes within the scenegraph, and therefore influence all other nodes in terms of reverb characteristics. In this manner appropriate equalizers can be chosen to affect all sounds belonging to this specific zone. This could add a lot of reverb in a cave.

Diffuse and Specular Sound. The division between diffuse and specular sound is not that obvious. We propose a user-centered design with distinction by user focus. By default every object within hearing range is giving diffuse sound (I_D). These sounds are created by our framework described above.

Separation of diffuse and specular sounds (I_S) is effected solely by user focus. To determine the object in focus different methods can be used. As a first approach, rough knowledge can be gathered only looking at the viewing direction (e.g. using head tracking). Surely if the user selects an object by using e.g. data gloves this is the object in focus. Better results are gathered by using eye tracking devices. If virtual reality applications are combined with an eye tracker many information about the user focus can be gathered and used.

The division between diffuse and specular sound is based on research proving the ability of auditory focus and ignoring background noises known as the cocktail party effect [6]. Humans are capable of focusing on one sound and ignoring other sounds around as well as reverberation. The auditory sense achieves a noise suppression between 9 and 15 dB, meaning the sound in focus is perceived two to three times louder than the surrounding background noises.

Summarizing these ideas, we propose formula (1) for sound rendering:

$$I = I_A - x + k \cdot (I_D - x + f \cdot I_S) \tag{1}$$

- I is the intensity of the sound
- I_A is the ambient part, I_D is the diffuse part, I_S is the specular part
- k is an environmental property
- f is 1 if an object in user focus is identified, 0 else
- x is 0 if f is 0, else between 9 and 15 dB

4 Conclusion

This paper proposes a concept for 3D-Audio generation, which utilizes some basic ideas of the Phong lighting model to approximate realistic auditory experience. Additionally we introduce environmental properties which allow for the modeling of different acoustic scenarios. While the Phong lighting model inspired this work, i.e. to strive for a decomposition method for sound the terms most adequate for 'diffuse and specular sound' need further discussion.

Our ultimate goal is to create virtual worlds in which the user can experience the environment he resides in, maybe even in the absence of graphical output. Smart graphics have a high potential in enriching virtual worlds not only with visual content, but furthermore generating multimodal output. In the real world, humans can easily distinguish between different surrounding scenarios, solely by acoustic clues. For example it is easy to tell if one is inside or outside a building. Thus a step towards a realistic environmental experience may be achieved.

5 Acknowledgments

This paper is a preprint version of an article published by Springer-Verlag. The original publication is available at http://link.springer.com/chapter/10.1007/978-3-642-22571-0_22.

References

1. J. Bates, "Virtual reality, art and entertainment," *Presence*, vol. 1, no. 1, pp. 133–138, 1992.
2. M. Luck and R. Aylett, "Applying artificial intelligence to virtual reality: Intelligent virtual environments," *Applied Artificial Intelligence*, vol. 14, no. 1, pp. 3–32, 2000.

3. M. E. Latoschik, P. Biermann, and I. Wachsmuth, "Knowledge in the loop: Semantics representation for multimodal simulative environments," in *Proceedings of the 5th International Symposium on Smart Graphics 2005*, pp. 25–39, 2005.
4. D. R. Begault, *3D Sound for Virtual Reality and Multimedia*. Academic Press, 1994.
5. M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer, 1st edition. ed., 2010.
6. E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.