Enhancing audio signal processing chains in virtual environments

Background and Motivation

In multisensory systems such as 3D applications and Virtual Reality, spatial cues describing the environment surrounding the user are carried through sound. Human hearing uses auditory stimuli to perceive space and identify entities that cannot be detected by other senses. Hence, in realistic representations of virtual environments (VEs), rendering spatial hearing enhances interaction and immersion. Spatial information of environments can be conveyed to listeners through real-time convolution of impulse responses in the signal processing chain (Vorländer et al, 2014). Siltanen (2005), in addition, proposes an approach that, based on reducing the geometry of the environment, allows real-time acoustic modelling. In virtual reality, presence is also induced by auditory information through spatialised audio (Larsson *et al* 2010, Stanney *et al,* 1995). A study is therefore proposed towards improvements of audio processing chains for virtual or augmented reality.

Aim

The aim is to improve perceived realism in VEs through the development of a system aware of objects and materials whose physical properties would be mapped to functions of audio rendering pipelines.

Objectives

- Review current research on the area of realistic sound propagation in virtual environments.
- Drawing on previous work, plan the creation of a system for realistic sound propagation in virtual environments.
- Explore and test methods for building a system aware of obstacles affecting sound propagation.
- Build a system that extract acoustics properties from object materials and physical attributes to calculate sound absorption.
- Integrate the spatial information extracted from the environment into the audio signal emitted from sound sources through real-time signal processing algorithms.
- Perform tests on a multisensory application conducting subjective experiments to assess whether there are noticeable improvements on perceived realism.
- Compare the results with state-of-the-art realistic audio renderers to evaluate quality and efficiency for potential real-world applications.

Research Questions

Wearable augmented reality (AR) devices offer depth estimation capabilities that provide information on the surrounding environment. During sound propagation, when waves encounter objects and surfaces, diffraction and reflections occur. There is constantly expanding research in the field of diffraction and scattering calculation for physical modelling and virtual acoustics using FDTD models and adaptive frustum tracing. Their real-time implementations are usually computationally expensive due to the number of reflections occurring in complex scenes. Since 2006, GPU-based parallel computing approaches such as CUDA have been employed to improve performance for real-time systems (Savioja 2010, Taylor *et al* 2009).

Schissler *et al* (2014), presented an approach to a real-time diffraction computation system demonstrating that audio rendering in complex scenes can be augmented with wavelength-dependent algorithms based on geometric acoustics and ray theory.

Image classification methods and computer vision have been proposed to estimate surfaces and predict their materials. This would enable the computer to gain an understanding of surfaces, objects and materials composing a scene. Such information could be fed in the signal processing chain to integrate spatial information to a sound signal emitted by a sound source in the virtual environment. In addition, Andrearczyk and & Whelan (2006) demonstrated how with Convolutional Neural Networks it is possible to classify object textures. Their model achieved over 70% accuracy on a total of 10 datasets and showed predictions capabilities for 1000 classes. Based on findings and recommended directions presented across the cited papers, the following research questions are proposed.

A. What methods can be employed to extract physical attributes of objects existing in complex scenes in order to calculate their impact on sound propagation and diffusion occurring in virtual environments?
B. How can sound absorption of surfaces and objects can be calculated in order to generate impulse responses encapsulating spatial information of complex scenes.
C. Based on findings and methods derived from research question A and B, would a real-time system, aware of obstacles and sound absorption in complex scenes, be able to process sound signal emitted from sound sources with spatial information? Would audio processing chains in virtual or augmented reality benefit from it? Would the system improve subject's task performance and presence evoked in virtual or augmented reality?

The diagram shown in Fig. 1 represents a potential system design that would test the hypotheses derived from the research questions.

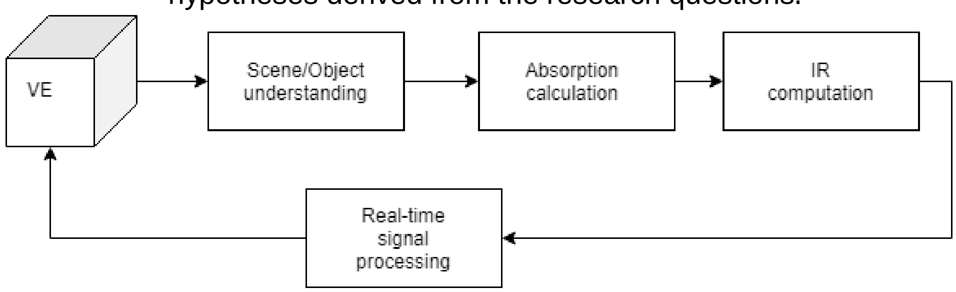

Fig. 1: a diagram showing a design for the system proposed. a method for object/scene understanding is applied in order to extract information that would be used to calculate sound absorption of surfaces and objects. As a result, impulse responses would be computed and integrated to the sound signal through real-time algorithms.

Significance and Impact of Research

A model for augmented audio would lead to realistic perception of virtual sound sources, as presence is also evoked by realistic audio rendering. This could also enable users of interactive applications to experience what a virtual object would sound like in a real environment. There is appetite for enhancements of realism and interactivity for design of serious games and informative experiences for Cultural Heritage contexts. Schofield *et al* (2018) proposed a design for a VR museum allowing visitors to experience virtual scenes built by a team composed by artists, archaeologists and historians. Such project would benefit from a model for augmented audio. In addition, predicting how a virtual room would sound like could contribute to 3D modelling and architectural planning.

Additional outputs and possible contributions include:

1. Improvement of subject's task performance in multisensory systems, through spatial audio. In interactive applications, augmented audio and spatial sound enable users to localise and identify objects and entities in environments without relying on visual information.
2. Realism in computer games can be improved with an audio render that is informed of materials, dimensions and acoustic features of environments.

References

Andrearczyk, V., & Whelan, P. F. (2016). Using filter banks in Convolutional Neural Networks for texture classification. Pattern Recognition Letters, 84, 63–69. https://doi.org/10.1016/j.patrec.2016.08.016

Larsson, P., Väljamäe, A., Västfjäll, D., Tajadura-Jiménez, A., & Kleiner, M. (2010). Auditory-Induced Presence in Mixed Reality Environments and Related Technology. 143–163. https://doi.org/10.1007/978-1-84882-733-2_8

Savioja, L. (2010). Real-time 3d finite-difference time-domain simulation of low-and mid-frequency room acoustics. Helsinki, Finland Department of Media Technology, Aalto University School of Science and Technology Espoo, Finland. Audio, 1–8.

Schissler, C., Mehra, R., & Manocha, D. (2014). High-order diffraction and diffuse reflections for interactive sound propagation in large environments. ACM Transactions on Graphics, 33(4). https://doi.org/10.1145/2601097.2601216

Schofield, G., Beale, G., Beale, N., Fell, M., Hadley, D., Hook, J., Thresh, L. (2018). Viking VR: Designing a virtual reality experience for a museum. DIS 2018 - Proceedings of the 2018 Designing Interactive Systems Conference, 2018(August), 805–816. https://doi.org/10.1145/3196709.3196714

Stanney, K. (1995). Realizing the full potential of virtual reality: human factors issues that could stand in the way. In Proceedings - Virtual Reality Annual International Symposium.

Taylor, M. T., Chandak, A., Antani, L., & Manocha, D. (2009). RESound: Interactive sound rendering for dynamic virtual environments. MM'09 - Proceedings of the 2009 ACM Multimedia Conference, with Co-Located Workshops and Symposiums, 271–280. https://doi.org/10.1145/1631272.1631311

Vedad Hulusic (2011). Auditory-Visual Interaction in Computer Graphics.

Vorländer, M., Schröder, D., Pelzer, S., & Wefers, F. (2015). Virtual reality for architectural acoustics. Journal of Building Performance Simulation, 8(1), 15–25. https://doi.org/10.1080/19401493.2014.888594