Dynamic auralisations of two simultaneous sound sources using ambisonics IR convolutions

Daniel Castro (1), Henry Andrew (2)

(1) Stantec, Melbourne, Australia
(2) Stantec, Manchester, UK

ABSTRACT
This paper presents a software that allows users to experience a high quality auralisation in a virtual reality scenario. The novelty of this new software iteration relies on the ability to simultaneously auralise two different noise sources while continuously and dynamically update the acoustic output at the user position within VR scenario. We use convolutions with pre-computed ambisonics impulse responses as the basis for our method.

1 INTRODUCTION
Virtual reality (VR) is an increasingly important technology. It has uses ranging from medical rehabilitation, to aiding scientists view data sets intuitively and consumer entertainment. One of the key concepts in virtual reality is immersion. When executed well, an immersive VR experience can make a user feel as though they are physically surrounded by and connected to the virtual world. This is achieved by presenting information to as many of the users senses as possible which is updated and reacts to the users input in real time. Visually, VR systems have become very realistic, with simulation of complex lighting effects and recent developments such as real time ray tracing. However, audio simulations still face many problems and struggle to render a realistic physically accurate soundscape in real time. This paper documents a real time auralisation method which makes use of pre-computation to produce a high quality auralisation of a room in VR with multiple static sources.

2 EXISTING AUDIO RENDERERS
Creating a renderer which is capable of real time high quality auralisation of life like scenes is extremely challenging and, with current computational power, there are always compromises. A common compromise is limiting the scene to be auralised to a simple setup. Using a single source, simple geometrical shapes for a room, stationary and omnidirectional sources or receivers are all common examples. A lower quality of auralisation may also be accepted to maintain the real time requirement. Geometrical acoustics simulations struggle to simulate low frequencies, replicate wave effects such as diffraction (Schissler, Manocha, 2017) and may use a lower number of rays to keep computation times low. Other methods including artificial reverberation may turn to parametric approaches, with drastically reduced computational times at the cost of only providing a rough approximation of how the acoustics might sound.

Another common compromise is doing the simulation, or even auralisation offline. This is often used by methods involving convolutions with an impulse response (IR) to factor out the typically long times needed to simulate a high-quality IR. The drawback is that the path a receiver takes through the scene is now fixed to points which have been pre-computed. An existing rendering method described by (Raghuvanshi, Snyder, 2018) pre-computes highly accurate simulated data, then distils this down to a parameter set stored at set positions. The parameter set is used in a filter network to approximate the acoustics. Though this method gives a very believable auralisation, and does not require much computation at run time, it may not be able to capture the full acoustics of a complex real-world scene.

3 PREVIOUS WORK
We previously proposed a method well suited to the real time auralisation of real-world scenarios (Castro et al., 2019). This method pre-computes a grid of ambisonics impulse responses (AIR) over the area to be auralised. The sound at any given receiver position is approximated with the following method. The convolution of the dry source signal is taken with each of AIR's from the closest 4 grid points. These 4 signals are then weighted proportionally to the receiver's proximity to that grid point and added together. This allows for the high quality auralisations from IR's calculated offline to be used without sacrificing freedom of movement. Another advantage of this method is that it allows the use of measured as well as simulated data. The implementation of this method was previously limited to auralisations of a single source.

4 IMPROVEMENTS

The software has been re-worked to improve its usability and add extra functionality. It is now capable of auralising multiple sources simultaneously. Our testing at this point has been limited, but the software is capable of auralising at least 2 simultaneous sources in real time while also reserving enough computational power to render the accompanying visuals. Figure 1 shows the DSP layout of the audio renderer when auralising 2 sources with the "ARIR convolution grid" technique previously developed. To our knowledge, this is currently the only system capable of auralising multiple sources in real time with the accuracy gained from using IR convolutions. The software has been implemented as a dll file with an emphasis on flexibility and ease of use. It can be used in game engines such as unity as well as called from programming languages such as MATLAB, c#, python etc. The software also features the functionality to swap between auralisations of different scenes instantly in real time as seen in figure 2. This is helpful for comparative purposes and could, for example, be used to compare between a simulated and measured IR of the same room.

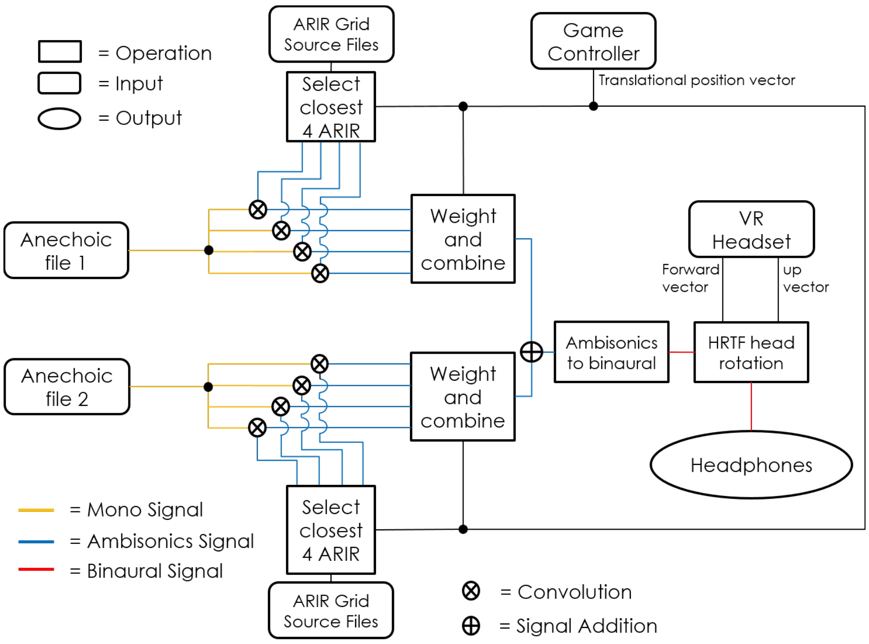

Figure 1: A signal processing view showing all the main operations on the signal from input to output.

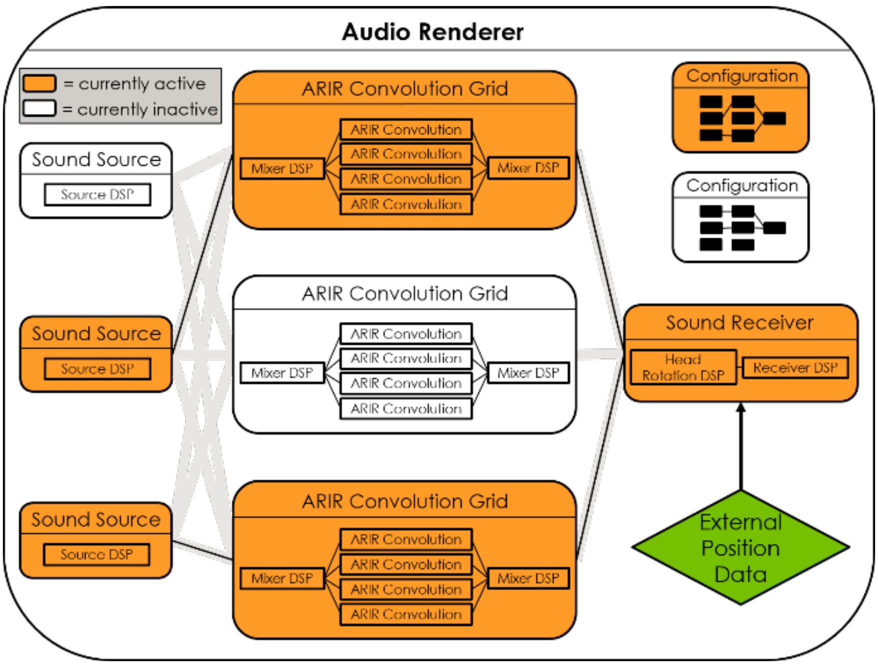

Figure 2: A software view showing how data can be loaded into the audio renderer and configured.

5 FUTURE WORK

The method we have used is limited in that it requires a large number of real time convolutions to be performed. This means that, with current computational limits, only a small number of sources can be simultaneously auralised. We can try to improve the efficiency of our algorithm to mitigate this as much as possible. Future work could also implement moving sources in addition to the moving receivers, and expand by including custom source directivity. Another improvement would be to implement the Doppler effect for fast moving receivers and/ or sources.

Another of the limitations is that 1st order ambisonics is not completely accurate for positioning the direction of a source. Adding an option in the software to use IR's of higher order ambisonics such as 2nd or 3rd order would provide greater positional feedback. This would however substantially increase the number of real time convolutions required, reducing the number of sources which could be simultaneously auralised.
Though our software has currently been designed to work with VR applications, another potential use is in Augmented Reality (AR). With the introduction of LIDAR sensors, improved microphones, speakers and processing power in some smart phones, this may become a possibility.

REFERENCES

Raghuvanshi, N. and Snyder, J. 2018. 'Parametric directional coding for precomputed sound propagation', *ACM Transactions on Graphics*, vol. 37(4), pp. 1-14. https://doi.org/10.1145/3197517.3201339

Schissler, C. and Manocha, D. 2017. 'Interactive sound propagation and rendering for large multi-source scenes', *ACM Transactions on Graphics*, vol. 36(4). https://doi.org/10.1145/3072959.2943779

Castro D., Verstappen, A. and Platt, S. 2019, 'Walk-through auralization framework for virtual reality environments powered by game engine architectures', In *Proceeedings of Acoustics 2019*, Victoria, Australia, November 2019. https://www.acoustics.asn.au/conference_proceedings/AAS2019/papers/p12.pdf