

Extension of the real-time Simulated Open Field Environment for fast binaural rendering

Tong Wang¹ and Bernhard U. Seeber

Audio Information Processing, Technical University of Munich, 80290 Munich

¹ E-Mail: Tong.Wang@tum.de

Background and Motivation

The “real-time Simulated Open Field Environment” (rtSOFE) [1,2] is the new real-time capable software component of the Simulated Open Field Environment (SOFE), the virtual acoustics facility developed and used by the Audio Information Processing group at the Technical University of Munich for hearing research. rtSOFE is used for simulating room acoustics, equalizing loudspeakers, convolving sound with room impulse responses and playing the sounds on a loudspeaker array in an anechoic chamber, thereby creating an interactive virtual acoustic scene for auditory research. rtSOFE is freely available and comes as a package with extensive documentation and configuration examples for various use cases in hearing research [3].

Recent advances in head mounted displays (HMD) for visual virtual reality provide researchers with a portable and affordable hardware tool for presenting visual scenes with high quality in experiments. To use rtSOFE with headphones and thus HMDs, we extended the rendering pipeline of rtSOFE to also synthesize binaural room impulse responses (BRIRs) based on filtering with head-related impulse responses (HRIRs) and implemented in the frequency domain via head-related transfer functions (HRTFs). In this paper we describe how we achieve the design target of a low latency binaural rendering system without introducing error-prone filter simplification. We provide also a benchmark of the implementation, discussions of the results and future development.

rtSOFE consists of two main programs, rtsofe.exe and convolver.exe, compiled from C/C++ code, and a collection of Python and MATLAB scripts [3]. This paper describes the extension of rtsofe.exe, the component that renders the room impulse response (RIR) or BRIR.

BRIR rendering workflow

Given a room acoustic model with surface material information, source and receiver positions and orientations and source directivity, rtSOFE simulates surface reflections with the image source method (ISM) [4], computes impulses contributed by the direct sound and each visible image source and synthesizes the multichannel impulse response.

Figure 1 shows rtsofe.exe’s workflow for interactive BRIR rendering. Routines highlighted in blue are implemented in this work.

Besides the HRTF rendering and BRIR creation routines, a position update of previously computed visible image sources is implemented. This allows the expensive image source computation to be skipped for small movements of the receiver or sound sources.

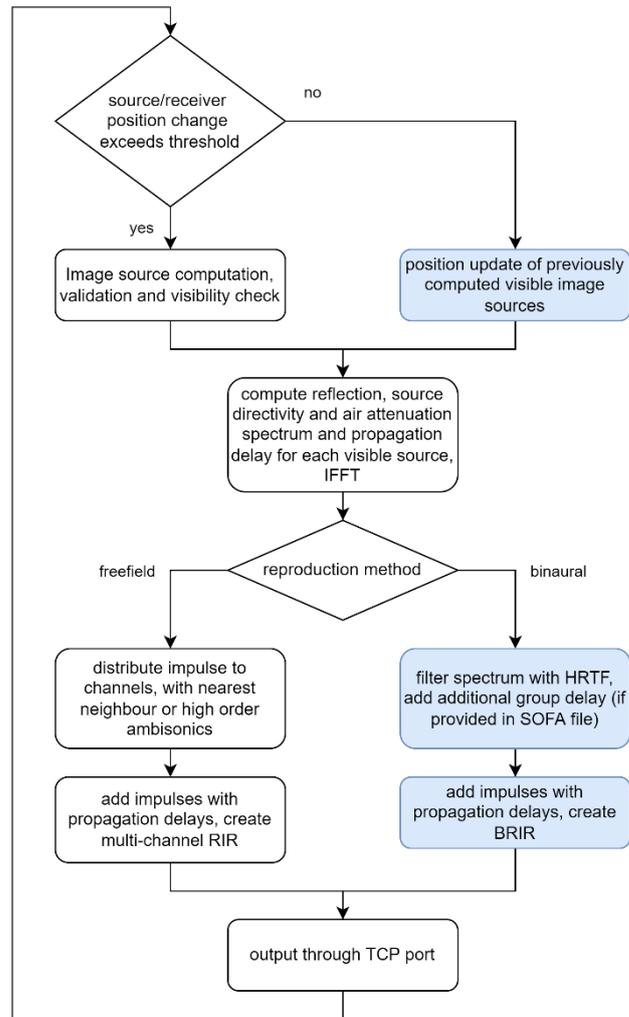


Figure 1: Workflow of rtsofe.exe for interactive rendering, with newly added routines in blue.

Design constraints and considerations

For 6 degree-of-freedom interactive BRIR rendering, using HRIR datasets with higher spatial resolution reduces artefacts due to spatial aliasing and thus gives the listener more accurate spatial cues for sound source localization. When using such a tool for auditory research, it must introduce as little systematic error as possible, going beyond the often required “plausibility” for normal hearing listeners. Hence rtSOFE was developed to render unaltered HRIRs of arbitrary length into the BRIR. In order to achieve real-time capability, an internal pre-interpolated HRTF set with fine grid is used, which can be loaded from a SOFA-file or interpolated from it at startup-time (c.f. Table 1). Alternately, a linear HRTF interpolation can be done at runtime if the HRTF set quantization or the memory overhead are unwanted. Our reasoning:

- Rendering unaltered HRIR provides full flexibility of the HRTF set used as it permits the use of individualized HRTFs and studies about HRTF features in dynamic rendering applications. For example, there is research on approaches to compute HRTF from a head model [5]. While the FEM simulation itself is not yet real-time capable, the simulation quality can already be interactively perceptually evaluated with rtSOFE when computed HRTFs are stored in the database and loaded into rtSOFE.
- There are algorithms to interpolate HRTF filters with small error [6], but efficient real-time implementations in C/C++ do not exist. However, they can be used offline to create the high-resolution HRTF dataset for use with rtSOFE.
- IIR filter based methods for HRTF interpolation may be efficiently implemented, but the IIR estimation of FIR filters introduces an error and the effect can be hard to predict in auditory experiments, hence we deem it unsuited for hearing research.
- Using FIR filters of the same length, a table lookup of nearest HRTF filter is, in theory, always faster ($O(1)$) than the most efficient real-time HRIR interpolation methods.
- Even consumer grade desktop computers have more than enough RAM to keep a high spatial resolution HRTF dataset in memory. Using HRIR filters of 256 taps and the spatial resolution in Table 1, we only need to keep about 150 MB of filter data in computer RAM.

Table 1: Azimuthal resolution of the internal pre-interpolated HRTF set as a function of elevation angle range. Elevation resolution is 1° .

Elevation angle range (deg)	Angular distance between neighboring directions (deg)
± 90	360
[60, 90) and (-90, -60]	5
[30, 60) and (-60, -30]	1
(-30, 30)	0.5

Benchmark

Benchmark conditions

The benchmark focuses on refresh rate of rtSOFE.exe's BRIR rendering. Test scene and rendering parameters are listed in Table 2. The source moves in a circular trajectory of 1 meter radius around the fixed receiver in the center. Position updates are sent by a Python script over OSC protocol. The benchmark is conducted on a computer equipped with a Ryzen 9 3900XT CPU with 12 physical cores (24 logical cores) and 64 GB RAM. The operation system is Windows 10. Logical cores #0 ~ #19 are assigned to the process rtsofe.exe through CPU affinity settings provided by the operation system. Logical cores #20 ~ #23 are assigned to other processes, including the Python process and convolver.exe.

Table 2: Benchmark conditions

Attenuation threshold for reflection synthesis	-60dB
Source movement speed	1 m/s
HRIR filter length	256 samples
Receiver position	(2,2,2)
Initial emitter position	(2,3,2)
Emitter position update rate	350 Hz
Number of surfaces in room acoustic model	22
Source movement threshold for IS validity and visibility	0.1 m

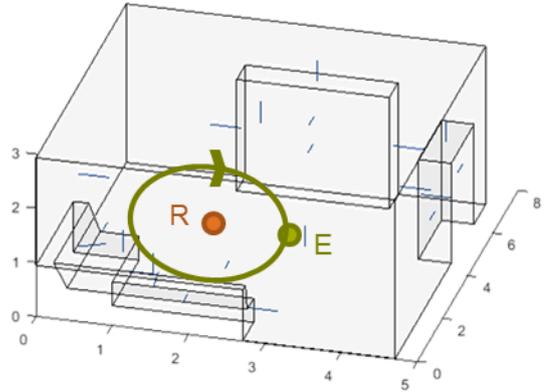


Figure 2: Room acoustic model

The “Source movement threshold for IS validity and visibility” parameter means that a full image source (IS) computation will only take place when the source moves more than 0.1 m from the place of the last full IS computation. With a source movement speed of 1 m/s, a computation of the full IS tree is expected to happen 10 times per second.

The benchmark has been performed with both pre-interpolation and real-time interpolation; for both conditions, rtSOFE is set to render reflections from order 4 to 7. rtSOFE's HRIR spatial interpolation is provided by “mysofa_getfilter_float” function of the “libmysofa” library [7]. “Libmysofa” implements a panning with Delaunay triangulation and inverse distance weighted sum of three filters.

Results & Discussion

Table 3: BRIR update rate of rtsofe.exe for given reflection orders with pre-interpolated HRTF and HRTF interpolated in real-time.

Rendered reflection orders		0-4	0-5	0-6	0-7
Average refresh rate (Hz)	Pre-interpolation	350	348	241	9
	Real-time interpolation	350	301	208	7.45

Table 3 presents the update rates achieved in this configuration. The target update rate 350 Hz (i.e. the rate of our motion tracker in the anechoic chamber) can be achieved for 5th order reflection computations. This means that with an audio buffer size of 128 and a sampling rate of 44100Hz, rtsofe.exe can provide a new BRIR for every input buffer.

Interpolation method

From rendering order 5, the impact of the real-time interpolation with the “libmysofa” library [7] becomes visible compared to the table lookup approach, creating an almost 14% decline in rendering rate. Since the HRTF-interpolation is only a small step in a long computational pipeline, the impact of this simple linear interpolation appears substantial. While rtSOFE allows the user to load a SOFA-file with HRIRs interpolated to a high spatial resolution, incorporating a more advanced interpolation method would benefit rtSOFE’s general handling and rendering quality for sparse data sets, however, the computational cost might be permissive for real-time applications.

Acknowledgments

rtSOFE v1.1 is published and available for download at <https://zenodo.org/record/5648305>.

The rtSOFE hard- and software was funded by BMBF 01 GQ 1004B.

rtSOFE application development is funded by Project C5 in the Collaborative Research Center SFB 1330 “Hearing Acoustics: Perceptive Principles, Algorithms and Applications” (HAPPAA), Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 C5.

References

- [1] Seeber, B.U., Kerber, S., & Hafter, E.R. (2010). A System to Simulate and Reproduce Audio-Visual Environments for Spatial Hearing Research, *Hearing Research* 260, 1-10.
- [2] Seeber, B. & Clapp, S. (2017). Interactive simulation and free-field auralization of acoustic space with the rtSOFE. *The Journal of the Acoustical Society of America*. 141. 3974-3974. 10.1121/1.4989063.
- [3] Seeber, B. U., & Wang, T. (2021). real-time Simulated Open Field Environment (rtSOFE) software package (1.1). Zenodo. <https://doi.org/10.5281/zenodo.5648305>.
- [4] Borish, J. (1984). Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6), 1827-1836.]
- [5] Ziegelwanger, H., Kreuzer, W., & Majdak, P. (2015, July). Mesh2hrtf: Open-source software package for the numerical calculation of head-related transfer functions. In *22st International Congress on Sound and Vibration*.
- [6] Bernhard, J., Gomez, G., & Seeber, B. U. (2015). Time-domain interpolation of head-related transfer functions with correct reproduction of notch frequencies. In *Fortschritte der Akustik - DAGA'15*, pp. 1126-1127.
- [7] Hoene, C., Patino Mejia, I. C., & Cacerovschi, A. (2017, May). MySofa—Design Your Personal HRTF. In *Audio Engineering Society Convention 142*. Audio Engineering Society.