ICA 2022
24th INTERNATIONAL CONGRESS ON ACOUSTICS
October 24th – 28th, 2022 Gyeongju, Korea

PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

ABS-0439

Perceptually informed interpolation and rendering of spatial room impulse responses for room transitions

Thomas McKENZIE[1]; Nils MEYER-KAHLEN[1]; Rapolas DAUGINTIS[2]; Leo McCORMACK[1]; Sebastian J. SCHLECHT[1,3]; Ville PULKKI[1]

[1] Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

[2] Dyson School of Design Engineering, Imperial College London, London, UK

[3] Media Lab, Department of Art and Media, Aalto University, Espoo, Finland

ABSTRACT

The acoustics of coupled rooms is often more complex than single rooms due to the increase in features such as double-slope decays, direct sound occlusion and anisotropic reverberation. For directional capture, analysis and reproduction of room acoustics, spatial room impulse responses (SRIRs) can be utilised, but measuring SRIRs at multiple positions is time consuming, and thus it may be desirable to interpolate between a sparse set of measurements. This paper presents a perceptually informed interpolation method for higher-order Ambisonic SRIRs that is robust for coupled rooms. It uses minimum-phase magnitude interpolation of the direct sound which is steered to the relative direction of arrival, sector-based early reflection interpolation in the frequency domain, and relative RMS matching for late reverberation. A method for rendering up to six degrees-of-freedom datasets of SRIRs is then presented using a time-varying partitioned convolution audio plugin, which is open-source and has made available for download. Finally, a listening test is conducted to assess the perceptual quality of interpolating between coupled room SRIR measurements with varying inter-measurement distance. The results suggest that for the tested scenario, using the presented interpolation method, a 50 cm inter-measurement distance is perceptually sufficient.

Keywords: Spatial room impulse response, six degrees-of-freedom, SRIR interpolation

1. INTRODUCTION

The human auditory system retrieves important spatial cues from the acoustics of a room. Several characteristics of reverberation are dependent on the source and receiver positions in a room, such as direct-to-reverberant ratio, early reflections and modal coupling, while reverberation time remains largely constant. Coupled room acoustics is more complex, with the emergence of double-slope decays in the room response, edge diffraction and portalling effects (1, 2), all of which vary with inter-room listener and source positions and coupling aperture size. Portalling is used in this paper to refer to scattering and diffraction around the coupling aperture, which gives the perception of the sound source location at the coupling aperture (3). When a listener moves in a simple shoebox room, acoustical changes tend to be smooth and gradual. In the transition between coupled rooms, however, rapid changes in acoustics can occur with small positional changes (4).

Recent literature has investigated how room acoustics measurements, known as room impulse responses (RIRs), can be used to evaluate the acoustical changes with different receiver positions inside a single room, both for virtual reality (5, 6) and dereverberation applications (7). For 6DoF rendering of sound scenes with Ambisonic spatial room impulse responses (SRIRs) at multiple positions in space, a convolution plugin that can switch between SRIRs in real time is needed, followed by an auralisation method such as binaural rendering (8, 9, 10).

Measurement or simulation of SRIRs at multiple positions at a high measurement resolution can be time consuming and computationally expensive, and therefore it can be desirable to interpolate between

[1] thomas.mckenzie@aalto.fi

a sparse set of measurements. The perceptual requirements for inter-measurement distance vary with auditory stimuli (11, 12), whereby sounds with limited frequency bandwidth can forgive larger distances between measurements (13), and the greater diffuseness of late reverberation allows for different measurement distances for different parts of the impulse response (5). For different receiver positions inside coupled rooms, however, the requirements may vary due to the increased acoustical complexity.

Interpolation of mono or binaural RIRs has been approached in many ways in the past: 1) Dynamic time warping (14), where the time axes of the nearest RIRs are stretched until they align; 2) Modal interpolation using a general solution to the Helmholtz equation (15), which is effective for non-uniform spatial distributions of RIRs at low frequencies; and 3) A combination of plane wave decomposition and time-domain equivalent source methods (11). Moving into SRIRs, a first-order interpolation method is presented in (16), which separates input SRIRs into specular parts, which are the direct sound and early reflections, and the diffuse parts. These are interpolated separately, where the specular parts are interpolated individually using direction of arrival estimations. In (12), a similar method is presented for early reflection interpolation between the nearest three receivers, with simpler interpolation of residual signals.

As the transition between coupled rooms is highly complex, it requires great care in reproduction. Interpolation between two coupled room RIRs is likely to be a more demanding task than for two RIRs inside the same room. This paper presents a perceptually informed SRIR interpolation method for higher-order Ambisonic SRIRs, which utilises minimum-phase magnitude interpolation of the direct sound steered to the estimated direction of arrival, sector-based early reflection interpolation in the frequency domain, and relative RMS matching late reverberation interpolation. A method for rendering up to six degrees-of-freedom datasets of SRIRs is then presented using a time-varying partitioned convolution audio plugin. Finally, a listening test is conducted in virtual reality to assess the perceptual quality of interpolating between coupled room SRIRs with varying inter-measurement distance using a previously measured dataset of the transition between coupled rooms.

The paper is laid out as follows: Section 2 details the interpolation method and Section 3 describes the convolution plugin and dynamic binaural rendering. Section 4 then presents the methodology and results of a listening test conducted in virtual reality to evaluate different inter-measurement resolutions of a dataset of coupled room SRIRs. Finally, Section 5 presents concluding remarks and proposes further work, and MATLAB code for the interpolation method and an open-source virtual studio technology (VST) plugin for the 6DoF convolution are made available for download, with the links provided at the end of the paper.

2. PERCEPTUALLY INFORMED SRIR INTERPOLATION

This section describes the method of perceptually informed interpolation between SRIRs. The method is designed for 3D sets of SRIRs. The maximum spherical harmonic (SH) order is denoted in this paper as N, with the order of an individual SH component denoted by n and the degree denoted by m. A MATLAB implementation of the interpolation method is available for download (see Section 6 for download link). In this paper, the Ambisonic Channel Numbering (ACN) and semi-normalised (SN3D) conventions are employed.

The SRIR measurements were made at J points at coordinates $\mathcal{P}_J \subset \mathbb{R}^3$. At each of those points a directional room impulse response was measured, which was encoded to the SH domain and is denotes as $\mathbf{h}_j(t) \in \mathbb{R}^{(N+1)^2}$. These responses are then interpolated to a dense set of $I > J$ points at positions $\hat{\mathcal{P}}_I \subset \mathbb{R}^3$. The distance between a point from the set of measurement points $\mathbf{p}_j \in \mathcal{P}_J$ and a point from the set of interpolation points $\hat{\mathbf{p}}_i \in \hat{\mathcal{P}}_I$ is denoted as

$$v_{i,j} = |\hat{\mathbf{p}}_i - \mathbf{p}_j|_2, \tag{1}$$

where $|\hat{\mathbf{p}}_i - \mathbf{p}_j|_2$ denotes the Euclidean distance. With the definition of the distance, it is possible to find the subset of $J' = 2^D$ measurement points, which contains the measurements closest to any interpolation point $\hat{\mathbf{p}}_i$, where D is the dimensionality in which the measurement points are arranged. Therefore, a 1D set of SRIRs in a line will have two nearest measurements; a 2D set of SRIRs in a grid will have four nearest measurements, and a 3D set will have eight nearest measurements. This gives a subset of nearest points $\mathcal{P}_{J'}^{(i)} \subset \mathcal{P}_J$ for each interpolation point. As all steps described next are carried out for each interpolation point, the index i is omitted for readability.

2.1 Direct Sound

In this study, the direct sound is taken as the first 4.17ms of the input SRIRs (200 samples at 48 kHz), though this value is adjustable. The method assumes RIR onsets are time-aligned. The direction of arrival (DoA) of the direct sound in each input SRIR is first estimated using the time-averaged pseudointensity vector, $\mathbf{i} \in \mathbb{R}^3$, which is derived from the first-order SH components as

$$\mathbf{i} = \sum_{t=1}^{200} [h_1(t)h_4(t),\ h_1(t)h_2(t),\ h_1(t)h_3(t)]^{\mathrm{T}}, \tag{2}$$

where superscript $^{\mathrm{T}}$ denotes transposition.

For each interpolation point, the direct sound direction $\hat{\theta} \in \mathcal{S}^2$ needs to be determined. For a non-occluded source, a geometrically correct method would be to estimate the sound source location based on the direct sound DoAs observed at the measurements. This can be done by finding the point that is closest to all lines along the DoAs, ideally their intersection point. Then, the direct sound direction could be computed at the interpolated point. In coupled rooms, where the sound source can potentially be occluded, see for example loudspeakers 2 and 3 in Fig. 2a, this procedure may cause problems. Between two measurements, the location of the first sound energy will change and such geometrical solutions may give arbitrary results. Therefore, a simpler approximate algorithm was used in this study to estimate the sound source location. The direct sound direction at each interpolation point was set to

$$\hat{\theta} = \sum_{j'} \theta_{j'} g_{j'}, \tag{3}$$

where $g_{j'}$ are distance weights obtained from the inverse distances between the interpolated positions and the nearest measurement positions

$$g_{j'} = \frac{v_{j'}^{-1}}{\sum_{j'=1}^{J'} v_{j'}^{-1}}. \tag{4}$$

When the direct sound coincides with the measurement position, the direction is correct. Also, when the sound source is sufficiently far away, or the spacing of measurement points is small, the error introduced by this simplification is small. In case of occlusion at some position, a smooth interpolation curve emerges between the direction of a visible direct sound, and the first energy arriving from an occluded sound source.

For the minimum phase direct sound interpolation, the omnidirectional channels of the nearest input SRIRs are first converted into the frequency domain. The spectra are then 1/3 octave smoothed, magnitude weighted based on the gains $g_{j'}$, and made minimum phase. The spectra are then summed, and encoded into SH at the target angle $\hat{\theta}$. The interpolated direct sound is then amplitude normalised based on the gain weighted RMS of the nearest measurements. This procedure ensures that the effect of the sound source directivity is accounted for at the interpolated position.

2.2 Early Reflections

For the early reflection interpolation, firstly the transition time t_{EL}, which is the cutoff between early reflections and late reverberation, is calculated separately for each input SRIR based on the energy decay curve passing a set threshold value (17). The omnidirectional channel of each SRIR is first bandpass filtered at 1 kHz, then normalised to a maximum amplitude of 1, and Schroeder integration is used to obtain the energy decay curve (EDC):

$$D(t) = \int_t^{\infty} h^2(\tau) d\tau. \tag{5}$$

In this study, values of t_{EL} are calculated as $t_{\mathrm{EL}} = D(t)/10$, rounded to the nearest 1000 samples, which generally fall between 80ms and 250ms for the room transition dataset (4). This is on the higher end of typical early reflection cutoff times reported in the literature (17, 18, 19).

The early reflections are interpolated and equalised at different directions on the sphere by using beamforming and reconstruction (20). For this, the measured SH domain responses are analysed with

a set of max-$\mathbf{r}_\mathrm{E}$ beams directed to a dense set of L quasi-uniformly arranged directions in a so called t-design $\mathbf{\Theta}_\mathrm{t}$ (21),

$$\bar{\mathbf{h}}_j(t) = \frac{4\pi}{L}\mathbf{Y}_N \mathrm{diag}_N\{w_n\}\mathbf{h}_j(t), \tag{6}$$

where $\mathbf{Y}_N \in \mathbb{R}^{L\times(N+1)^2}$ is a matrix of real spherical harmonics evaluated at directions $\mathbf{\Theta}_\mathrm{t}$, and $\mathrm{diag}_N\{w_n\}$ is a diagonal matrix of beamforming weights, with one unique weight for all SH components belonging to each order. The t-design with the least number of points that fulfills $T \geq 2N+1$ is selected. For the fourth-order SRIRs used in this paper for example, the t-design has 48 points. The beam signals are weighted with the distance weights and summed together

$$\bar{\mathbf{h}}(t) = \sum_{j'} g_{j'}\bar{\mathbf{h}}_{j'}(t). \tag{7}$$

Next, the summed signals are equalised to match the weighted sum of the magnitude spectra in each direction. Equalisation is needed to rectify any comb filtering artefacts that may arise from the summing of correlated signals, and is most apparent when the SRIRs to be interpolated are a greater distance apart. Every beamformed signal is equalised separately, such that colouration is removed in each direction. The equalisation is performed in the frequency domain in equivalent rectangular bandwidth (ERB) frequency bands (22):

$$\mathrm{BW}_\mathrm{ERB} = 24.7(4.37\times10^{-3}f_c + 1), \tag{8}$$

where f_c is the centre frequency. In this study, 48 frequency bands are employed with the lowest frequency at 10 Hz, which approximates to 1/3rd octave bands. For each ERB band, the target RMS is a sum of the RMS of each amplitude-weighted nearest SRIR beam divided by the current RMS of the interpolated beam. An equalisation curve is then calculated by linear interpolation of each ERB band target RMS, between 20 Hz and 20 kHz. After the directionally equalised responses for every directional response in $\bar{\mathbf{h}}^{(\mathrm{EQ})}(t)$ are obtained, they are brought back in the SH domain using

$$\mathbf{h}(t) = \mathrm{diag}_N\left\{\frac{1}{w_n}\right\}\mathbf{Y}_N^\mathrm{T}\bar{\mathbf{h}}^{(\mathrm{EQ})}(t). \tag{9}$$

2.3 Late Reverberation

The late reverberation interpolation follows much of the same method as used for the early reflections, but without the beamforming. The final interpolated SRIRs are a sum of the interpolated direct sound, early reflections and late reverberation, with cosine shaped amplitude windows used to fade between sections: 20 samples for direct sound to early reflections, and 10 ms for early reflections to late reverberation (both values are configurable). The interpolated set of SRIRs can then be saved as a spatially oriented format for acoustics (SOFA) file (23), in the same format as the input set, which makes it directly compatible with the convolution plugin to be described in the following section.

3. RENDERING

To auralise the SRIRs, a virtual studio technology (VST) plugin was developed, which allows for a monophonic input signal to be convolved with a specified SRIR from a set of input SRIRs. The plugin uses fast partitioned time-varying convolution in the frequency domain (24) with the overlap-add method to allow for real-time switching between input SRIRs, with minimal perceptual switching artefacts. It is based on a MATLAB prototype presented in a previous study (8), for which the reader is directed to for a more detailed description. The plugin is freely available as part of the SPARTA plugin suite (25) (see Section 6 for a download link), and the plugin graphical user interface (GUI) is presented in Figure 1.

To summarise the method, input SRIR filters are divided into blocks based on the digital audio workstation (DAW) block size and placed into a filter matrix. Each block is then zero padded with the same number of samples as the block size, and converted to the frequency domain using the discrete Fourier transform (DFT). Only the first half of the result is saved, which reduces computational load. The input monophonic signal to be convolved with the SRIRs is also converted into the frequency domain, and a

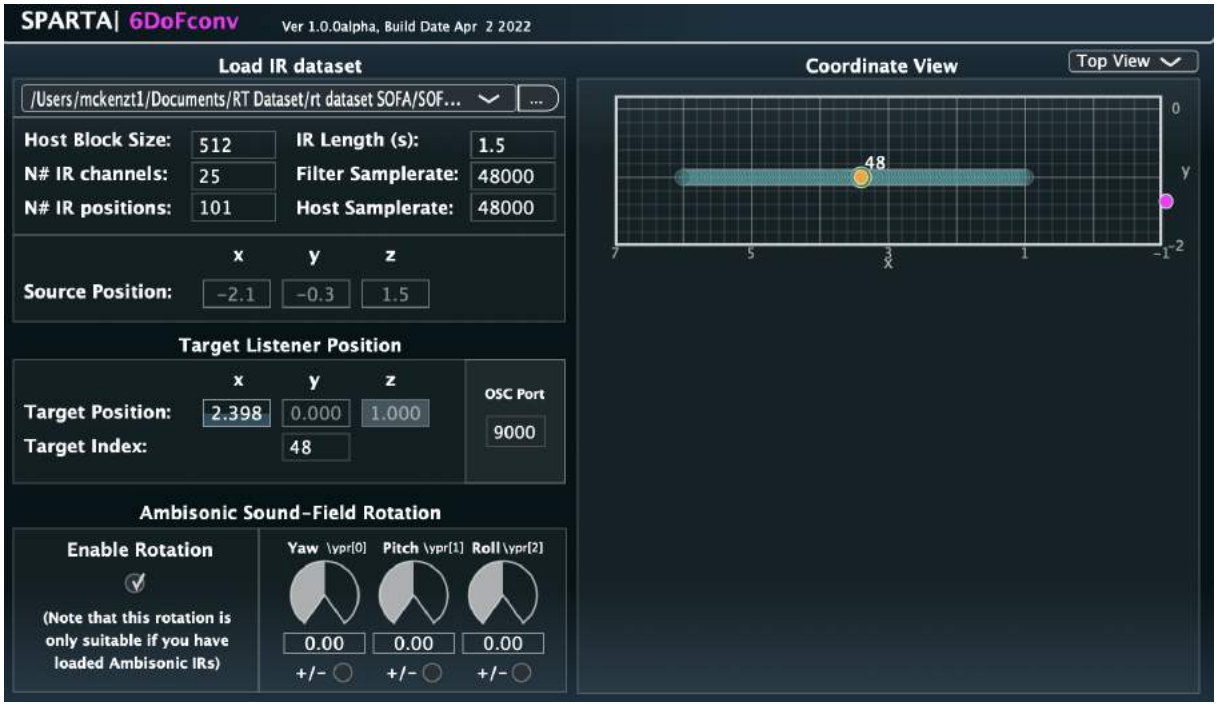


Figure 1 – The graphical user interface of the 6DoFconv VST plugin.

signal matrix is constructed of the input signal blocks, whereby for each input block time period, the input signal matrix is shifted by one block, dropping the oldest input block and placing the current signal block to the front.

Each block of the signal matrix is then multiplied by each block of the filter matrix corresponding to the chosen filter selection, and the results are then summed. The block is then duplicated, flipped and the complex conjugate is taken to rebuild the second half of the frequency domain signal, before being converted into the time domain using the inverse DFT. Convolution artefacts caused by the signal discontinuities when switching between the SRIRs are mitigated by cross-fading the convolved signals across multiple convolution blocks: the block convolved with the currently selected SRIR is saved for the next time period while the current output block is constructed from a linear cross-fade between the second half of the convolution block before the last and the first half of the last convolution block.

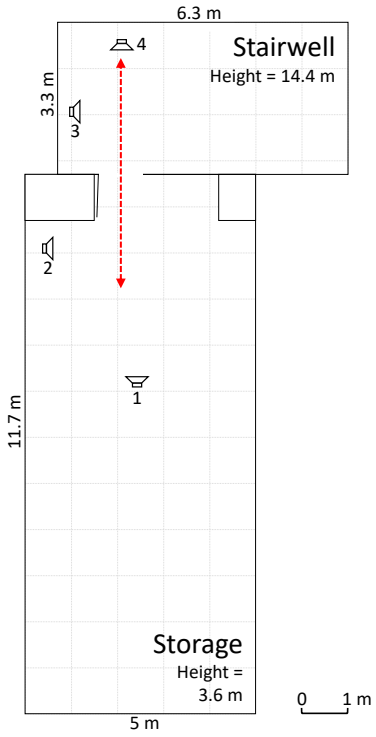
The GUI of the VST plugin allows for a user to select a path to a SOFA file with SRIRs and their associated listener positions encoded. It then loads the SRIRs and displays their positions in the Coordinate View on the right (the dimensions of the view window are determined by the coordinate range of the source and listener positions). The view can be chosen to be from the top or from the side using a drop-down menu on the top right corner. The listener position can be changed by dragging the orange dot in the coordinate view or moving the target position sliders on the left. When the position is changed, the plugin finds the nearest neighbouring SRIR based on the smallest Euclidean distance. The listener position can also be controlled via open sound control (OSC) messages from an external device, such as a head tracker. Additionally, the plugin includes an Ambisonic sound-field rotator for Ambisonic SRIRs.

4. EVALUATION

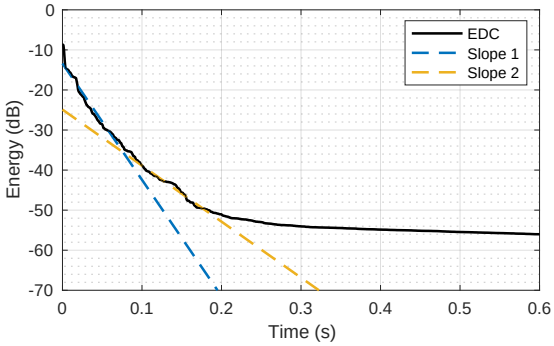
This section details the evaluation of the interpolation algorithm, which was carried out both numerically and perceptually. The set of measurements used in the evaluation was the *storage to stairwell* measurements from the Room Transition dataset of SRIRs at $N = 4$ (4), available under a Creative Commons license[1]. The room transition investigated in this paper is from a dry storage space to a more reverberant stairwell, with measured background noise levels of 32.8 dBA and 35.2 dBA and RT60s of 0.29 s and 0.73 s, respectively (8).

Figure 2a presents the room geometry and loudspeaker positions of the measurements, with four

[1] http://doi.org/10.5281/zenodo.4095493



(a) Room geometry and loudspeaker locations



(b) EDC (energy decay curve) for loudspeaker 2 at 150 cm inside storage space

Figure 2 – Room geometry and loudspeaker locations of the coupled room transition, and an energy decay curve illustrating the double-slope decay of the room transition. Measurements denoted by dashed arrow; loudspeaker numbers 1 and 4 retain a continuous line-of-sight between the loudspeakers and microphone for all measurement positions, 2 and 3 feature occlusion at some measurement positions (4).

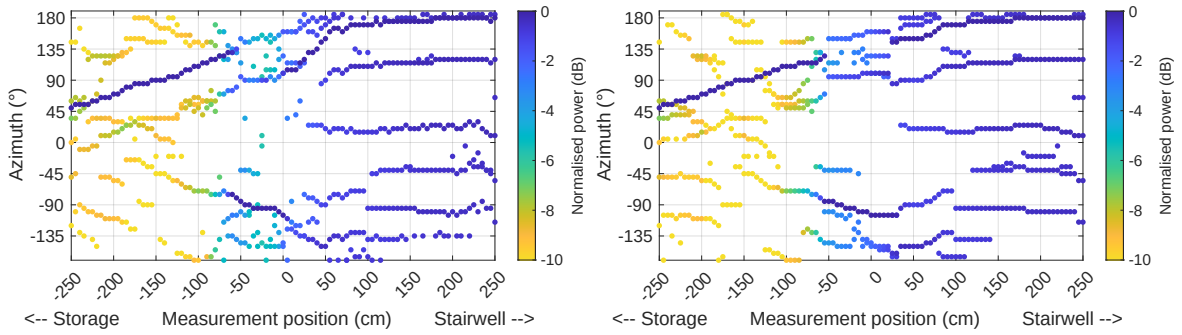
loudspeakers: two in each room; for which one retains a continuous line-of-sight (CLOS) between the source and receiver for all receiver positions, and two without CLOS. Figure 2b shows the EDC, calculated using equation 5, for loudspeaker 2 at receiver position 100 cm, which is 150 cm inside the storage space. The EDC illustrates the double-slope nature of the energy decay, caused by the combination of the reverberation times of the coupled rooms, whereby the amplitude of each room's single-slope decay is the only feature that is considered to change with receiver position (26).

To assess the interpolation, test sets of SRIRs were calculated from the original dataset of measured SRIRs, which has a 5 cm inter-measurement distance (IMD). This was done by interpolating (at 5 cm intervals) sparse versions of the original dataset with new IMDs of 10 cm, 20 cm, 50 cm, 100 cm, 200 cm and 500 cm (where the 500 cm case is just two SRIRs - one at either end). This was repeated for the measurements at the four loudspeaker positions illustrated in Figure 2a.

4.1 Numerical Evaluation

To numerically evaluate the interpolation method, the DoA of the room transition SRIRs was estimated first for the original dataset (with an IMD of 5 cm), and then for the test sets of SRIRs calculated from interpolation of the original dataset with a reduced IMD. DoA was estimated above 3 kHz, due to the order dependent filtering necessary for higher-order spherical microphone arrays (27), using a fourth-order SH steered plane-wave decomposition beamformer, that calculates the power at each chosen location on the sphere (28). DoA was estimated in five degree resolution for seven arrivals, referring to the direct sound and loudest early reflections.

The error in DoA was then calculated as the difference in azimuth angle between the DoAs calculated from the reference dataset and the interpolated datasets. A single azimuth error value E_θ for each interpolated dataset was then calculated as the mean of the absolute difference in estimated azimuth. Table 1



(a) Original 5 cm interval

(b) Interpolated from 50 cm interval

Figure 3 – Estimated direction of arrival of direct sound and early reflections for LS 2 (in storage, no continuous line-of-sight between the source and receiver, see Fig. 2a). Azimuth values are presented from $-170°$ to $190°$ for aided visibility around $\pm 180°$, and colour intensity is normalised separately to each measurement's maximum power value.

presents the results. In general, E_θ increases with higher IMD, which is expected. Some interesting results emerge when considering the differences between LS 1 and 2, in the storage space, and LS 3 and 4, in the stairwell. LS 3 and LS 4 have considerably lower E_θ for the low IMD sets, which may be explained by the higher reverberation time of the stairwell, leading to higher energy throughout the room transition. The E_θ significantly jumps at IMD = 500, suggesting the interpolation method is unable to accurately reconstruct the room transition acoustics at this distance.

Table 1 – Mean estimated DoA error E_θ in degrees between the reference dataset and the test SRIR datasets. IMD refers to inter-measurement distance of the test SRIR datasets, and LS X refers to the loudspeaker positions as illustrated in Figure 2a.

IMD (cm)	10	20	50	100	200	500
LS 1	11.0	14.0	16.2	21.0	18.7	24.5
LS 2	13.6	16.9	18.3	19.4	21.1	35.7
LS 3	4.88	7.63	10.5	14.6	12.0	22.0
LS 4	3.40	5.48	6.88	7.35	19.6	18.6

To better illustrate the DoA of the interpolated SRIR sets, the horizontal DoA for all source locations and measurement positions of LS 2 (in storage, no CLOS between the source and receiver) is presented in Figure 3, for the original SRIRs and for the interpolated SRIRs from 50 cm IMD. In the plot, a positive increase in azimuth denotes anticlockwise movement, and colour intensity is normalised separately for each measurement to the maximum power detected in that measurement, in order to illustrate the relative intensity of the dominant source direction to the other reflections. The overall trends are largely retained with the interpolation, though some details around the coupling aperture are somewhat less accurately captured.

4.2 Subjective Evaluation

To perceptually evaluate the quality of the SRIR interpolation, a listening test was conducted in virtual reality. The test paradigm was MUSHRA-like, with a hidden reference but no anchor. Participants were presented with seven conditions for which they could select one condition at a time, and were asked to walk the transition and rate the sound quality in terms of overall perceived similarity to the reference, with instructions to listen for all of localisation accuracy, colouration and reverberation. The reference condition was the original dataset of SRIRs, and the test conditions were the interpolated SRIR sets at different IMDs.

Two test stimuli were used: a dry recording of a drumkit, chosen for its transients, sharp attacks and wide range of frequency content, and an anechoic violin recording, chosen for its smooth and pe-

riodic waveform[2]. The 6DoFconv plugin was used to convolve the test stimuli with the set of SRIRs, whereby the SRIR was switched depending on the participant's position. The convolved signals were then rendered binaurally using the parametric higher-order DirAC binaural decoder (29). Mysphere 3.2 headphones were used for playback, which have been shown to offer high levels of passive transparency (30) which makes them suitable for experiments with both real and virtual sources (8, 9). Audio processing and programming of the listening test was conducted in Cycling 74 Max.

To display the room transition in virtual reality, three-dimensional models of the two rooms were captured using LiDAR technology from an Apple iPad Pro, with certain features enhanced in post processing, such as the doors and windows, using high resolution two-dimensional textures and sharper edges. Unity was used to render the visuals, which were displayed on an Oculus Rift S. The loudspeaker model was movable in the environment, such that whichever loudspeaker was currently playing was displayed (as determined in Max, and sent to Unity via OSC). User position and orientation data, for SRIR selection and sound field rotation in the 6DoFconv plugin, was sent from Unity to Max via OSC.

The listening test instructions and MUSHRA-like user interface were shown in the Unity virtual environment: the position of these was controlled by the Oculus left hand controller, and interactions made using the trigger on the Oculus right hand controller. To ensure participants stayed within the bounds of the SRIR measurements, a guiding line was placed at 1.2 m above the ground in the Unity scene, from 2.5 m inside the storage space to 2.5 m inside the stairwell, corresponding to the positions of the measurements. In the case that the participant strayed more than 25 cm from the guiding line in the X or Z axis, the screen flashed red and the audio cut out.

The listening test consisted of a total of eight trials: the four loudspeaker positions presented once with the drumkit and once with the violin. No repeats were conducted. Trial and condition ordering was randomised and double anonymous. The tests were conducted on 13 participants aged between 24 and 31 (11 male, 2 female) with self reported normal hearing and prior critical listening experience (such as education or employment in audio or music engineering).

4.2.1 Results and Discussion

The results of the listening test are presented as violin plots in Figure 4. Violin plots display both the density trace and box plot, which better illustrates the structure of the data over traditional box plots (31). The violin widths represent the density of data, median values are presented as a white point, interquartile ranges are marked using a thick grey line, the ranges between the lower and upper adjacent values are marked using a thin grey line, and individual results are displayed as coloured points.

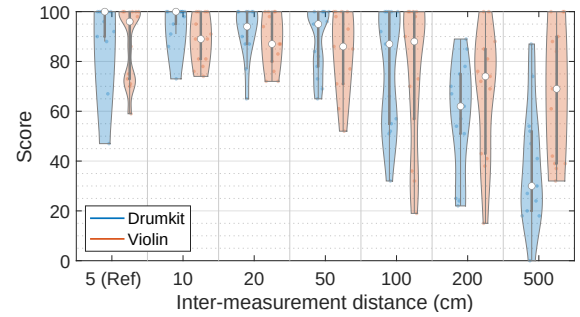
The results generally show that, with the presented SRIR interpolation method, IMDs up to 50 cm produced perceptually comparable results to the reference at 5 cm IMD. Even for the 100 cm IMD, median values were above 80 for 7 out of 8 tested conditions. At 200 cm and 500 cm IMD, scores were significantly lower, especially for LS 2 and LS 3, where there was no CLOS between the source and receiver, and the largest angular errors in the direct sound direction occur due the choice of direct sound location estimation. This is in fitting with the results shown in (8), which showed that a linear interpolation between the first and last measurements was rated as higher in naturalness for the two sound sources with CLOS (LS 1 and LS 4) than those without (LS 2 and LS 3).

To test the statistical significance of the results, the data was first tested for normality using the Shapiro-Wilk test, which showed not all data to be normally distributed, even when excluding the reference condition. Therefore, statistical analysis was conducted using non-parametric methods. Friedman's tests showed that the conditions were statistically significantly different ($p < 0.001$) for all stimuli and loudspeaker pairs except LS 1 with the violin stimulus: $\chi^2(6) = 8.32, p = 0.21$; in this configuration both 200 cm and 500 cm IMDs performed relatively well, with median values of 74 and 69, respectively.

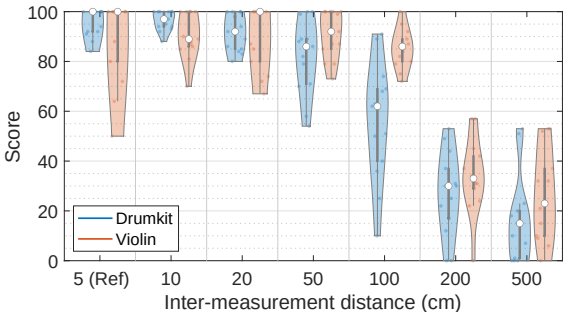
To look in more detail at the statistical significance of the difference between results, post-hoc pairwise Wilcoxon signed-rank tests with the Bonferroni-Holm correction were conducted: the results are presented in Figure 5. These confirm that the main differences in results are caused by the 200 cm and 500 cm IMDs in most cases. They suggest an IMD of 100 cm is sufficient for most cases, apart from LS 2 with the drumkit stimulus.

The different stimuli, a drumkit and a violin, on the whole produced relatively similar results, though at IMD $\geq$ 100 cm, the median rating of the drumkit was lower for 11 out of 12 cases. This suggests that the drumkit stimulus showed the artefacts of interpolation better, and could suggest that the choice of

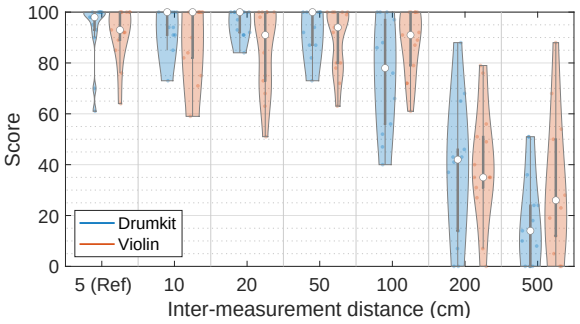
[2]Downloaded from `https://www.openair.hosted.york.ac.uk/`



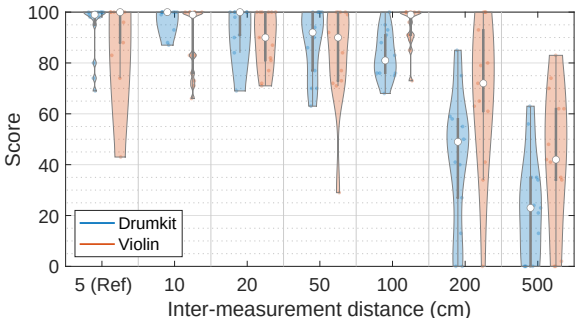
(a) LS 1 (in storage, with CLOS)



(b) LS 2 (in storage, no CLOS)



(c) LS 3 (in stairwell, no CLOS)



(d) LS 4 (in stairwell, with CLOS)

Figure 4 – Violin plots of the MUSHRA-like listening test results. CLOS refers to a continuous line-of-sight between the loudspeaker and listener for all listener positions (refer to Fig. 2a for loudspeaker positions and room geometries). Median values are a white point, interquartile range a thick grey line, the range between lower and upper adjacent values a thin grey line, and individual results are coloured points.

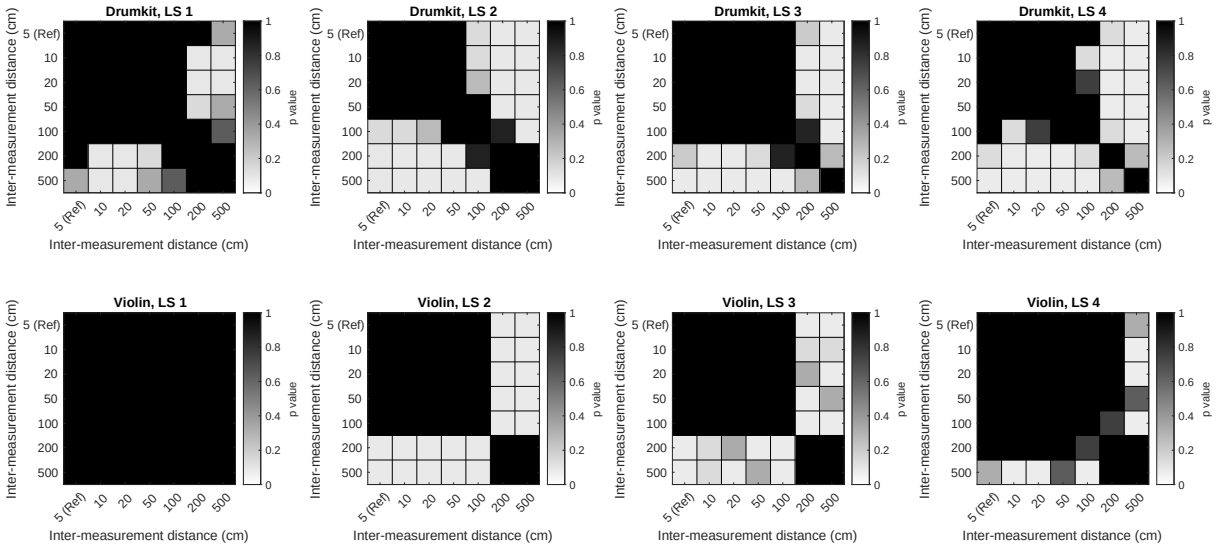


Figure 5 – Wilcoxon signed-rank (with Bonferroni-Holm correction) matrices of the listening test results between different conditions.

IMD in measuring could be influenced by the stimuli of the application.

5. CONCLUSIONS

This paper has presented an interpolation method for higher-order Ambisonic spatial room impulse responses (SRIRs), suitable for up to six degrees-of-freedom datasets and robust in interpolating measurements in the transition between coupled rooms. A time-varying partitioned convolution method then allows for real-time switching of SRIRs.

The system has been evaluated numerically, using direction-of-arrival (DoA) analysis, which shows that the interpolation seems relatively accurate even at an inter-measurement distance (IMD) of 10 times the original. A dynamic listening test has then been conducted in virtual reality, using visuals of three-dimensional models from room scans using LIDAR technology and parametric binaural decoding, where participants were able to walk through the transition in real time. The results showed that, using the presented interpolation method, IMDs up to 50 cm or in some cases 100 cm were rated as highly similar to the reference (IMD of 5 cm).

The evaluation showed that, even for a demanding acoustic scenario such as a room transition, the presented SRIR interpolation method is able to reduce the necessary inter-measurement distance, which allows for time and cost saving in measurements.

Further work will compare the presented SRIR interpolation method to other available methods, and quantify the improvements over a basic linear interpolation method. Additionally, the method should be used to interpolate between measurements in a single room, and the results compared to the evaluation in this study, to assess the feasibility of interpolating between measurements at a higher IMD when the acoustical changes are smaller.

6. DOWNLOAD

The presented interpolation method is available for download as MATLAB code[3], along with demonstration and analysis scripts, and the 6DoFconv VST plugin is now freely available as part of the SPARTA suite[4].

ACKNOWLEDGEMENTS

This research was supported by the Human Optimised XR (HumOR) Project, funded by Business Finland, and the EU's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 812719.

REFERENCES

1. Xiang N, Jing Y, Bockman AC. Investigation of acoustically coupled enclosures using a diffusion-equation model. J Acoust Soc Am. 2009;126(3):1187-98.
2. Billon A, Valeau V, Sakout A, Picaut J. On the use of a diffusion model for acoustically coupled rooms. J Acoust Soc Am. 2006;120(4):2043-54.
3. Raghuvanshi N, Snyder J. Parametric directional coding for precomputed sound propagation. ACM Trans on Graphics. 2018;37(4):1-14.
4. McKenzie T, Schlecht SJ, Pulkki V. Acoustic analysis and dataset of transitions between coupled rooms. In: IEEE Int. Conf. on Acoust., Speech and Sig. Proc. Online; 2021. p. 481-5.
5. Neidhardt A, Tommy AI, Pereppadan AD. Plausibility of an interactive approaching motion towards a virtual sound source based on simplified BRIR sets. In: AES 144th Conv. Milan; 2018. p. 1-11.
6. Stein E, Goodwin MM. Ambisonics depth extensions for six degrees of freedom. In: AES Int. Conf. on Headphone Technology. vol. 2019. San Francisco; 2019. p. 1-10.
7. Jeub M, Schäfer M, Vary P. A binaural room impulse response database for the evaluation of dereverberation algorithms. In: IEEE Int. Conf. on Digital Sig. Proc. Santorini; 2009. p. 1-5.
8. McKenzie T, Schlecht SJ, Pulkki V. Auralisation of the transition between coupled rooms. In: Immersive and 3D Audio: From Architecture to Automotive (I3DA). Online: IEEE; 2021. p. 1-9.

[3] https://github.com/thomas-mckenzie/srir_interpolation

[4] https://leomccormack.github.io/sparta-site/docs/plugins/sparta-suite/#6dofconv

9. Meyer-Kahlen N, Amengual Garí S, McKenzie T, Schlecht SJ, Lokki T. Transfer-plausibility of binaural rendering with different real-world references. In: Jahrestagung für Akustik - DAGA 2022. Stuttgart; 2022. p. 1-4.
10. McCormack L, Politis A, McKenzie T, Hold C, Pulkki V. Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple Ambisonic receivers. Journal of the Audio Engineering Society. 2022;70(5):355-72.
11. Antonello N, De Sena E, Moonen M, Naylor PA, Van Waterschoot T. Room impulse response interpolation using a sparse spatio-temporal representation of the sound field. IEEE/ACM Trans on Audio, Speech and Lang Proc. 2017;25(10):1929-41.
12. Müller K, Zotter F. Auralization based on multi-perspective Ambisonic room impulse responses. Acta Acustica. 2020;6(25):1-18.
13. Neidhardt A, Reif B. Minimum BRIR grid resolution for interactive position changes in dynamic binaural synthesis. In: AES 148th Conv. Online; 2020. p. 1-10.
14. Masterson C, Kearney G, Boland F. Acoustic impulse response interpolation for multichannel systems using Dynamic Time Warping. In: AES 35th International Conference. London; 2009. p. 1-10.
15. Das O, Calamia P, Gari SVA. Room impulse response interpolation from a sparse set of measurements using a modal architecture. In: IEEE International Conference on Acoustics, Speech and Signal Processing. Online; 2021. p. 960-4.
16. Zhao J, Zheng X, Ritz C, Jang D. Interpolating the directional room impulse response for dynamic spatial audio reproduction. Applied Sciences. 2022;12(4).
17. Hidaka T, Yamada Y, Nakagawa T. A new definition of boundary point between early reflections and late reverberation in room impulse responses. Journal of the Acoustical Society of America. 2007;122(326).
18. Meesawat K, Hammershøi D. An investigation on the transition from early reflections to a reverberation tail in a BRIR. In: International Conference on Auditory Display. Kyoto; 2002. p. 5-9.
19. Campos A, Sakamoto S, Salvador CD. Directional early-to-late energy ratios to quantify clarity: A case study in a large auditorium. In: Immersive and 3D Audio. Online; 2021. .
20. Hold C, McKenzie T, Gotz G, Schlecht SJ, Pulkki V. Resynthesis of spatial room impulse response tails with anisotropic multi-slope decays. Journal of the Audio Engineering Society. 2022;70(6):526-38.
21. Hardin RH, Sloane NJA. New spherical designs in three and four dimensions. In: IEEE Int. Symposium on Information Theory - Proceedings. vol. 441; 1995. p. 181.
22. Moore BCJ, Glasberg BR. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am. 1983;74(3):750-3.
23. Majdak P, Iwaya Y, Carpentier T, Nicol R, Parmentier M, Roginska A, et al. Spatially Oriented Format for Acoustics: A data exchange format representing head-related transfer functions. In: AES 134th Conv. Rome; 2013. p. 1-11.
24. Wefers F, Vorländer M. Efficient time-varying FIR filtering using crossfading implemented in the DFT domain. Proceedings of Forum Acusticum. 2014;2014-Janua(August 2015).
25. McCormack L, Politis A. SPARTA and COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods. In: Proceedings of the AES Int. Conf.. vol. 2019-March; 2019. p. 1-12.
26. Götz G, Hold C, McKenzie T, Schlecht SJ, Pulkki V. Analysis of multi-exponential and anisotropic sound energy decay. In: Jahrestagung für Akustik - DAGA 2022. Stuttgart; 2022. p. 1-4.
27. Daniel J, Moreau S. Further study of sound field coding with higher order Ambisonics. In: Proc. of the 116th Conv. of the Audio Eng. Soc. Berlin; 2004. p. 1-14.
28. Politis A. Microphone array processing for parametric spatial audio techniques [PhD Thesis]. Aalto University; 2016.
29. Politis A, McCormack L, Pulkki V. Enhancement of Ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing. In: IEEE Workshop on Applications of Sig. Proc. to Audio and Acoustics; 2017. p. 379-83.
30. Llado P, McKenzie T, Meyer-kahlen N, Schlecht SJ. Predicting perceptual transparency of headworn devices. Journal of the Audio Engineering Society. 2022;70(7/8):585-600.
31. Hintze JL, Nelson RD. Violin plots: a box plot-density trace synergism. The American Statistician. 1998;52(2):181-4.