Walkable auralizations for experiential learning in an immersive classroom

Samuel Chabot and Jonas Braasch
School of Architecture, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA

(Dated: 9 January 2023)

This paper is part of a special issue on Education in Acoustics. It proposes an experiential method for learning acoustics and consequences of room design through the rapid creation of audio-visual congruent walkable auralizations. An efficient method produces auralizations of acoustical landmarks using a 2D ray-tracing algorithm and publicly available floor plans for a 128-channel wave field synthesis (WFS) system. Late reverberation parameters are calculated using additional volumetric data. Congruent visuals are produced using a web-based interface accessible via personal devices which automatically formats for and transmits to the immersive display. Massive user-contributed online databases are harnessed through APIs such as those offered by the Google Maps Platform to provide near-instant access to innumerable locations. The approach allows the rapid sonic recreation of historical concert venues with adequate sound sources. The listeners can walk through these recreations over an extended user area (12 m × 10 m).

[https://doi.org(DOI number)]

[XYZ] Pages: 1–13

I. INTRODUCTION

Strong evidence supports the growing shift in science, technology, engineering, and math (STEM) education toward the inclusion of more active learning opportunities (Freeman *et al.*, 2014). Students engage with hands-on demonstrations and activities to apply traditional classroom knowledge and terminology to specific tasks. There exists much pedagogical research supporting these trends. The concept of "situated learning" is the locating of the learning experience in a specific setting, environment, or activity (Lave and Wenger, 1991). An essential message of this theory is that students "learn through doing." A report by the Education Resource Information Center (ERIC) states: "Applied to the classroom, situated learning is not only reflecting upon and drawing implications from previous experiences but is immersion in and with the experience" (Stein, 1998). A variety of examples of active learning demonstrations and environments to better understand classroom concepts exists in acoustics education. Students are presented demonstrations on acoustic beamforming (Carman, 2012), loudspeaker line arrays (Anderson *et al.*, 2012), and auralizations of acoustical enclosures (Vorländer, 2020), as proposed in this paper. Instructors note student engagement and enjoyment in active-learning introductory acoustics courses (Neilsen *et al.*, 2012).

William Bricken's 1990 report "Learning in VR" is perhaps the seminal work to describe the potential utilization of virtual reality for educational scenarios (Bricken, 1990). As technology has caught up with theory, VR applications today, including head-mounted displays (HMDs) and CAVE systems (Cruz-Neira *et al.*, 1993), can place students in active scenarios and often focus on either situating learners in their appropriate visual context (Chabot *et al.*, 2020) or visualizing learning concepts in novel ways (Limniou *et al.*, 2008). This project proposes the usage of a CAVE-like environment to place students within appropriate visual and, importantly, aural contexts to reinforce learning performed in a traditional classroom setting.

Through architectural acoustics education students are to become familiar with acoustic impulse responses and their relationship with the built environment, its materials, and subsequent perceptual outcomes. Room acousticians often utilize impulse responses for the characterizing and auralizing of interior spaces, including both existing and developing sites such as concert halls, churches, and mosques and historical and archaeological recreations (Murphy, 2006). Impulse responses are adequate for describing linear, time-invariant systems (LTI) with defined in- and outputs, such as those found in audio amplifiers. Acoustical enclosures can be treated as LTI systems, but lack defined in- and outputs. Therefore, in- and outputs are created by specifying source (loudspeaker) and receiver (microphone) positions within the enclosure. Its impulse response is then measured between these two locations. These measurements can also be estimated by defining source and receiver positions in computer-generated space, as is described in this paper. Sufficiently dry audio files can be convolved with these impulse responses to produce recordings as if they had been captured in the enclosure at the microphone position (Vorländer, 2020). For an educational introduction see (Braasch, 2018). These auralizations utilized in the classroom create dynamic examples through which students experience the complex relationships that shape perceptual room parameters (Pelzer *et al.*, 2014). In

subsequent discussions students are able to directly apply classroom knowledge and terminology to their experiences.

These auralizations are most often source-receiver specific: each loudspeaker/microphone pair describes the specified positions. To audition multiple locations and experience spatially-dependent acoustical characteristics, the calculation of multiple receiver positions is required. However, these still describe specific and discrete points in the environment. Additionally, their presentation binaurally over headphones or via techniques such as stereophony and higher-order ambisonics (HoA) over loudspeakers confines the listener to a smaller sweet spot – the area where spatial sound is accurately reproduced (Wierstorf *et al.*, 2017). These drawbacks make it difficult for groups to collectively "walk" through and explore auralizations together.

Wave-field synthesis (WFS) is a notable exception among spatial-audio techniques. Here, Huygen's concept of elementary waves (Huygens, 1690) and Fresnel's principle of interference (Fresnel, 1819) are used with a large array of loudspeakers to simulate the rectilinear propagation of sound waves of a virtual sound source behind the array. Similar work emerged as early as the 1930s: Steinberg and Snow used a linear "curtain" of equidistant microphones to capture a wavefront and reproduce it at a remote site with loudspeakers spaced at the same distances (Steinberg and Snow, 1934). A research team at Delft University was the first to design and implement a WFS system in practice (Berkhout, 1988; Berkhout *et al.*, 1993). This method allows for accurate sound reproduction across larger listening areas below a calculable spatial aliasing frequency. In this relatively larger region, listeners are able to localize recreated sound sources with greater accuracy than with HoA or stereophonic methods (Wierstorf *et al.*, 2017).

While much research has been devoted to accurately reproducing sound waves emitted from direct sound sources (Spors *et al.*, 2008; Ziemer, 2018), comparatively little focus has been set on the faithful reproduction of sound reflected from boundaries (Hulsebos *et al.*, 2002). Instead, wall reflections are often simulated using traditional stereophonic methods or spherical harmonics, methods built on a single point-like receiver position.

This project is realized over the extended listening area (12 m × 10 m) of Rensselaer's *Collaborative Research Augmented Immersive Virtual Environment* (CRAIVE-Lab), depicted in Figs. 1 and 2. Its concept uses a 2D method to render the direct sound signals and early reflections, a feasible approach considering that wave field synthesis (WFS) systems are usually 2D setups, where loudspeakers are placed along the perimeter of an extended listener area. The proposed method is compatible with the 2.5D operator for WFS (Verheijen, 1997). Section II discusses in detail the proposed auralization method, which allows users to auralize historic buildings from scaled floor plans, often—in contrast to full three-dimensional computer-aided design (CAD) models—publicly available.

The CRAIVE-Lab's design ideology seeks to put equal emphasis on the audio-visual experience. This is facilitated by eight short-throw projectors and a 360° micro-perforated front-projection screen which surrounds the listening area up to a height of 4.3 m. Located behind the screen at ear-height is a dense horizontal audio array of 128 loudspeakers. Six additional speakers are hung from the ceiling for elevation cues. Together, these two systems present congruent audio-visual content with balanced audio-visual fidelity. The balance of loudspeaker size, quantity, and density allows for low frequency reproduction to a cutoff at 37 Hz, 360° horizontal coverage, and a spatial aliasing frequency of approximately 900 Hz. With applied acoustical panels and curtains, the background noise of the equipment in the space measures approximately 38 dBA with the projectors off and 45 dBA with the projectors on, and the room's reverberation time measures about 0.89 seconds. While the WFS-array enclosing room is fairly large (floor dimensions: 16 m × 13.5 m), we plan to bring down the reverberation time considerably in the future by adding absorptive wall material.

The extensive floor area of the CRAIVE-Lab can host up to 49 participants at a time (per fire code), providing the opportunity to experience an extended section of the simulated acoustical enclosure. A trade-off of the larger floor plan is a greater distance to the screen, creating a more focused vertical field-of-view with a maximum of approximately 46 degrees (measured from the center of the room). While this often accounts for the most relevant content in a scene, it can mean the truncating of floor, ceiling, or vertically tall content, such as in a cathedral.

The 2D approach has some educational advantages over traditional 3D methods. Firstly, it allows for the creation of a model from less than an hour to a few hours given the complexity of the floor plan and additional volumetric estimations. A 3D model typically requires significantly more time to program and the necessary information is often unavailable. The rendering process is shorter, allowing students to render a 128-channel simulation in class, while a 3D method requires much more time for the same number of channels. Given that the CRAIVE-Lab and most other WFS systems position all loudspeakers on a 2D plane, the compromises of the dimensional reduction are acceptable, as long as the volume of the enclosure is considered to calculate the parameters for the late reverberation. For three-dimensional loudspeaker setups (e.g., dome configurations for higher-order ambisonics), the 2D approach has more setbacks, although the annotated floor-plan approach might still be useful for educational purposes. The biggest side effect for dome configuration will be the lack of ceiling reflections, which have been deemed important in concert-hall acoustics. For these setups, it might be useful to estimate and add at least one, first order ceiling reflection, for example using another 2D approach in the vertical plane that bisects a concert hall from front to back.

The process for creating paired visual content for environments such as the CRAIVE-Lab can be laborious and difficult for end users, and content for virtual environments is often bespoke and not easily transferred between spaces (Müser and Fehling, 2022). This is a design challenge. Stone states "an intuitive interface between man and machine is one which requires little training ... and proffers a working style most like that used by the human being to interact with environments and objects in his day-to-day life" (Stone, 1993). Section III discusses in greater depth the proposed interface for streamlining the creation of immersive visuals. Standardized inputs, such as equirectangular imagery, as well as access to massive online crowd-sourced datasets such as Google Street View allow users to contribute content without prior tailoring. Such an interface should be accessible via those devices which users are most comfortable and familiar with, their own personal devices, to reduce any necessary training.

Towards these goals, this research creates walkable room simulations over an extended listening area for student demonstrations to meet the following design objectives:

1. Create walkable auralizations to experience changes in room acoustics over distance for a classroom size group

2. Simulate congruent audio-visual experiences of concert halls and other venues

3. Provide students and instructors with the user interface to rapidly produce congruent visuals without prior tailoring or training

In the next section, the basic concept of our proposed auralization technique is discussed, followed by a discussion of the interface for the rapid creation of visuals. Analysis of the walkable auralizations follows, before a brief conclusion and outlook.

II. ACOUSTIC RAY-TRACING METHODOLOGY

Figure 3 shows the system architecture of the ray-tracing program. All input variables and data are shown at the bottom. The basis for auralizing a building or outdoor venue is a floor plan with a scale or known dimensions. Outdoor venues have to be approximated as LTI systems, as the program cannot compute temperature variations and wind conditions. A Deutsches Institut für Normung (DIN) database with wall and floor materials is utilized to determine absorption coefficients (Deutscher Normenausschuß, 1968). The user determines the locations of sound sources and receivers and provides sufficiently anechoic sound files for auralization. The receiver positions determine the locations of virtual microphones for the WFS loudspeaker locations, the location of a virtual dummy head (rendered through head-related transfer functions), or the position of a virtual spherical microphone array.

A. Floor-plan annotation

The first step in the auralization process is the annotation of floor plans. Standard floor plans such as those found on the world wide web can be processed as long as the user has information about its dimensions in the form of a scale or known overall dimension of the building (e.g., length and width). If needed, dimensions can also be estimated from satellite data or street maps. For the auralization preparation, the user marks up a floor plan using a bitmap editor (e.g., GIMP or Photoshop), annotating corners with red dots and columns with green dots. Further, the floor plan scale has to be annotated, and the original dimensions need to be handed over to the program as well, for example, 0 m and 5 m for two scale points. The auralization program automatically enumerates the corner positions and creates pairs of corners spanning wall elements. The user corrects and edits these pairs to finalize the digitized floor plan. Using this approach, a user can create a walkable auralialization of a large venue within a few hours – for example for the Cologne Cathedral as shown in Fig. 6.

The user can either assign a single wall-material ID to all walls or specify individual wall-material IDs for each wall – see Fig. 3, bottom, 2$^{\text{nd}}$-from-left box. The wall material data taken from the DIN database contains a short description of the material, frequency-specific absorption coefficients and Finite-Impulse-Response (FIR) filter coefficients. The latter is needed to compute adequate room impulse responses.

B. Ray-tracing algorithm

For each sound-source position, the program sends out rays at equidistant azimuth angles covering the full horizontal plane (360° circle) – see Fig. 3, 3$^{\text{rd}}$-left bottom box. Intersection points are computed for each ray and boundary object using the solution of the line–line intersection problem in Euclidean geometry. For each ray, the closest boundary intersection is determined. At the intersections, the reflection angle is calculated using Ibn Sahl/Snell's law (Rashed, 1990; Zghal *et al.*, 2007), which predicts that the reflected angle measured from the normal of a plane surface equals the incoming angle: $\cos(\alpha_r) = \cos(\alpha_i)$, and the next order ray is sent out into the new direction until the maximum order (e.g., the number of reflections) specified by the user is reached or the ray exits the floor plan – see Fig. 4. The outgoing rays are stored as a sequence of ray elements containing the intersection points and the boundary material identifiers. Since the initial angles are stored with the rays, a source-specific directivity pattern can be simulated after all rays have been traced.

1. Creating a room impulse response

Next, the rays are collected by receivers in the module *impulse response (IR) estimation* – see Fig. 3, rightmost bottom box. Receivers can be located anywhere in the rendered room. For WFS, a virtual array is

FIG. 1. The Collaborative-Research Augmented Immersive Virtual Environment (CRAIVE-Lab) at Rensselaer is equipped with a nearly-360° front-projection screen. It is microperforated so as to remain acoustically transparent to the dense 128-loudspeaker array located behind it. These systems together present congruent audio-visual content and are utilized for producing walkable auralizations for the purpose of experiential acoustics education.

placed into the floor plan section that should be auralized. The receivers (virtual microphones) are positioned at the WFS loudspeaker positions. So for a 128-channel loudspeaker system, 128 receivers are placed at the loudspeaker positions. For each receiver, a separate impulse response is created. To catch the rays, a virtual circle with an adjustable diameter is positioned at each receiver location. Then, the algorithm calculates which ray elements intersect the circle, and for all positive cases, the total ray distance between sound source and receiver is calculated. All listed values are stored together with the final angle of incidence, the angle the ray was initially sent out, the reflection order, and the sequence of wall identifiers. Based on these data, the impulse response is computed. The direct sound and reflections are computed as delta peaks at the delays that correspond to the rays' path length from the source to the receiver. In addition, each impulse is transformed the following way – see also Fig. 5:

1. The intensity of the ray is reduced based on the inverse-square law. Consequently, the sound pressure magnitude decreases with $1/r$.

2. The high frequencies are filtered out based on dissipation effects in air.

3. The absorption effects of the walls and other boundaries are simulated using a cascaded Finite Impulse Response (FIR) filter. The material-specific filters are chosen from the DIN database. The number of cascaded filters matches the order of the reflection.

4. In the final step, incoming direct sound and reflections are selected by their close proximity passing a receiver position within the virtual circle that was placed around it. The virtual receivers are directionally sensitive using the positive lobe of a figure-of-eight pattern. The latter points outward the array to avoid echo artifacts resulting from rays passing through the array at the opposite end of the array. An overlap-add method ensures that the delayed reflections can partially overlap, which can happen because of the cascading FIR filters' prolonged effect that simulates the wall reflections.

2. Simulation of late reverberation

Late diffuse reverberation is computed in addition to the early reflections generated by the ray-tracing model. This process is described in Fig. 3 (second and third rows from bottom). In order to calculate the frequency-specific reverberation times via the Sabine equation, we first need to calculate the volume as well as the individual areas of walls, floor, and ceiling. Alternatively, the reverberation time can be calculated using the Eyring equation or other methods such as those derived for es-

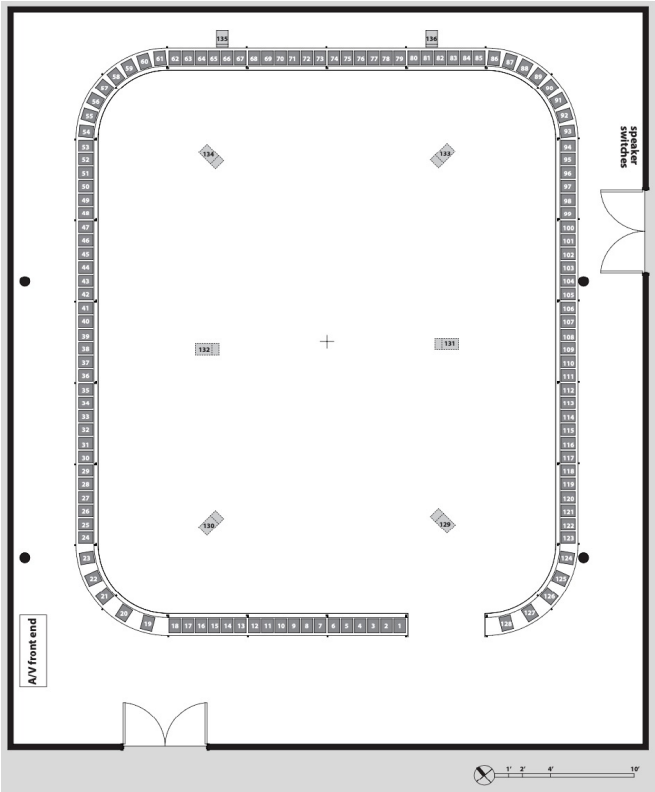

FIG. 2. Floorplan of the CRAIVE-Lab, measuring 12 m × 10 m. Depicted here is the dense horizontal 128-loudspeaker array and the six additional ceiling speakers.

timating the double-decay reverberation slope for coupled volumes (Bradley and Wang, 2010). For the Sabine equation, the average height of the enclosure has to be calculated or estimated—see Fig. 3, bottom/left. The floor area, A_F, (center-left box in Fig. 3) is calculated from the annotated floor plan and multiplied with the average height, h, to estimate the volume: $V = A_F \cdot h$. (top-left box in Fig. 3). The wall areas are estimated as the product of the average height and wall-element length—see Fig. 3, top, 2$^{\text{nd}}$-left box.

Using these dimensions, as well as the absorption coefficients the frequency-specific reverberation times are computed via the Sabine equation:

$$T_{60} = 0.161 \cdot V/(A + 4m \cdot V)\text{s}. \tag{1}$$

The total absorption, A, is defined as the sum of all surface elements, S_n, multiplied with their specific absorption coefficient, α_n:

$$A = \left(\sum_{k=1}^{K} \alpha_k \cdot S_k \right). \tag{2}$$

Since the late reverberation tail is formed by a stochastic process with an underlying Gaussian distribution, the fine structure of the simulated reverb tail is constructed from a Gaussian noise sample. The duration of the Gaussian noise sample is adjusted to twice the value of the maximum reverberation time. Next, the noise sample is processed through an octave-band filter bank with nine bands. An exponentially-decaying window, y_k, adjusted to the frequency-specific reverberation time, is calculated for each octave band, k:

$$y_k = e^{\frac{-t \cdot 20 \cdot \log(10)}{T_{60,k} \cdot 60}}, \tag{3}$$

with the reverberation time, $T_{60,k}$, in the k^{th} frequency band, and the time, t, in seconds. Afterwards, the total exponentially-decaying noise signal, x_t, is reassembled by summing the octave-filtered noise samples multiplied with the exponentially decaying window sample by sample:

$$x_t = \sum_{k=1}^{K} x_k \cdot y_k. \tag{4}$$

The process is repeated for each loudspeaker channel using independent Gaussian noise samples for each channel while keeping all other parameters constant. It has been shown that the described method benefits from having a large number of independent reverberation channels (Agrawal and Braasch, 2018).

The absorptive surface area is calculated from the ray-tracing model's wall elements; each multiplied with the average height. The frequency-specific absorption coefficients are determined through the values stored in the DIN database via the wall material identifiers. A linear onset ramp is calculated to gradually blend in the late reverberation tail with the direct sound and early reflections. The onset ramps' start and end points can be adjusted by the user.

Two methods are available to calculate the direct-to-reverberant energy ratio. The first method estimates the critical distance, the distance from the sound source at which the sound pressure levels of the direct sound match the sound pressure level of the reverberant field. For an omnidirectional sound source, the critical distance can be calculated using this equation (Kuttruff, 2000, p. 317):

$$r_c = 0.057 \cdot \sqrt{\frac{\gamma V}{T_{60}}}, \tag{5}$$

with the volume, V, the reverberation time, T_{60}, and the directivity factor, γ. For an omnidirectional sound source that we will assume in the subsequent calculation, the directivity factor is one.

In the next step, the impulse response is calculated at a receiver position at the critical distance. The overall energy of the impulse response, E_T, is the sum of the direct sound energy, E_D, the early reflection energy, E_E, as well as the late reverberant energy, E_L:

$$E_T = E_D + E_E + E_L. \tag{6}$$

At the critical distance, the following condition has to be met for an omnidirectional source and receiver pair:

$$E_D = E_E + E_L. \tag{7}$$

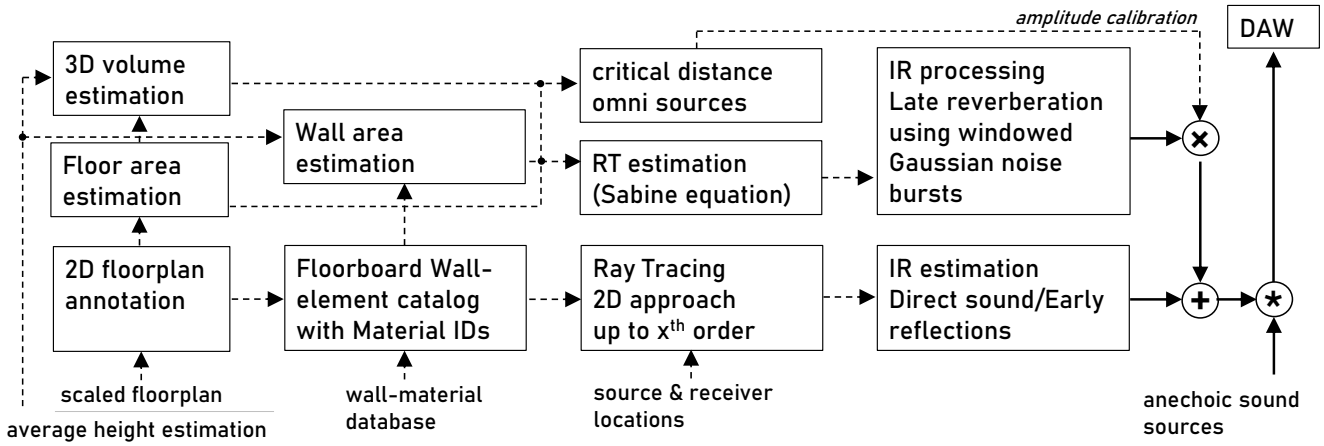

FIG. 3. System architecture for the auralization algorithm.

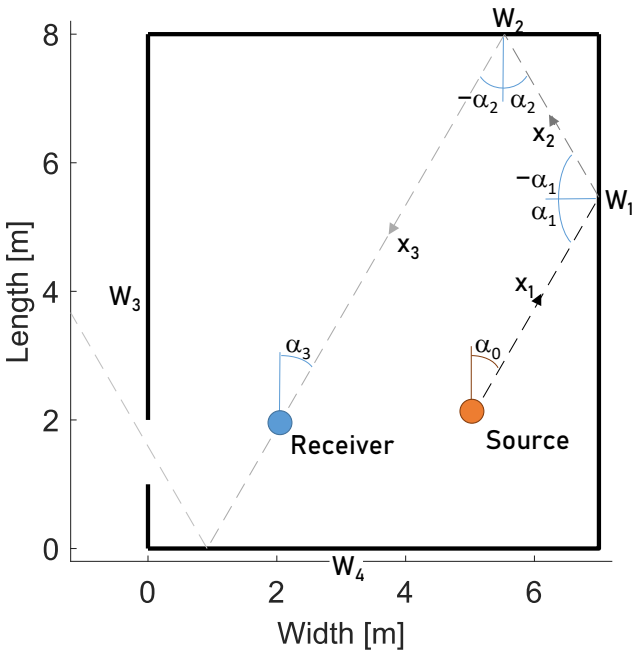

FIG. 4. Example of ray-tracing pathway.

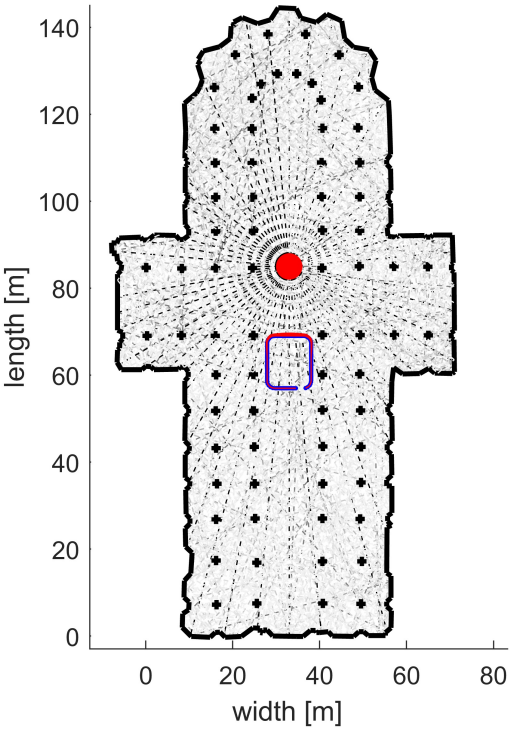

FIG. 6. Raytracing simulation of the cathedral in Cologne based on an annotated floor plan.

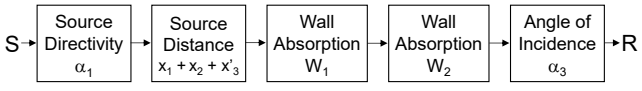

FIG. 5. Ray-tracing signal flow.

Consequently, the energy of the late reverberation has to be adjusted to:

$$E_L = E_D - E_E, \tag{8}$$

$$\sum_{t=0}^{2T_{60}} p_L^2 = \sum_{t=0}^{2T_{60}} p_D^2 - \sum_{t=0}^{2T_{60}} p_E^2, \tag{9}$$

with the sound pressure, p, which is, of course, proportional to the digital signal amplitude.

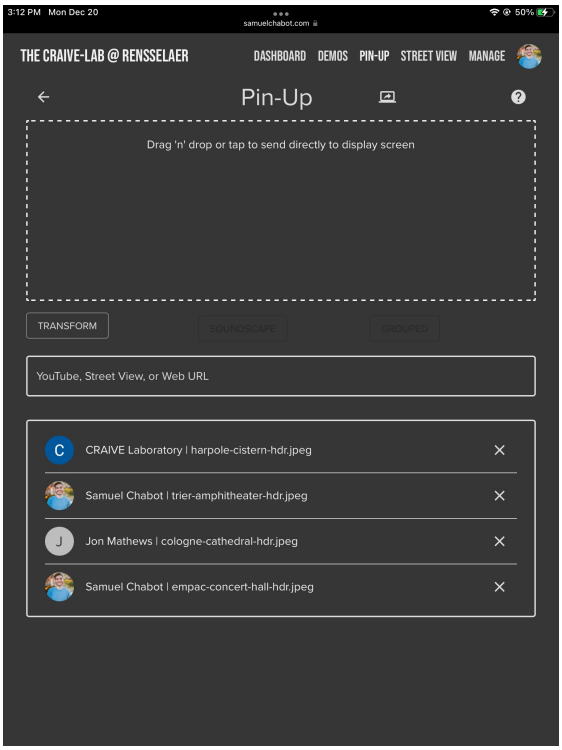

FIG. 7. The web interface for automatic formatting for and transmission to the immersive display. Students and instructors can access this from the devices they are most comfortable with, their own personal ones, as is shown here from a tablet view.

In the second method, the exponentially-decaying amplitude of the reverberation tail is fitted to the exponentially-decaying amplitudes of the reflection pattern. For this purpose, both signals are logarithmized so the decaying impulse response can be fitted by a linear regression curve. The amplitudes of the decaying slope are then matched, and a cross-fade method is used to blend out the early reflections while gradually blending in the late reverberation.

The CRAIVE-Lab's spatial aliasing frequency of 900 Hz limits the faithful spatial reproduction of soundfields to below this frequency. To mitigate this effect, the impulse responses can be generated with higher-order ambisonics using a virtual spherical receiver that is placed in the center of the CRAIVE-Lab position for the auralization. The impulse responses can be combined using the WFS method at and below the spatial antialiasing frequency and HoA above. It is also an interesting demonstration for students to compare the outcome of the WFS, HoA, and hybrid simulations that can be processed using the same annotated floor plan.

III. RAPID CREATION OF CONGRUENT VISUALS

Rapid prototyping is highly necessary in the CRAIVE-Lab; immersive experiences live and die by available content. Especially in the context of education, content must be created on a quick and affordable basis to keep up with the curriculum. This section describes the intuitive interface by which instructors and students can create visuals nearly instantly which are properly formatted for the display and require no tailoring or training on the user's part. The immersive display offers the opportunity for the presentation of visuals congruently with the corresponding synthesized auralization. Previous research has unveiled the importance of visual representations of concert venues for the acoustic quality judgement of these (Cabrera *et al.*, 2004; Chen and Cabrera, 2021; Jeon *et al.*, 2005; Maempel and Horn, 2018; Valente and Braasch, 2008, 2010). As a single continuous PC desktop, imagery should adhere to the ultrawide dimensions and aspect ratio of the screen: 15360 pixels × 1200 pixels and 12.8:1, respectively. However, there are additional considerations. Due to the blended overlap between projectors, the system renders the horizontal 15360 pixels over the footprint of 11636 square pixels, introducing a squeezing effect. Additionally, the nonuniform nature of the rectangular screen means the distance from the center of projection to the screen is not equal at all points across a horizontal cross-section. This introduces distortion to equirectangular imagery, requiring a perspectival transformation. Counteracting these forms of distortion is necessary to maintain the audio-visual congruency; the possible deviations of up to 200 pixels translate into over half a meter of discrepancy on-screen.

Asking a user to understand and apply these corrections on their own is an unnecessary hindrance. Further, the transferring itself of corrected content to the display machine creates additional hurdles: users unevenly resort to email, flash drives, or online file transfers. As tedious, confusing, and universal considerations, the transfer process and necessary corrections are prime candidates for programmatic abstraction.

To facilitate a smoother and more rapid content creation experience, a browser-based, device-agnostic interface has been developed. Importantly, users can access this interface via the devices they are most comfortable with: their own personal ones. It should be noted that immersive systems cannot be easily operated through a separate control room, as is standard, for example, for sound recording studios; operators should be present and immersed as well. System operation needs to work tetherlessly and/or through personal devices because anything larger would be counterproductive to the immersive experience. For the developed interface, a simple web page provides familiar drag-and-drop upload functionality, shown in Fig. 7. Users can submit panoramic imagery in standard equirectangular format. Upon doing so, the image is passed to the backend Node.js server via the RESTful POST method. Here, the image is fed through Python scripts loaded with the corrections and geometrically-determined perspectival transform. While this perspectival transform is made with respect to the center of the room, the produced effect is still an improvement across the space: audio remains more congruently

FIG. 8. A standard equirectangular image, above, before it is processed at the back-end through the required perspectival transform and horizontal stretch, below. This correction counteracts any introduced distortion due to the screen's irregular geometry (i.e. rectangular with round corners rather than cylindrical).

aligned with on-screen visuals and the inward bulging effect is removed. Figure 8 shows an example of a standard equirectangular image and its corresponding transformed image. After this step, the image is forwarded via a WebSocket connection directly to the display for presentation. This method requires no tailoring or training, asks nothing unreasonable from even the least-technical users, and results in nearly instantaneous display of content. As well as providing the upload functionality, the interface also lists all available previously uploaded images for the current session. Users can quickly toggle between each available image on the display. The complete signal flow from personal devices to backend network transformations to final display output is shown in the diagram in Fig. 9.

A. Crowd-Sourced Online Imagery

While the ability for users to submit their own imagery is a helpful and necessary feature, the capturing and subsequent processing of these photos can be time-consuming and geographically limiting: through this method, users must travel to the desired locations, snap the appropriate photos, and stitch them together using the appropriate software. There exist innumerable panoramic images in user-contributed online datasets such as Google Street View. However, these images also require special consideration for proper presentation on the immersive display. Applications such as Street View expect a flat plane display (i.e. a single desktop monitor), and consequently only display a limited field of view. Opening these in a fullscreen window stretches and distorts the incomplete horizontal field-of-view across the entire display. In order to view the entirety of a panorama, the Street View Static API of the publicly available Google Maps Platform is utilized and built into the web interface. This application programming interface receives RESTful GET requests for, and subsequently returns, specified imagery at explicit fields-of-view and compass headings. For users, this simply requires the submission to the interface (shown in Fig. 7) of the URL of their desired panorama. The interface parses the embedded unique panorama ID from the URL. This ID is paired with compass headings (in degrees) and the two are used as parameters in requests to the API. The returned images are then positioned at their respective location on the screen to present the complete field of view. This panorama is then added to the running list of those available in the current session. This method vastly increases the imagery available, offering instant and effortless access to countless locations.

IV. EXAMPLES AND EVALUATION

This section deals with concrete auralization examples that were established using the ray-tracing method described in the previous sections.

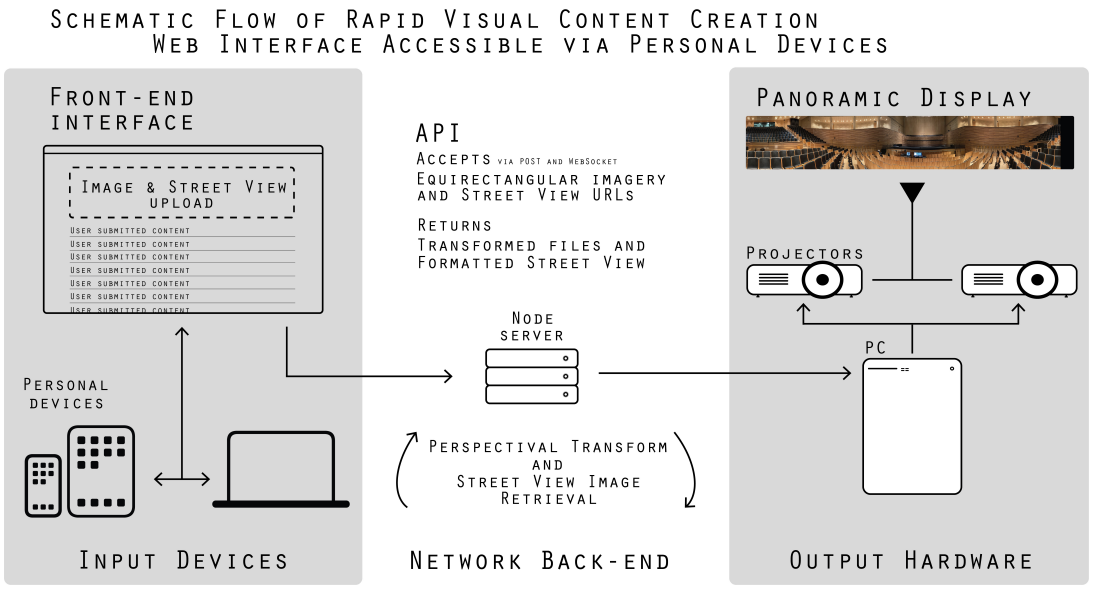

FIG. 9. The signal flow for rapid visual content creation for the immersive display. Content is uploaded at a web interface accessed via personal devices. The backend Node server receives this content and processes the perspectival transform or retrieves the request Street View imagery. The then properly formatted content is automatically transmitted to the display for near-instant presentation.

A. Haydnsaal, Eisenstadt

The first example is a recreation of the *Great Hall of Schloss Esterházy* (now Haydnsaal), where Joseph Haydn inaugurated several of his symphonies. The hall is located in Eisenstadt (now Austria) at the seat of the Esterházy family that Haydn worked for. The hall was built during the baroque building phase of the castle (1663–1672). The floor plan was scanned and manually annotated from Jürgen Meyer's paper on Haydn's concert venues (Meyer, 1978), which also graphed the reverberation times for the unoccupied hall, approximately:

f	0.125	0.25	0.5	1	2	4	8
RT	3.5	3.2	3.3	3.25	2.7	2.0	1.4

The virtual receivers for the WFS loudspeaker locations are depicted in blue in Fig. 10. A virtual sound source was positioned at the top left corner, indicated by the red-filled circle. The thin black lines show the simulated rays which are sent out in equirectangular directions.

Figure 11 demonstrates the effectiveness of the walkable-auralization design of the CRAIVE-Lab. The left graph shows the outline of the CRAIVE-Lab loudspeaker array (blue dots) within the floorplan of the Eisenstadt concert venue. The red dot shows the position of the virtual sound source sending out rays indicated by thin, solid lines. The surf plot within the loudspeaker array shows the relative sound pressure levels of the direct signal which, as expected for a point source, decreases with distance with the inverse square law for intensity or $1/r$ law for sound pressure – see also the color bar at the very right of the figure. In contrast to the direct signal,

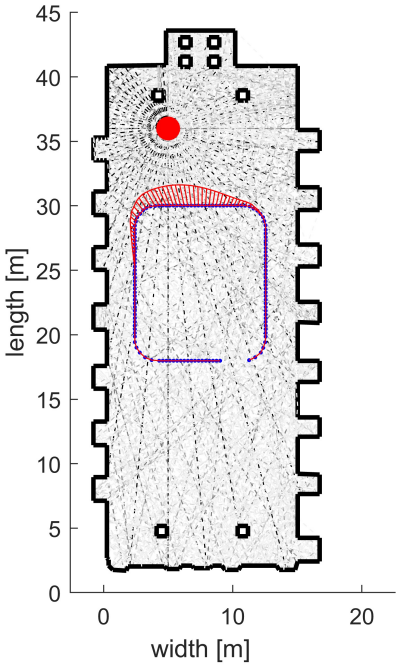

FIG. 10. Raytracing simulation of the Haydnsaal in Eisenstadt based on an annotated floor plan.

the sound pressure level of the diffuse reverberant sound field will remain steady for a typical concert hall. This is the case for the diffuse reverberant field created by 128

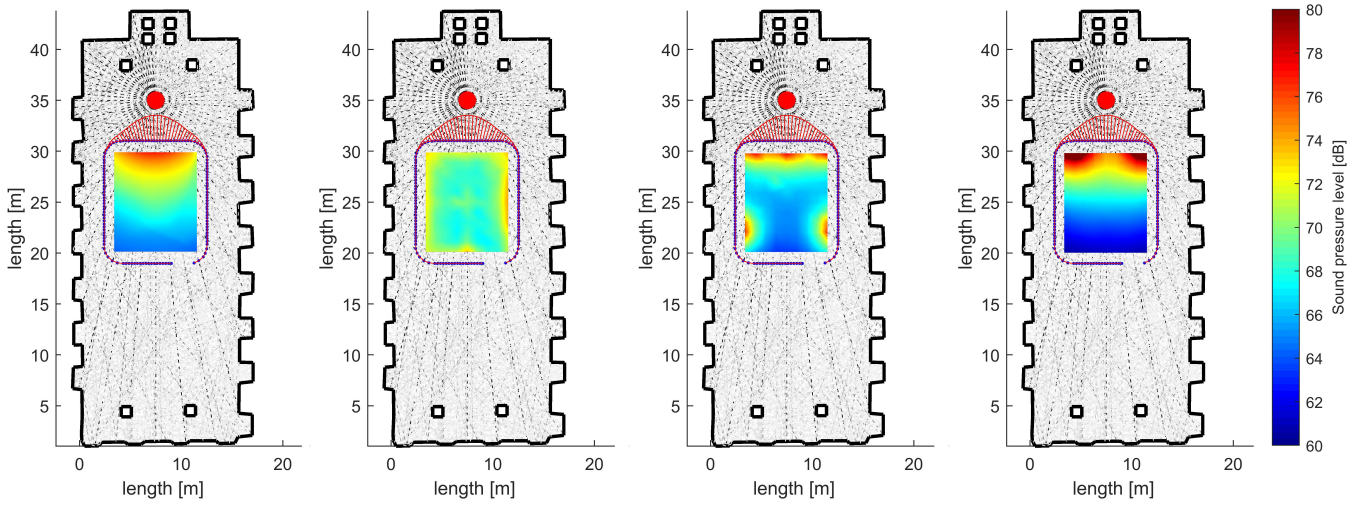

FIG. 11. Spatial simulation results for the Haydnsaal in Eisenstadt for a single sound source are shown in red. The individual graphs show the calculated sound pressure levels within the 128-channel CRAIVE-Lab array, which is virtually positioned within the Haydnsaal, from left to right: direct signal simulated using WFS, reverberant sound field created with 128 loudspeakers, reverberant sound field created with a 5-channel surround subset 0° , ±30°, ±120° azimuth, and reverberant sound field created with a 2-channel stereo subset (±30° azimuth). The colorbar indicates the decibel changes for all four graphs.

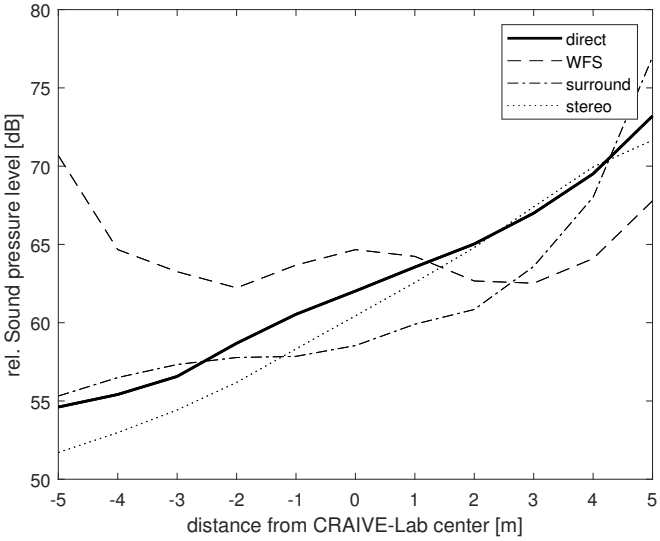

FIG. 12. Calculated sound pressure levels within the 128-channel CRAIVE-Lab array for a the diffuse reverberant sound field in the simulated Haydnsaal for the midline along the front-back direction as shown in Fig. 11. The legend indicates the individual cases for direct signal and reverberant sound field.

loudspeakers (Fig. 11, second graph from left). Over a large floor area of approximately 7 m × 9 m, the sound pressure level for the diffuse reverberation only varies by about 2 dB. This exemplifies the case made for the walkable design. The students can experience the change of the direct-to-reverberant signal levels by walking from front to back, including the position of the critical distance, where both levels cross each other.

Figure 12 shows this effect in more detail for the front/back midline of the loudspeaker array. In the depicted case, the sound pressure level of the direct signal (solid line) crosses the sound pressure level of the reverberant signal level (dashed line) 1 meter in front of the center. This way, the students can experience the drastic acoustical changes that occur when the direct signal level falls below the reverberant level for speech intelligibility. Using traditional stereophonic approaches will not provide a similar educational experience. The reverberant sound field will roll off like the direct signal for a traditional 2-channel loudspeaker set up (±30° azimuth measured from CRAIVE-Lab center) as shown in Figs. 11 and 12 (rightmost graph). The reverberant sound field provided by a 5.1-surround loudspeaker array has its own set of inconsistencies – see Figs. 11 and 12 (second from right). Like the classic stereo set up, it will roll off by over 10 dB on the midline but increase in sound pressure level when one of the surround loudspeakers is approached (which then will dominate the sound image, while still not providing the experience of a decorrelated sound field). The WFS simulation is much more consistent, with a variation of less than 2 dB within the center range from −4 to 4 m. The simulation in Fig. 12 was produced under the ideal assumption that the reverberant field is spatially diffuse and the levels are homogeneous. Research has shown that this is not always the case. Barron and Lee (1988) have measured that the level of the reverberant field drops by about −0.5 dB/10 m with distance from the sound source for most concert

halls, when using the 80-ms threshold from measured impulse responses. This effect is currently not implemented in the simulation. It can be assumed that the sound field is spatially more homogeneous, when a larger threshold is defined.

B. Acoustical Recreation of the Dan Harpole Cistern

The second auralization example focuses on an acoustical enclosure used for avant-garde music. The Dan Harpole Cistern was built in 1907 as an underwater tank for Fort Worden, WA, to provide water to extinguish fires that result from enemy attacks. Its substantial volume of 7,600 m^3 (2-million-US gallons) provides an anecdotally reported reverberation time of 45 seconds. The construction of Fort Worden occurred between 1898 and 1917. Together with Fort Flagler and Fort Casey, it served as a system of three forts called the *"The Triangle of Fire"* to defend the Pacific coast. It was named after Admiral John L. Worden and was an active military installation until 1953. It is now a recreational park for the State of Washington.

Interestingly, the underwater tank received its acoustical significance after its official purpose when the emptied tank was discovered by visiting artists. In June 2007, the tank was named the Dan Harpole Cistern to honor the name-giver for his efforts to obtain a designation as "Washington's Official Instrument" for the cistern (Herald staff, 2008). The Dan Harpole Cistern has been a recording venue for several landmark recordings, starting with an album from 1989 by Pauline Oliveros, Stuart Dempster, and Panaiotis called *Deep Listening*. The Dan Harpole Cistern has a diameter of 56.7 m and a floor-to-ceiling height of 4.3 m. It is made of concrete and contains 89 square columns (about 38 cm wide).

An impulse response of the cistern was recorded for the 1991 Deep Listening album: *The Ready Made Boomerang*. The following reverberation times were computed from analyzing this impulse response (frequency in kHz, reverberation times in seconds):

f	0.125	0.25	0.5	1	2	4	8
RT	63.2	41.7	28.8	18.5	12.5	5.7	2.4

The measured reverberation time is longer than 45 seconds at very low frequencies and much shorter at high frequencies. Circular buildings are often troublesome because the curved surfaces create focused echoes, especially for sound sources at the center. The large number of columns in the cistern mitigates this effect by creating a high level of diffusion. Jürgen Meyer notes the importance of columns in churches. He points out that, depending on the thickness of the column, its effectiveness as a reflector is limited to above 1000 Hz to 1500 Hz, but that complete bending around columns can be expected only below 200 Hz to 300 Hz (Meyer, 2009). This phenomenon can be demonstrated to students using this ray-tracing approach. Figure 13 shows the result of the ray-tracing process for the cistern. The left graph shows the reproduction with columns; the right graph shows the same simulation but without columns. Without the columns, the outer, circular wall will produce audible focusing effects, which can be heard in the walkable auralization as sharply localizable focal points.

In the future, we plan to extend the current raytracing algorithm to simulate edge diffraction (Calamia and Svensson, 2006; Tsingos *et al.*, 2001), which is an audible effect in room acoustics. It should also be noted that the physical width of the columns has an audible effect which is not faithfully simulated with raytracing.

The reflection of sound waves depends on their frequencies. Fresnel zones describe the minimum area needed to reflect sound off a surface, with increasing surface area required for lower frequencies. One approach to improve the accuracy of the acoustical demonstration is to employ the Boundary Element Method (BEM) (Kirkup, 2019; Vorländer, 2020) to calculate the wave field for low frequencies, which can be done as a 2-dimensional model using the students' floorplan-based models. Currently, we demonstrate two architectural models using wave-based computations that are implemented as part of Master's theses: Rensselaer's Concert Hall at EMPAC (Hochgraf, 2015) and Jazz at Lincoln Center's Dizzy's Club in New York City (Scott, 2019).

The EMPAC concert hall auralization works well in the CRAIVE-Lab, but the acoustics of the Dizzy's Club are too dry and are overshadowed by the CRAIVE-Lab's own reverberation (0.89 s, broadband). These and other small-venue auralizations can be better simulated in a small studio of RPI's acoustics program (4 m × 5 m floor area T_{30}=0.160 s, 1 kHz), which is equipped with a 20-channel dome-style spherical loudspeaker array for HoA renderings. In the foreseeable future, we plan to retrofit the CRAIVE-Lab enclosing space to reduce the reverberation time substantially.

Based on the geometrical data, the volume was estimated to be 10,770 m^3. The value is larger than the volume of 7,600 m^3 (2-million US gallons) provided in the literature. However, it is not clear if the cistern was fully filled with water or if the 2-million US gallons are a rough estimate. The following values were obtained from the geometrical model using the Sabine equation and absorption coefficients for smooth unpainted concrete (Vorländer, 2020).

f	0.125	0.25	0.5	1	2	4	8
RT	27.5	27.5	13.7	13.7	13.7	5.5	5.5

These critical distances were estimated in meters:

f	0.125	0.25	0.5	1	2	4	8
d_c	1.13	1.13	1.59	1.59	1.59	2.52	2.52

Since the estimated reverberation times are too short in the low-frequency range, absorption coefficients were computed for 250 Hz and 1000 Hz. For 1000 Hz, an absorption coefficient of 0.015 produces a reverberation time of 18.3 seconds. For 250 Hz, an absorption value of 0.0066 results in a reverberation time of 41.6 seconds. These values correspond to critical distances of 0.91 m and 1.38 m. The simulated impulse response for the cistern with columns is plotted in Fig. 14.

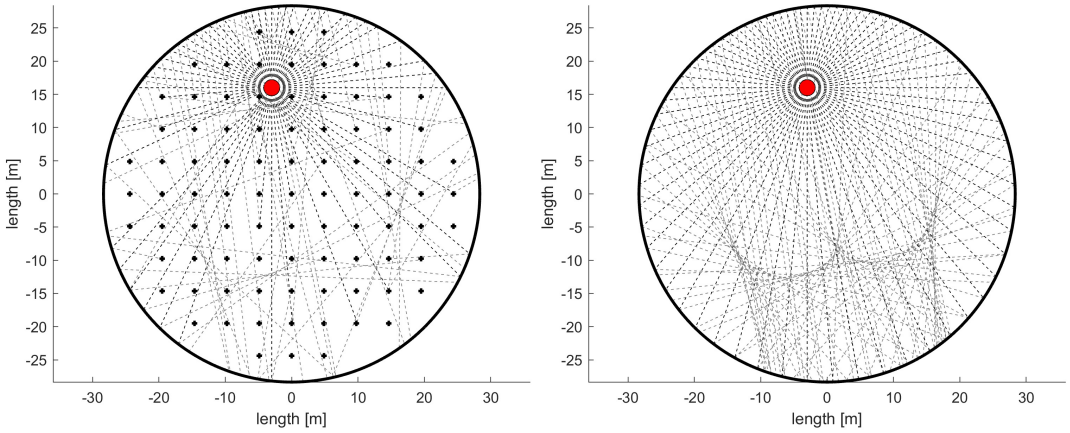

FIG. 13. Raytracing results for the Dan Harpole Cistern; left: with simulated columns, right: without simulated columns.

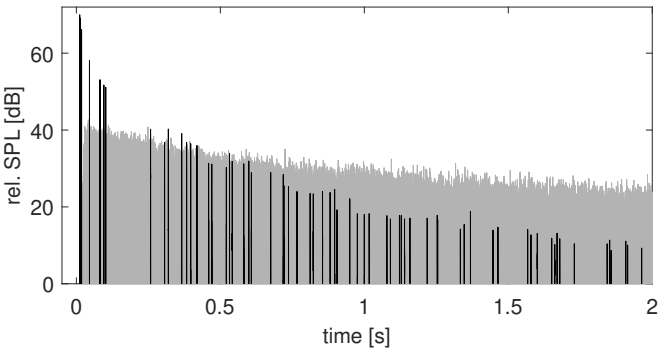

FIG. 14. Simulated impulse responses for the Dan Harpole Cistern. The black lines show the results for the direct sound and early reflection at one measurement position, 4 m from the sound source. The gray curve depicts the results for the late reverberation. Both curved were normalized via the critical distance at the 2-kHz octave band according to Equations 5 & 7.

V. CONCLUSION

Rapid auralization and visualization methods in an immersive classroom were proposed in this paper. Annotated floor plans and a ray-tracing method allow for rapid prototyping of acoustical enclosures, a process that usually requires only between one and two hours on a typical desktop computer. These walkable auralizations provide students of acoustics and disciplines beyond with an experiential method for learning acoustics and the consequences of room acoustic design. In the course *Aural Architecture*, students choose a site and create their own walkable auralizations. The web interface allows for the rapid creation or retrieval of context congruent visuals. Users can submit their own panoramic imagery or those available in online databases for automatic formatting and near-instant display on the screen. The library of simulated venues continues to grow and includes the Cologne Cathedral, Savoy Ballroom Jazz Club, Columbia 30^{th} Street Studio, St. Mark's cathedral in Venice, St. Patrick's Cathedral in New York City, Pantheon in Rome, and the Amphitheatre in Trier. Analyses of these auralizations performed with a binaural manikin validate the walkable nature of the experience.

VI. ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant IIS-1909229 and by the Cognitive and Immersive Systems Laboratory (CISL) and an RPI Brown's Fellowship.

Agrawal, S., and Braasch, J. (**2018**). "Impression of spatially distributed reverberation in multichannel audio reproduction," in *Audio Engineering Society Convention 145*, Audio Engineering Society, p. 10076, http://www.aes.org/e-lib/browse.cfm?elib=19802.

Anderson, B. E., Moser, B., and Gee, K. L. (**2012**). "Loudspeaker line array educational demonstration," J. Acoust. Soc. Am. **131**(3), 2394–2400, doi: 10.1121/1.3676723.

Barron, M., and Lee, L.-J. (**1988**). "Energy relations in concert auditoriums. I," The Journal of the Acoustical Society of America **84**(2), 618–628.

Berkhout, A. J. (**1988**). "A holographic approach to acoustic control," Journal of the audio engineering society **36**(12), 977–995.

Berkhout, A. J., de Vries, D., and Vogel, P. (**1993**). "Acoustic control by wave field synthesis," The Journal of the Acoustical Society of America **93**(5), 2764–2778.

Braasch, J. (**2018**). "Convolution, fourier analysis, cross-correlation and their interrelationship," in *Springer Handbook of Systematic Musicology*, edited by R. Bader (Springer, Cham, Switzerland), pp. 273–284.

Bradley, D. T., and Wang, L. M. (**2010**). "Optimum absorption and aperture parameters for realistic coupled volume spaces determined from computational analysis and subjective testing results," The Journal of the Acoustical Society of America **127**(1), 223–232.

Bricken, W. (**1990**). "Learning in Virtual Reality," Technical Report 9.

Cabrera, D., Nguyen, A., and Choi, Y. J. (**2004**). "Auditory versus visual spatial impression: a study of two auditoria," in *Proc. of ICAD*, Sydney, Australia.

Calamia, P. T., and Svensson, U. P. (**2006**). "Fast time-domain edge-diffraction calculations for interactive acoustic simulations," EURASIP Journal on Advances in Signal Processing **2007**, 1–10.

Carman, J. C. (**2012**). "Classroom demonstrations of acoustic beamforming," J. Acoust. Soc. Am. **131**(3), 2401–2404, doi: 10.1121/1.3677242.

Chabot, S., Drozdal, J., Peveler, M., Zhou, Y., Su, H., and Braasch, J. (**2020**). "A collaborative, immersive language learning environment using augmented panoramic imagery," in *Proc. of the Immersive Learning Res. Netw. Conf. 2020*, IEEE, San Luis Obispo, CA.

Chen, Y., and Cabrera, D. (**2021**). "The effect of concert hall color on preference and auditory perception," Applied Acoustics **171**, 107544.

Cruz-Neira, C., Sandin, D. J., and DeFanti, T. A. (**1993**). "Surround-screen projection-based virtual reality: The design and implementation of the cave," in *Proc. of the 20th Annu. Conf. on Comput. Graph. Interact. Techn.*, SIGGRAPH '93, ACM, New York, NY, USA, pp. 135–142, doi: doi.acm.org/10.1145/166117.166134.

Deutscher Normenausschuß, ed. (**1968**). *Schallabsorptionsgrad-Tabelle* (Beuth-Vertrieb GmbH, Berlin).

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., and Wenderoth, M. P. (**2014**). "Active learning increases student performance in science, engineering, and mathematics," Proc. Nat. Acad. of Sci. **111**(23), 8410–8415, doi: 10.1073/pnas.1319030111.

Fresnel, A. (**1819**). "Mémoire sur la diffraction de la lumière [memorandum on light diffraction]," Mémoires de l'Académie des Sciences, da p. 339 a p. 475: 1 tav. ft; AQ 21 339.

Herald staff (**2008**). "Music will float from cistern to the heavens," August 14, Herald Net, Everett Herald and Sound Publishing, Inc., https://www.heraldnet.com/life/music-will-float-from-cistern-to-the-heavens/, last retrieved May 1, 2020.

Hochgraf, K. (**2015**). "Auralization of concert hall acoustics using finite difference time domain methods and wave field synthesis," Master's thesis, Rensselaer Polytechnic Institute.

Hulsebos, E., de Vries, D., and Bourdillat, E. (**2002**). "Improved microphone array configurations for auralization of sound fields by wave-field synthesis," Journal of the Audio Engineering Society **50**(10), 779–790.

Huygens, C. (**1690**). *Traité de la lumière où sont expliquées les causes de ce qui luy arrive dans la reflexion, & dans la refraction, et particulierement dans l'etrange refraction du cristal d'Islande [Treatise on light in which are explained the causes of that which occurs in reflection, in refraction, and particularly in the strange refraction of Iceland spar]* (chez Pierre Vander Aa marchand libraire).

Jeon, J. Y., Kim, Y. H., Kim, S. Y., Cabrera, D., and Bassett, J. (**2005**). "The effects of visual input on the evaluation of the acoustics in an opera house," in *Forum Acusticum*, pp. 2285–2289.

Kirkup, S. (**2019**). "The boundary element method in acoustics: A survey," Applied Sciences **9**(8), 1642.

Kuttruff, H. (**2000**). *Room Acoustics* (Spon Press, London, New York).

Lave, J., and Wenger, E. (**1991**). *Situated Learning* (Cambridge Univ. Press, New York, NY, USA).

Limniou, M., Roberts, D., and Papadopoulos, N. (**2008**). "Full immersive virtual environment CAVE in chemistry education," Computers and Education **51**(2), 584–593.

Maempel, H.-J., and Horn, M. (**2018**). "The Virtual Concert Hall - A Research Tool for the Experimental Investigation of Audio-visual Room Perception," International Journal on Stereo & Immersive Media **1**(1), 79–98.

Meyer, J. (**1978**). "Raumakustik und orchesterklang in den konzertsälen joseph haydns," Acta Acustica united with Acustica **41**(3), 145–162.

Meyer, J. (**2009**). *Acoustics and the Performance of Music*, 5th ed. (Springer, New York, NY).

Murphy, D. T. (**2006**). "Archaeological acoustic space measurement for convolution reverberation and auralization applications," Proc. 9th Int. Conf. Digit. Audio Effects, DAFx 2006 221–226.

Müser, S., and Fehling, C. D. (**2022**). "AR/VR.nrw–Augmented und Virtual Reality in der Hochschullehre," HMD Praxis der Wirtschaftsinformatik **59**(1), 122–141, doi: 10.1365/s40702-021-00815-y.

Neilsen, T. B., Strong, W. J., Anderson, B. E., Gee, K. L., Sommerfeldt, S. D., and Leishman, T. W. (**2012**). "Creating an active-learning environment in an introductory acoustics course," J. Acoust. Soc. Am. **131**(3), 2500–2509, doi: 10.1121/1.3676733.

Pelzer, S., Aspöck, L., Schröder, D., and Vorländer, M. (**2014**). "Integrating real-time room acoustics simulation into a CAD modeling software to enhance the architectural design process," Buildings **4**(2), 113–138, doi: 10.3390/buildings4020113.

Rashed, R. (**1990**). "A pioneer in anaclastics: Ibn Sahl on burning mirrors and lenses," Isis **81**(3), 464–491.

Scott, E. E. (**2019**). "Simulated impulse responses of small rooms using hybrid methods: optimizing crossover frequency," Master's thesis, Rensselaer Polytechnic Institute.

Spors, S., Rabenstein, R., and Ahrens, J. (**2008**). "The theory of wave field synthesis revisited," in *In 124th Convention of the AES*, Citeseer.

Stein, D. (**1998**). "Situated learning in adult education," Technical Report ED418250 (ERIC, Columbus, OH).

Steinberg, J. C., and Snow, W. B. (**1934**). "Auditory perspective – physical factors," Electrical Engineering **Jan**, 12–17.

Stone, R. (**1993**). "Virtual Reality: A Tool for Telepresence and Human Factors Research," in *Virtual Reality Systems*, edited by R. A. Earnshaw, M. Gigante, and H. Jones, Academic Press, London, p. 183.

Tsingos, N., Funkhouser, T., Ngan, A., and Carlbom, I. (**2001**). "Modeling acoustics in virtual environments using the uniform theory of diffraction," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 545–552.

Valente, D. L., and Braasch, J. (**2008**). "Subjective expectation adjustments of early-to-late reverberant energy ratio and reverberation time to match visual environmental cues of a musical performance," Acta Acustica United with Acustica **94**, 840–855.

Valente, D. L., and Braasch, J. (**2010**). "Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues," The Journal of the Acoustical Society of America **128**(4), 1952–1964.

Verheijen, E. (**1997**). "Sound reproduction by wave field synthesis," Ph.D. thesis, Technical University Delft, The Netherlands.

Vorländer, M. (**2020**). *Auralization* (Springer).

Wierstorf, H., Raake, A., and Spors, S. (**2017**). "Assessing localization accuracy in sound field synthesisa)," J. Acoust. Soc. Am. **141**(2), 1111–1119, http://dx.doi.org/10.1121/1.4976061, doi: 10.1121/1.4976061.

Zghal, M., Bouali, H.-E., Lakhdar, Z. B., and Hamam, H. (**2007**). "The first steps for learning optics: Ibn Sahl's, Al-Haytham's and Young's works on refraction as typical examples," in *Education and Training in Optics and Photonics*, Optical Society of America, p. ESB2.

Ziemer, T. (**2018**). "Wave field synthesis," in *Springer handbook of systematic musicology* (Springer), pp. 329–347.