WYSAHIWYG (What You See And Hear Is What You Get): Learning from photocartography in mapping the cross-modal features of the soundscape

J. Parkman CARTER[1]; Jonas BRAASCH[2]

[1,2] Rensselaer Polytechnic Institute, USA

ABSTRACT

"Measuring" the soundscape is like trying to measure the ocean, the forest, or the city—one dimensional metrics and color-coded maps are inadequate to meaningfully describe the myriad complex conditions, relationships, and effects which comprise the soundscape's impact upon us, and our impact upon it. In order to better understand the subjective experience of the soundscape, we need new methods to capture and represent the multisensory extents of the sonic environment. Much can be learned from the recent revolution in photocartography—which uses various forms of satellite and street view photography to generate interactive online maps. Photocartography is the new WYSIWYG (What You See Is What You Get) interface, and has rapidly proven to be more data-rich and more broadly accessible than conventional symbolic maps. This project proposes an interdisciplinary method to integrate multi-directional Ambisonic audio recordings of an environment with high dynamic range (HDR) spherical panoramic photography in the form of interactive maps and virtual tours. Case studies from both urban and park environments will demonstrate how this cross-modal mapping method can capture and reveal specific soundscape phenomena which elude conventional acoustic measurements.

Keywords: Soundscape, Cross-modal, Mapping, Aural Ecology
I-INCE Classification of Subject Numbers: 56.3, 68.3, 76.9, 79.9

1. INTRODUCTION

In order to improve our understanding of the impacts from and upon the sonic environment—or "soundscape"—we require a new means of representing *in situ* data that is full-range, multivalent, and legible to a wide variety of stakeholders (e.g. non-acousticians). Current mapping of the soundscape is primarily confined to "noise-mapping", which tends to rely on predictive models representing one-dimensional metrics with old modes of cartographic abstraction (e.g. color-coding), and simply does not convey a soundscape's temporal, spectral, or, most importantly, its contextual complexities. Multisensory environmental context, after all, is largely responsible for determining whether a particular sound is welcome or not.

WYSIWYG refers to a representational approach where forms of working abstraction are avoided in favor of direct simulations of the end-product. The classic example was popularized when word processing applications in the 1980s leveraged developments in programming and computer graphics to display on screen a reliable simulation of the printed page. Such intuitive interaction via "what you see is what you get" is largely credited with the rapid growth of desktop publishing (1) and even the popular adoption of personal computing in general. Today WYSIWYG interfaces are common across a variety of media-based applications, where the power and efficiency of constructing and/or evaluating content directly in its final form (whether for print, web, audio/video, etc.) enables content creators and consumers alike to enjoy media literacy and transparency while minimizing the obscuring effects of abstraction (e.g. computer code, or misleading quantifications).

Cartography has a long history of utilizing graphic abstractions and symbols to render geographical environments. Such representation has evolved significantly in recent years, however, with the advent

[1] cartej8@rpi.edu
[2] braasj@rpi.edu

of extensive satellite and interactive "street view" photography supplanting the conventional, symbol-based abstractions of static cartography. But these extensive advancements have only evolved our representation of the visual environment. How might acousticians and soundscape researchers achieve this level of immersive and interactive documentation of our diverse sonic environments? Could we not supplement increasingly popular user-navigated spherical panoramic photography with an associated layer of steerable directional audio? Could we not integrate the processes of recording such audio and photocartography?

In order to facilitate productive discussion among acousticians, ecologists, policy-makers, designers, and citizens, this project develops a WYSAHIWYG (What You See *And Hear* Is What You Get) method to document, integrate, and represent the full-range aural and visual fields. This method produces interactive documents that are simple to access and navigate, providing opportunities for soundscape analysis and discussion among acousticians and non-acousticians alike. By placing emphasis on design representation tools and applications, this method encourages designers and planners—those who often inadvertently wield disproportionate influence over the sonic environment and our experience of it—to incorporate soundscape considerations into their decision-making processes.

Several current issues would benefit from improved soundscape documentation: acoustic ecologists working to analyze and conserve endangered habitats; noise pollution policy-makers requiring more than the typical one-dimensional SPL metric to understand annoyance factors within communities; and sustainable designers and engineers working to define "acoustic comfort" for the purposes of evolving best design practices and building performance evaluation criteria.

2. PRECEDENTS & THEIR LIMITATIONS

Since being classified as a pollutant, noise has challenged acousticians to quantify it, resulting in a number of metrics which give some indication of equivalent sound pressure level exposure but do little to address the real issues of subjective and contextual annoyance (2). While "noise maps" produced by proprietary software such as SoundPLAN (3) are broadly legible (red = loud) and provide colorful visual aids in the noise pollution discussion, they tend to tell us little beyond what we already know (e.g. roadways are loud).

R. Murray Schafer's work on soundscape research attempted to shift the discussion from a negative one about noise to a positive one about aural awareness (4). In order to improve cataloging and analysis efforts, forms of visual notation and classification were needed to be able to compare different elements within a soundscape, as well as to compare different soundscapes. The abstract notation method Schafer developed relies on graphic abstractions to describe certain sonic attributes, but like other reductive and esoteric notation systems, his system leaves much room for improvement, given the capabilities of today's tools.

Brigitte Schulte-Fortkamp and Bennett M. Brooks are currently leading a focused effort to standardize soundscape research in order to correlate objective physical measures with subjective experience. Together they have co-chaired multiple ASA (Acoustical Society of America) sessions to establish the definitions, metrics, and protocols for soundscape applications. While the conceptual framework is to establish that soundscape requires more than physical and psychoacoustical definitions, a recent workshop (5) to generate a soundscape lexicon by collectively ranking different descriptors produced nebulous and inadequate results. If nothing else, the attempt to create such a reference lexicon confirms the need for improved soundscape representation which goes beyond numbers and words.

3. AN INTERDISCIPLINARY SOLUTION

Scientific method attempts to measure and predict physical and perceptual phenomena by isolating variables from the complexities of context in order to identify knowable elements and behaviors. Design process, on the other hand, attempts to understand and anticipate the irreducible complexities of context which act upon a site in order to produce appropriate responses in the built environment. It can be said, then, that "design science" is an oxymoron, given the distinctly different approaches to addressing the broad, interwoven impacts of context on a perceiver. Of the myriad different specialized branches of acoustics, soundscape research is perhaps the most difficult to approach solely scientifically, due to the relatively vast scale and complexity of the source, receiver, and environmental relationships involved. It can be argued that the present insufficiencies in soundscape documentation are a direct result of attempting to isolate too many variables from their fundamental contextual factors. What is needed is an approach which does not consider cross-modal (i.e. multisensory) effects as “interference”—as is typical in much scientific perception research—but rather recognizes that cogent interpretations of the soundscape must integrate different modes of perception and experience.

Traditionally and lamentably, architectural designers are far better trained in visual observations and communication than in matters of sound. Conversely, acousticians tend to use abstract data visualizations with little, if any, foundation in aesthetics or visual communication. The approach taken here aims to remedy the seemingly intractable communication divide between designers and acousticians by combining observation skills of both science and design, and by employing interdisciplinary tools in the construction of interactive, multisensory documents. Integrating audio capture with the accessibility of photography allows designers who are not otherwise incorporating soundscape into their research to promote better soundscape discussion with their clients and the public. An interdisciplinary approach is, therefore, a form of outreach and advocacy.

4. THE METHOD

The cross-modal soundscape mapping method developed in this project entails integrating Ambisonic multi-channel audio with high dynamic range (HDR) spherical panoramic photography, and making that integrated information legible and interactive. Procedures for capturing this data *in situ* and then processing it are summarized below, followed by an overview of techniques for generating interactive multi-media maps using commonly available (if atypical for soundscape research) software.

Figure 1 – Image of author’s cross-modal soundscape recording configuration comprised of Ambisonic audio capture and spherical panoramic photography.

4.1 Full-Range Photography

In audio recording, we take for granted that it is straightforward to have an omnidirectional microphone capsule capable of capturing nearly the entire dynamic range and frequency spectrum of sound experienced by humans over the entire spatial field. To replicate this in photography is a technical challenge, but recent developments in high dynamic range (HDR) digital photography and spherical panography (the seamless stitching of several photographs covering the complete visual field around a single point) enable us to produce the equivalent of a full-range "omni" photograph.

Because even professional DSLR camera sensors cannot capture the vast light-to-darkness range (contrast ratio) dynamically perceived by the human visual system, HDR photography relies on multiple exposures which are "tonemapped" in post-processing to generate an image which is variably and accurately exposed throughout the scene, approximating how the eye works to manage different light levels when scanning the environment (6). This project uses HDRsoft's Photomatix Pro software to perform HDR processing (7).

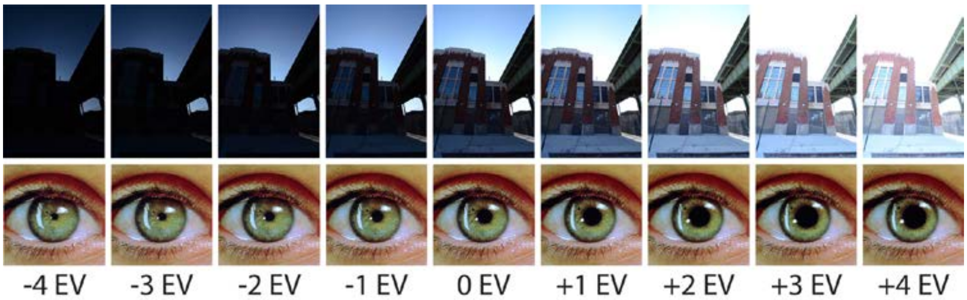

Figure 2 – Varying a camera's aperture is similar to the eye's dilation of the pupil to control the amount of light entering the visual system. While the eye does this dynamically in real-time, the limitations of a camera sensor require use of several different exposures to accurately expose all areas of a scene.

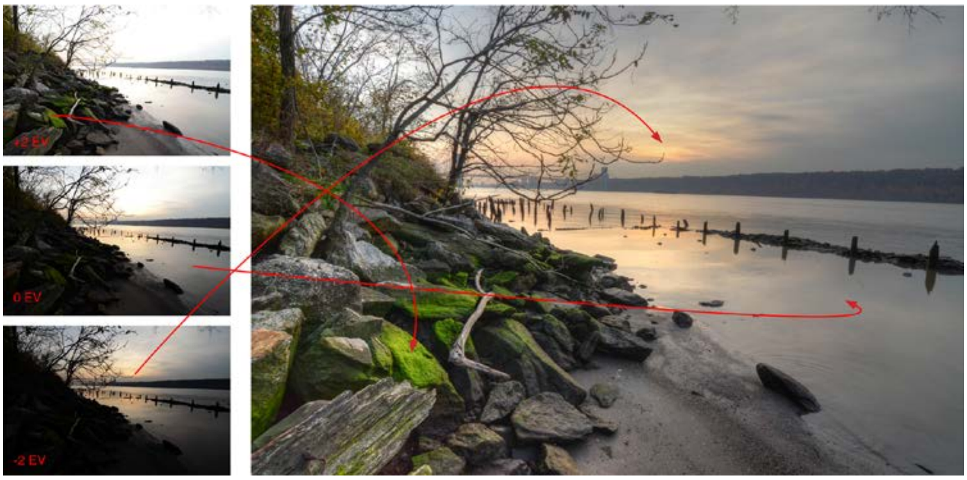

Figure 3 – In this simple example using three exposures, it is apparent how "tonemapping" algorithms choose appropriate exposure levels throughout the scene to generate a single image which approximates the eye's experience.

Spherical panoramas are similar to the popular technique of blending multiple overlapping photographs into a single image with a wider field of view than could be captured directly by typical photographic optical lenses. The distinction of the spherical panorama entails making certain to capture the entire visual field around a fixed point known as the "no-parallax point" (NPP) in order to achieve seamless stitching accuracy in post-production. To create an HDR spherical panorama typically requires 140-160 images at different exposures and orientations to cover the entire visible range in terms of light contrast and spatial extents. However, this can be a relatively rapid procedure in practice (5-10 minutes), aided by dedicated panorama tripod heads with fixed stops such as those by Nodal Ninja (8), and a remote control which can be programmed to quickly sequence the necessary exposures, such as the Promote Control (9).

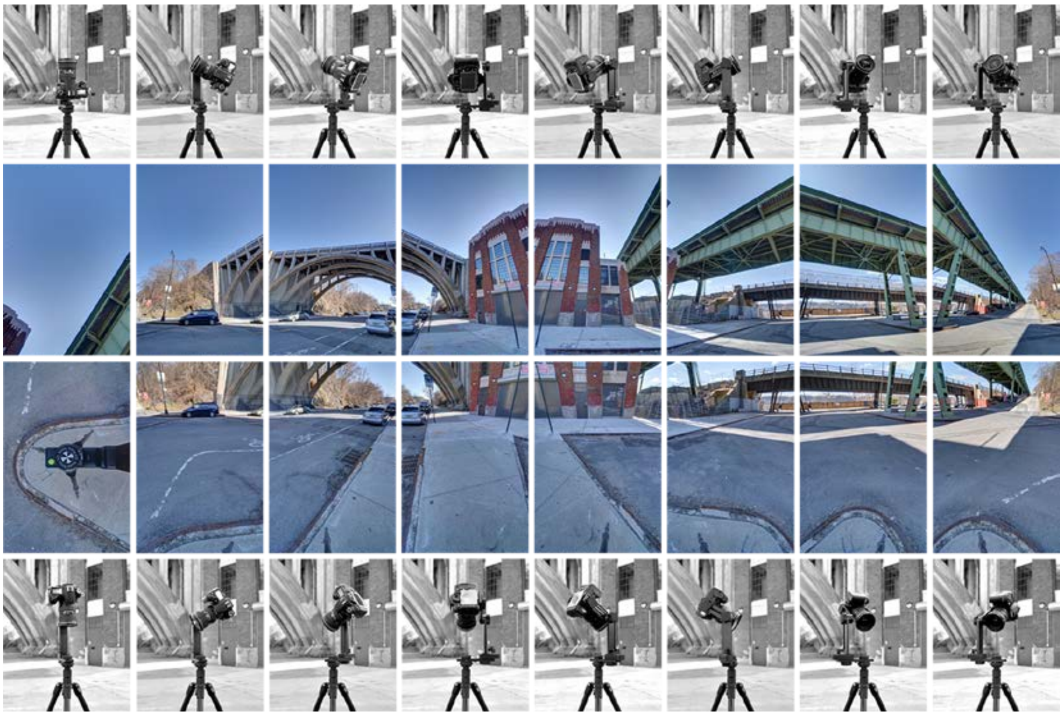

Figure 4 – Each camera position above corresponds to the resulting HDR image generated from that orientation. The separate orientations are stitched together to create a seamless projection of the entire visual field from a given location (see Figure 5 below).

There are a number of ways to project such an "omni" photograph. The most useful for our purposes is the 2:1 aspect ratio "equirectangular" projection, which is a flattened representation covering 360° in the horizontal dimension and 180° in the vertical dimension (from looking straight down at the nadir to looking straight up at the zenith). The equirectangular projection allows us to overlay an orthogonal grid representing angular degrees, which is useful for locating the directional Ambisonic audio files. It is also possible to generate an interactive "steerable" file in the Apple QuickTime Virtual Reality (QTVR) format, which is the basis of the virtual tours discussed in section 4.3 below. This project uses PTGui Pro for stitching panoramas and generating the various projections (10).

Figure 5 – The stitched HDR spherical panorama is shown here in the "equirectangular" projection, overlaid with an orthogonal grid for locating directional audio feeds relative to the visual field.

4.2 Full-Range Audio Recording

There are a number of ways to record environmental sound in the field; however, in order to maintain flexibility in post-processing, this project makes use of Ambisonic audio capture. It is beyond the scope of this paper to fully describe the methods of Ambisonic recording and processing; for our purposes, the key advantages are: 1) a single A-format microphone (four tetrahedrally mounted cardioid capsules) such as the Core Sound TetraMic (11) (seen mounted to the author’s custom armature in Figure 1) has an extended range in a small form factor requiring only four recording channels in the field; 2) thanks to mathematical principles developed by Michael Gerzon (12), it is possible to reorient the microphone’s directionality and change the microphone’s pickup pattern after the recording has been made; and 3) thanks to Svein Berge’s High-Angular Resolution Planewave Expansion (HARPEX) B-format decoding, it is now possible to derive a wider variety of microphone pickup patterns from first-order systems with spatial specificity approaching a third-order system (13). The HARPEX plug-in (14) allows us to use a single, highly portable TetraMic in the field, and later derive 26 discreet channels of “shotgun” virtual microphones oriented every 45° in all three dimensions (15).

Figure 6 – Screenshot of Svein Berge’s HARPEX-B plug-in which is capable of deriving eight virtual shotgun microphones per processing pass from a single A-format capture.

Figure 7 – With four processing passes of the A-format audio data through HARPEX-B, it is possible to approximate the above pickup pattern with 26 channels of directional shotgun feeds.

This Ambisonic approach to capturing the directional audio information of a soundscape from a specific location in space provides the structure for overlaying the discreet spatial audio feeds onto the full visual field (as represented by the HDR spherical panorama) from the same point in space. Linking the aural and visual domains in this way creates the “What You See And Hear Is What You Get” effect: multisensory, steerable information from a single point of perception allowing us to recreate and analyze the cross-modal spatial phenomena experienced in the visual soundscape.

4.3 Interactive Multimedia Map Generation

Though not a commonly used feature, it is possible to embed multiple discreet audio files into widely accessible documents. In order to leverage popularity with designers and researchers from a variety of disciplines, the project has thus far focused on using the multimedia features in Adobe’s standalone PDF documents (generated by Adobe’s InDesign software (16)) and Microsoft’s ubiquitous PowerPoint presentations. Embedding location-specific and discreet directional audio files into photographic maps allows the user to interact with the aural dimension simply by using mouse rollovers and clicks to select different spatial aspects of the soundscape recordings.

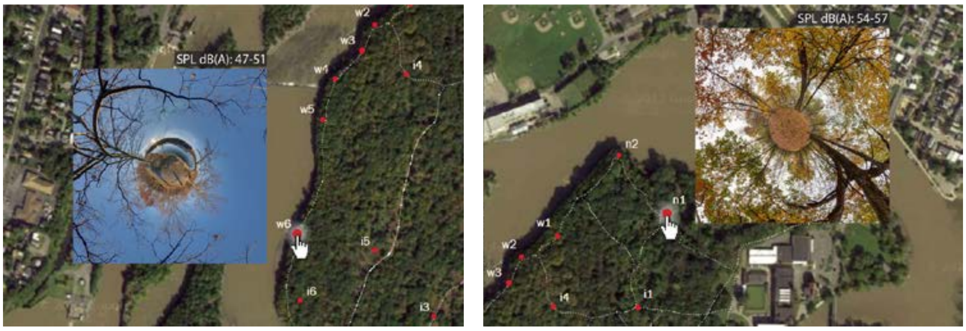

Figure 8 – Examples from a soundscape map which uses mouse rollovers to trigger specific audio files, allowing rapid juxtaposition of different locations within a soundscape.

Figure 9 – Example of a soundscape map from a single location, based on the equirectangular projection of a spherical panorama with embedded Ambisonic audio feeds, allowing comparison of directional sound arrival with visual context.

Currently the project is utilizing “virtual tour” authoring packages to generate more engaging and interactive documents. This software is commonly used in marketing, hospitality, and real estate photography, and produces results similar to Google’s popular “street view” photocartography, albeit with considerably better resolution, exposure quality, and seamless image stitching. Kolor’s Panotour Pro 2.0 (17) also has a feature which allows the embedding of audio files. Normally this functionality is used to add a music soundtrack or narration to the photography, but using the directional audio captured on site has many powerful potential applications. By essentially converting a spherical panorama into a panning tool for the directional Ambisonic audio feeds, we can use such documentation to identify specific sound sources and relationships within a dense acoustic environment and correlate such sound signatures with other visual and ecological attributes. We could also use such documentation indoors, to aid in understanding different architectural reflection contributions in room acoustics.

5. CASE STUDIES

To test the cross-modal soundscape mapping method developed in this project, case studies were chosen which exemplify the complex sonic and environmental interactions that are difficult, if not impossible, to measure or adequately describe using conventional methods and quantitative metrics. The goal was to construct interactive maps which demonstrate that integrating cross-modal data is a viable and effective means for representing large, complex soundscapes. Parks were chosen because of the generally heightened sensitivity to soundscapes within parks—whether local, state, or national—and the expectation that parks provide both humans and animals with restorative habitats which contrast the sensory palette of urban life. As soundscape advocacy gains momentum, parks are an obvious first focus for efforts to confront noise pollution, and it is therefore important that we develop means of mapping and representing park soundscape conditions for effective communication and policy-making.

Two park scenarios in New York State have been studied and documented using the proposed cross-modal soundscape mapping method. Peebles Island State Park, located 8 miles north of Albany at the confluence of the Mohawk and Hudson rivers, was chosen because of its complex combination of sounds coming from isolated dense natural forest, moving water through natural and man-made systems, roadway noise, and industrial factory noise. Peebles Island State Park is an excellent example of soundscape complexity: it enjoys many protected natural views unencumbered by the built environment; however it is impacted significantly by nearby sounds of the built environment. This makes it a perfect candidate for a cross-modal study which can describe the disconnect between visual and aural expectations with more clarity and efficacy than typical unimodal metrics.

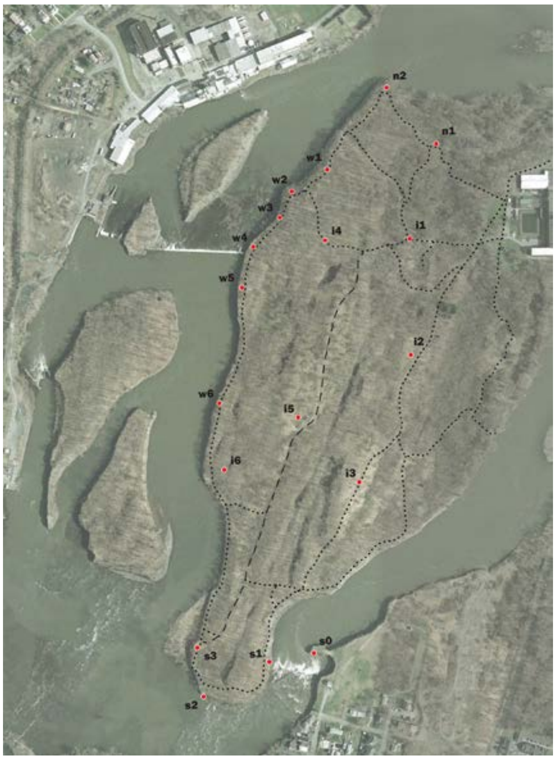

Figure 10 – This Google Earth photo of Peebles Island State Park in Cohoes, NY, is overlaid with the locations where cross-modal soundscape recordings have been made to analyze the complex combined effects of broadband waterborne sound from dams and waterfalls, and industrial machine noise from the paper plant to the north, along with surrounding forest sounds and distant roadway noise.

The second park scenario studied is a comparison of the pair of northern Manhattan parks along the Hudson River side of the island—Ft. Tryon Park and Inwood Hill Park. Both parks have proximity to the traffic sounds of the Henry Hudson Parkway, but they are quite different in terms of their visual experience. Ft. Tryon affords more open views and expansive vistas, and Inwood Hill has more dense and enveloping forest scenarios. Counterintuitively, the intrusion of the highway sounds is more pronounced in the enveloping forest of Inwood Hill Park, given that one's expectations are considerably more disconnected from the highway due to its being out of view than is the case in Ft. Tryon. Furthermore, the ability to view great distances from the vistas of Ft. Tryon diminishes the relative aural impact from the nearby highway. It is precisely such cross-modal phenomena which often prevent our environmental noise metrics from adequately describing or predicting annoyance factors. By combining immersive data for both the aural and visual components simultaneously, the cross-modal soundscape mapping method improves our ability to document and recreate such phenomena.

Figure 11 – This Google Satellite image shows the two northern Manhattan parks sitting along the Henry Hudson Parkway traffic noise. Inwood Hill Park (north) suffers more because its dense and enveloping forest sets up mismatched visual expectations, while the more open and expansive views within Ft. Tryon Park (south) assuage the highway's sonic presence.

6. ONGOING DEVELOPMENT

Previously, this work has focused on developing the necessary field recording techniques, conducting case studies using the new method, and producing soundscape maps in portable document formats. A fairly large library of soundscape recordings has been collected from a variety of environments over the past year. Presently, the project is shifting focus to 1) pursuing applications for hosting such collections online for wider public access, and 2) transposing the collections to a large-scale immersive projection environment hosted at Rensselaer Polytechnic Institute in Troy, NY.

The new Collaborative Research Augmented Immersive Virtual Environment (CRAIVE) Lab is a uniquely capable system comprised of a 14-foot high 360° projection screen with 128 audio channels surrounding viewer-listeners in a social and exploratory setting. By using the soundscape footage acquired in the field and controlling the degrees of visual and aural spatial extents experienced by users, we will be able to conduct subject studies focusing on cross-modal contributions to environmental sensitivities such as perceived loudness and annoyance, acoustic comfort, peripheral vision alertness, and sensorial retention and memory. Furthermore, the large scale immersive and multisensory experiences possible in the CRAIVE Lab will allow us to explore such burgeoning topics as "digital conservation", wherein high-fidelity representations of endangered or extinct environments allow historians, researchers, and students to gain and maintain sensorial insights into ecologies and built environments which cannot be experienced first-hand.

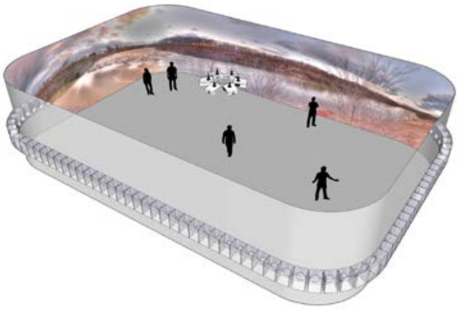

Figure 12 – Rendering of the new Collaborative Research Augmented Immersive Virtual Environment (CRAIVE) Lab at RPI which will allow large scale studies within the cross-modal soundscape recordings made in this project.

7. CONCLUSIONS

While controlled subject studies using these cross-modal soundscape maps are forthcoming, the method developed in this project has already proved to be a valuable approach to representing soundscape complexity and generating meaningful discussion of soundscape phenomena which otherwise escape conventional metrics and predictive models. It should be revealed here that the most profound—and profoundly simple—discovery made while using this method is that the field recording techniques actually become secondary to a larger initiative: to provide those of us researching acoustic ecology and soundscape issues with a structure to move beyond color-coded maps based on predictive algorithms, and to return us to the complexities of *in situ* observation of real-world conditions. The field techniques utilized in this method are essentially an opportunity to reengage with place and heighten one's direct experience and understanding of the environment under study. Such first-hand insights lead to more efficacious discussion with stakeholders.

ACKNOWLEDGEMENTS

This project is being supported in part by Rensselaer Polytechnic Institute and the National Science Foundation (NSF, #12293911).

REFERENCES

1. http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Desktop_publishing.html
2. M. Goldsmith. Discord: The Story of Noise, Oxford, UK: Oxford University Press, 2012.
3. http://www.soundplan.eu/english/soundplan-acoustics/
4. R. M. Schafer, The Soundscape: Our Sonic Environment and the Tuning of the World, Rochester, VT: Destiny Books, 1994.
5. B. Schulte-Fortkamp, and B. Brooks, "Soundscape workshop report: Perception lexicon," in 21st International Congress on Acoustics, Montreal, CA, 2013.
6. C. Bloch, The HDRI Handbook 2.0: High Dynamic Range Imaging for Photographers and CG Artists, Santa Barbara, CA: Rocky Nook, 2012.
7. http://www.hdrsoft.com/
8. http://www.nodalninja.com/
9. http://www.promotesystems.com/
10. http://www.ptgui.com/
11. http://www.core-sound.com/TetraMic/
12. M. Gerzon, "Periphony: With-height sound reproduction," J. Audio Engineering Society, vol. 21, pp. 2-8, 1973.
13. S. Berge, "High Angular Resolution Planewave Ex-pansion," Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics, Paris, FR, 2010.
14. http://harpex.net/
15. J. P. Carter, "Cross-Modal Soundscape Mapping," Master's thesis, Troy, NY: Rensselaer Polytechnic Institute, 2013.
16. http://www.adobe.com/products/indesign.html
17. http://www.kolor.com/ptp2