IMAGING CONCERT HALL ACOUSTICS USING VISUAL AND AUDIO CAMERAS

Adam O'Donovan, Ramani Duraiswami and Dmitry Zotkin

Perceptual Interfaces & Reality Lab., Computer Science & UMIACS, Univ. of Maryland, College Park

ABSTRACT

Using a recently developed real time audio camera, that uses the output of a spherical microphone array beamformer steered in all directions to create central projection to create acoustic intensity images, we present a technique to measure the acoustics of rooms and halls. A panoramic mosaiced visual image of the space is also create. Since both the visual and the audio camera images are central projection, registration of the acquired audio and video images can be performed using standard computer vision techniques. We describe the technique, and apply it to the examine the relation between acoustical features and architectural details of the Dekelbaum concert hall at the Clarice Smith Performing Arts Center in College Park, MD.

Index Terms— spherical microphone arrays, room acoustics, acoustical camera, acoustical scene analysis.

1. INTRODUCTION

Human listening enjoyment and our ability to localize sound and identify environments are greatly influenced (both positively and negatively) by the process of the source sound scattering. Scattering off the environment and off the human before it reaches the ear-canal for physiological transduction and scene interpretation allows for scene interpretation and source localization. The scattering off the listening space (such as an office space, concert hall, classroom, etc.) is influenced by its geometry and the materials of the walls and other scatterers in the space. Since the time of the early acousticians [7], numerous studies on how reverberation affects human perception of sound and music have been conducted. Since the reverberation properties of a room play extremely important role in determining the listening experience (e.g., [8]), architectural acousticians use design principles and measurements/simulation to assure that the room acoustics helps the perception of the performance rather than ruining it.

Room acoustics is generally evaluated in terms of various subjective characteristics expert musicians/listeners assign to sound received at a location in space such as liveness, intimacy, fullness/clarity, warmth/brilliance, texture, blend, and ensemble. Most of these criteria are related to the room impulse response between the sound sources (usually on stage, or from speakers distributed in the hall) and receiver locations (the two ears of the listener at a particular seat). The impulse response is in turn characterized by the direct path from the source to the receiver(s) and the scattered sound received at the received locations. The structure and the discreteness of the early reflections, the directions they arrive from (within about the first 80 ms of first arrival [9]) and the overall energy and structure and directionality of the later part of the response are all held responsible for the various listening characteristics of a space [2]. Modern listening spaces have various computer controlled reflecting elements (curtains, screens, reflectors), that can be placed to provide some control of the achieved nature of the impulse response.

In general the experimental characterization of a space is done via measurements of impulse responses, preferably binaural. A study of the impulse response, attributing various elements of it to architectural features, and the modification of the space to either eliminate or enhance some of the features of the impulse response, are all part and parcel of the work of an architectural acoustician. Of course, as every concert-goer knows, not all seats in a concert hall are created equal in terms of their listening characteristics, and the impulse response varies significantly as source and receiver locations change.

Spherical microphone arrays provide an opportunity to study the full spatial characteristics of the sound received at a particular location. Over the past few years there have been several publications that deal with the use of spherical microphone arrays (see e.g. [5, 12, 6]). Such arrays are seen by some researchers as a means to capture a representation of the sound field in the vicinity of the array [10], and by others as a means to digitally beamform sound from different directions using the array with a relatively high order beampattern [13].

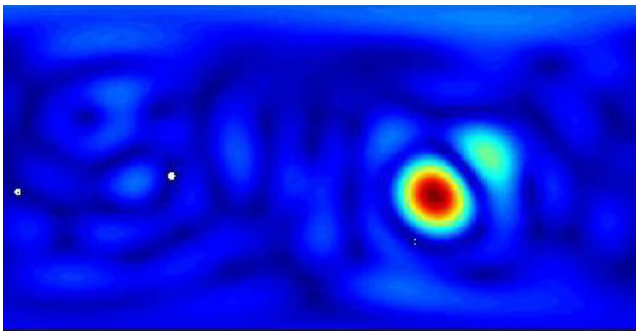

Fig. 1. A sound image created by beamforming along a set of 8192 directions (a 128×64grid in azimuth and elevation), and quantizing the steered response power according to a color map.

Audio Cameras for characterizing room acoustics: A particularly exciting use of these arrays is to steer it to various directions and create an intensity map of the acoustic power in various frequency bands via beamforming. The resulting image, since it is linked with direction, can be used to relate sources with physical objects and scatterers (image sources) in the world and identify sources of sound and be used in several applications, including the imaging of concert hall acoustics that we discuss in this paper.

We gratefully acknowledge the support of DARPA.

Such spherical camera images have already been used to preliminarily characterize concert hall responses [1], though in that paper the measurements were performed over extended periods of time, and the identification with physical objects was performed by interpretation. In effect we use our spherical array and its ability to generate images in real-time as an audio camera. For precision and automation the sound images must be captured in conjunction with a visual camera, and the two must be automatically analyzed to determine correspondence and identification of visual features and the acoustics of the space. For this a formulation for the geometrically correct warping of the two images, taken from an array and cameras at different locations is necessary. We use such a formulation, first presented in [3] that enables the use of a common geometry for analyzing visual and auditory images.

Paper Outline: In Sec. 2 we provide some background and notation for spherical arrays. In Sec. 3 we briefly describe the joint analysis of audio and visual images. In Sec. 4 we describe our measurements of the Dekelbaum theater, and discuss the measurements. Sec. 5 concludes the paper.

2. SPHERICAL MICROPHONE ARRAY AUDIO IMAGING

Beamforming with Spherical Microphone Arrays: Let sound be captured at N microphones at locations $\mathbf{\Theta}_s = (\theta_s, \varphi_s)$ on the surface of a solid spherical array. To beamform the signal in direction $\mathbf{\Theta} = (\theta, \varphi)$ at frequency f (corresponding to wavenumber $k = 2\pi f/c$, where c is the sound speed), we sum up the temporal Fourier transform of the pressure at the different microphones, d_s^k as

$$\psi\left(\Theta; k\right) = \sum_{s=1}^{S} w_N(\mathbf{\Theta}, \mathbf{\Theta}_s, ka) d_s^k\left(\mathbf{\Theta}_s\right). \tag{1}$$

The weights w_N are related to the quadrature weights C_n^m for the locations $\{\mathbf{\Theta}_s\}$, and the b_n coefficients obtained from the scattering solution of a plane wave off a solid sphere

$$w_N(\mathbf{\Theta}, \mathbf{\Theta}_s, ka) = \sum_{n=0}^{N} \frac{1}{2i^n b_n(ka)} \sum_{m=-n}^{n} Y_n^{m*}(\mathbf{\Theta}) Y_n^m(\mathbf{\Theta}_s) C_n^m(\mathbf{\Theta}_s). \tag{2}$$

For the placement of microphones at special quadrature points, a set of unity quadrature weights C_n^m are achieved. In practice, it was observed [13] that for $\{\mathbf{\Theta}_s\}$ at the the so-called Fliege points, higher order beampatterns were achieved with some noise (approaching that achievable by interpolation $(N+1) = \sqrt{S}$). In the beamformer used in this paper, we use one order lower than this limit, the Fliege microphone locations, and beamforming to a fixed $\mathbf{\Theta}$ grid of audio image pixel locations. This allows taking advantage of the spherical harmonic addition theorem which states that

$$P_n\left(\cos\gamma\right) = \frac{4\pi}{2n+1} \sum_{m=-n}^{n} Y_n^{-m}\left(\mathbf{\Theta}\right) Y_n^m\left(\mathbf{\Theta}_s\right) \tag{3}$$

where Θ is the spherical coordinate of the audio pixel and Θ_s is the location of the sth microphone, γ is the angle between these two locations and P_n is the Legendre polynomial of order n. This observation reduces the order n^2 sum in Eq. (2) to an order n sum. The image generation can be performed at a high frame rate using processing on a graphical processing unit [4].

3. COMBINING AUDIO AND VISUAL CAMERAS

Spherical Panorama of the Dekelbaum Theater: As discussed above the spherical array provides a spherical image of the intensities of planewaves from all directions. We needed to compute a similar visual spherical image of the space being measured. To do this, we took a regular digital camera, which we calibrated using standard computer vision procedures. Using this camera we took several overlapping pictures of the theater from near the locations where audio measurements were to be made. While the procedures for creating a panoramic mosaic are well described in the computer vision literature, we simply used a free version of ptGui, a panoramic toolbox available at http://www.ptgui.com/. It finds correspondences in the images automatically and stitches them into a (θ, φ) omnidirectional spherical image (Fig. 2).

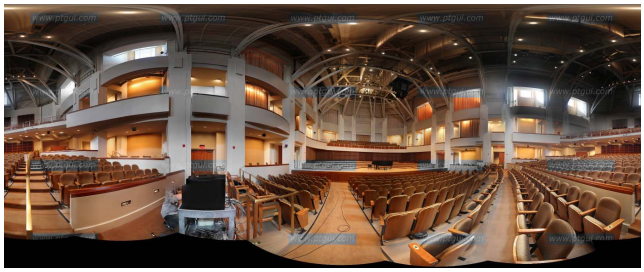

Fig. 2. A spherical panoramic image mosaic of the Dekelbaum Concert Hall of the Clarice Smith Center at the University of Maryland.

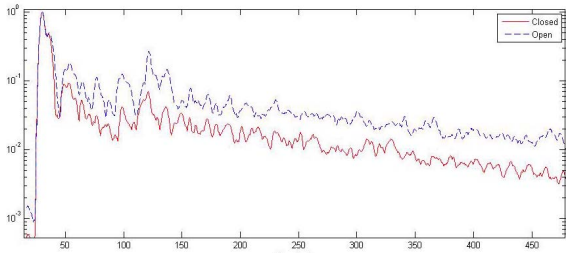

Fig. 3. Peak beamformed signal magnitude for each sample time for the case the hall is in normal mode, and it is in reverberant mode. Each audio image at the particular frame is normalized by this value.

Joint Audio-Visual processing and Calibration: In [3] we provide a detailed outline of how to use cameras and spherical arrays together and determine the geometric locations of a source. The key observation was that the intensity image at different frequencies created via beamforming using a spherical array could be treated as a central projection (CP) camera, since the intensity at each "pixel" is associated with a ray (or its spherical harmonic reconstruction to a certain order). When two CP cameras observe a scene, they share an "epipolar geometry" (see [11]). Given two cameras and several correspondences, it is possible to take points in one cameras coordinate system and relate them to directly to pixels in the second cameras coordinate system. Given a single spherical panoramic image and a corresponding audio panorama image, the transfer can be accomplished if we assume that the world is on the surface of a far sphere. Further cameras can make this transfer without this assumption, but we did not pursue this here.

4. ACOUSTICAL ANALYSIS OF A CONCERT HALL

Measurements: We performed several experiments at the Dekelbaum concert hall located at our university. We created the image

panorama at two different locations, one close to the stage and one towards the center of the hall, at the lower level. The spherical array was placed near where the locations where the panorama was built. For calibration between the visual and audio images, sounds were generated near prominent features in the visual image and the transformation between the audio and the visual panaoramic images obtained. All our measurements can be viewed as a 3D movie that can be navigated at www.umiacs.umd.edu/~odonovan/Visual_Reverb.htm

Next, a loudspeaker source was placed at center-stage and a chirp of length 10 ms played from it. The received data was collected at the microphone array and ten repetitions were taken. We allowed a waiting time of 5 s between measurements, to allow reverberations to die out. The Dekelbaum theater has computer controlled settings which allows various reflective and absorptive elements, at the windows, near the ceiling, and at the back of the hall to be spread out to achieve a "normal" and a "reverberant" setting (other settings are also available). The readings were taken in each of these two settings.

Results of the measurements: Since these measurements were of a somewhat preliminary nature, aimed at both convincing ourselves and others that joint audio-visual imaging can be used to reveal the acoustical features of a listening space, we will present a few observations that our measurements allowed us to make. These results are presented as images in which the acoustic camera image is warped on to the spherical panoramic image, using alpha-blending, with the value of the alpha blending parameter proportional to the peak. A greyscale colormap is used for the acoustical image, and the peak of this colormap is adjusted at each frame. Each individual image then displays the peaks in the sound at that time.

Identifying particular contributions to the impulse response: During the first 90 ms of the recording the acoustic energy highly localized in the images. These very distinct peaks correspond initially to first order reflections. The first major reflection which appears as a single peak in Fig. 4 occurring at 45-60 ms is actually a combination of 3 sequential reflections from the front face of the closest lower balcony and the join of the upper balcony and a support column. In the acoustic video the peak can be seen starting at the front face of the lower balcony sliding up the support column and remaining at the front face of the upper balcony for 5 ms. Approximately 4-5 ms later (1-2 m of sound travel time) the third components of this initial reflection can be seen originating at the back wall of the lower balcony which is consistent with the balconies depth. The next major peak, occurring from 80-90 ms, occurs on the wall directly across the concert hall and exhibits similar behavior starting first at the lower balcony and then sliding up to the second balcony front. After this point the acoustic energy becomes more diffuse and is distributed in several peaks.

Middle time response: From 100-150 ms a very strong peak can be seen in Fig. 5. This peak is associated with a focusing effect of the concave back balcony and lower back wall. The peaks can be seen dancing from left to right and peaking in the center of the wall.

Late time response: Beyond this time, the response is dominated by various pockets of resonant energy in open cavities formed by balconies and box seat areas. Fig. 6 shows a number of these effects.

Measurements in the reverberant condition: In the reverberant condition, with all of the acoustic curtains drawn up, the structure of the first 150ms is very similar to the damped case. The energy however, is much stronger in each of the reflections. After 150 ms, the energy in the hall remains much higher with all of the acoustic curtains drawn up but the structure of the peaks begins to change showing stronger effects resonances occurring at the balconies and the back corners of the ceiling. Fig. 3 shows a plot of the decay in energy from the initial direct sound intensity in both of the conditions.

Focusing effects: The focusing effects observed above are much stronger in the resonant condition, and the acoustical energy dances around the region beneath the balcony.

5. CONCLUSIONS

While the various mechanisms by which sound waves interact with structures are well understood, the acoustics of a listening space such as a concert hall is a complex mixture of these interactions. The spherical array based audio camera can be an extremely useful tool to study the acoustics, and manipulate and understand this acoustics. In conjunction with visual cameras we can make precise identification of the causes of various interactions. As mentioned the audio system is capable of real-time operation. Real-time visual panoramic mosaic generators (e.g., from PointGrey Research and Immersive Media) are also available, and can be combined with our real-time spherical audio image generator to achieve a straightforward implementation that can allow for the interactive imaging and understanding of the acoustics of spaces. Measurements of several others spaces are planned in the near future, as are collaborations with room acousticians.

6. REFERENCES

[1] M. Park and B. Rafaely. Sound-field analysis by plane-wave decomposition using spherical microphone array. J. Acoust. Soc. Am., 118:3094-4003, 2005.

[2] M. Barron and A. H. Marshall, "Spatial impression due to early lateral reflections in concert halls: the derivation of physical measure," J. Sound Vib., 77:211–232 1981.

[3] Adam O'Donovan, Ramani Duraiswami, Jan Neumann. "Microphone Arrays as Generalized Cameras for Integrated Audio Visual Processing." Proc. IEEE CVPR. 1:1 - 8, 2007

[4] Adam O'Donovan, Ramani Duraiswami, Nail A. Gumerov, "Real Time Capture of Audio Images and Their Use with Video," accepted, to appear Proc. IEEE WASPAA, 2007.

[5] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," Proc. ICASSP, 2:1781–1784, 2002.

[6] B. Rafaely, "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Proc., 13, 135–143 2005 .

[7] W. C. Sabine (1900). "Reverberation", originally published in 1900 and reprinted in Acoustics: Historical and Philosophical Development, ed. by R. Lindsay. Dowden, 1972.

[8] H. Kuttruff. *Room acoustics (3rd edition)*, Elseiver, 1991.

[9] D. R. Begault (1994). *3D sound for virtual reality and multimedia*, Academic Press Professional, Boston, MA.

[10] R. Duraiswami et al., "System for capturing of high-order spatial audio using spherical microphone array and binaural head-tracked playback over headphones with HRTF cues," Proc. 119th convention AES, 2005.

[11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[12] Z. Li, R. Duraiswami, E. Grassi and L.S. Davis, "Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays," ICASSP2004, IV:41-44, 2004.

[13] Z. Li and R. Duraiswami. "Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming," IEEE Trans. Audio, Speech and Lang. Proc., 15:702-714, 2007

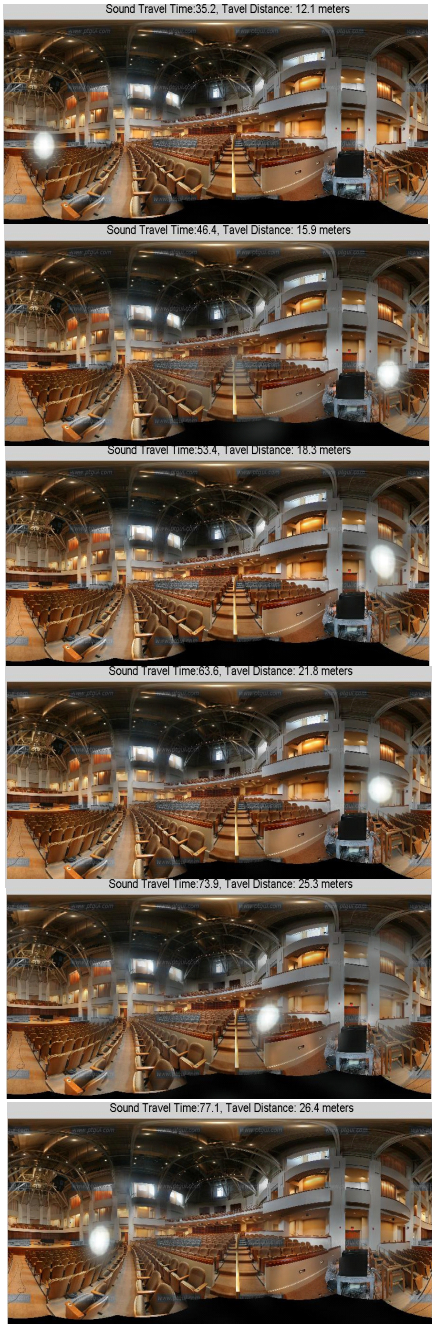

Fig. 4. The frame corresponding to the arrival of the source sound at the array located at the center of the hall, followed by the first five reflections. The sound images are warped on to the spherical panoramic mosaic and display the geometrical/architectural features that caused them.

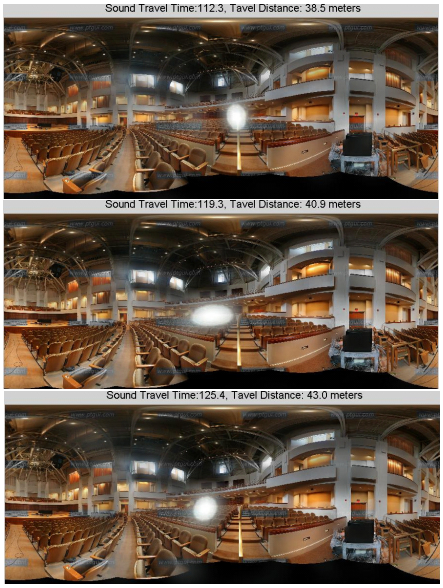

Fig. 5. In the intermediate stage the sound appears to focus back from a region below the balcony of the hall to the listening space, and a bright spot is seen for a long time in this region.

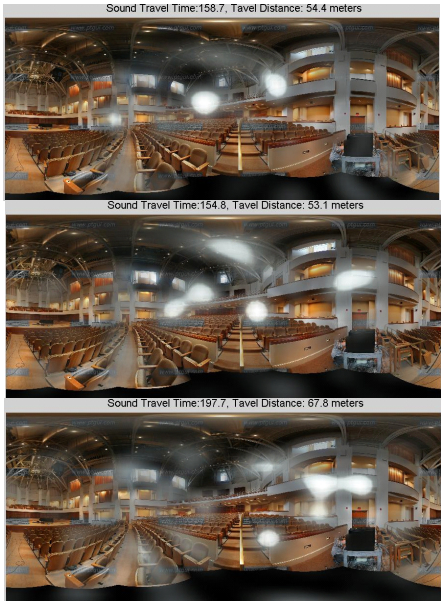

Fig. 6. In the later stages, the hall response is characterized by multiple reflections, and "resonances" in the booths on the sides of the hall.