Acoustic visualizations using surface mapping

Samuel Siltanen, Philip W. Robinson, Jukka Saarelma, Jukka Pätynen, Sakari Tervo, Lauri Savioja, and Tapio Lokki
Aalto University School of Science, Department of Media Technology, P.O. Box 15500, FI-00076 Aalto, Finland
Samuel.Siltanen@aalto.fi, Philip.Robinson@aalto.fi, Jukka.Saarelma@aalto.fi, Jukka.Pätynen@aalto.fi, Sakari.Tervo@aalto.fi, Lauri.Savioja@aalto.fi, Tapio.Lokki@aalto.fi

Abstract: Sound visualizations have been an integral part of room acoustics studies for more than a century. As acoustic measurement techniques and knowledge of hearing evolve, acousticians need more intuitive ways to represent increasingly complex data. Microphone array processing now allows accurate measurement of spatio-temporal acoustic properties. However, the multidimensional data can be a challenge to display coherently. This letter details a method of mapping visual representations of acoustic reflections from a receiver position to the surfaces from which the reflections originated. The resulting animations are presented as a spatial acoustic analysis tool.

© 2014 Acoustical Society of America
PACS numbers: 43.55.Cs, 43.55.Gx, 43.58.Fm, 43.60.Jn [NX]
Date Received: March 11, 2014 Date Accepted: April 28, 2014

1. Introduction and background
Since the pioneering Schlieren photography of ultrasonic wave fronts in scale model auditoria by Sabine (1922), room acousticians have developed creative methods to visualize sound’s propagation within, and interaction with enclosures. In the century after this still-impressive technical accomplishment, advances in computational simulation and measurement techniques have facilitated ever more detailed and revealing demonstrations of room acoustics. Additionally, spatial hearing studies have revealed perceptually relevant spatial features of the sound field.

Some limitations of the cumbersome Schlieren photography methods have been overcome, in particular, ease of measurement and the possibility of computational simulations. However, display media are still predominately flat, and investigators remain faced with the challenge of projecting high-dimensional data upon planar screens and print media. To gain the most complete picture, an acoustic analyst would like to display as many of the acoustic dimensions human hearing is sensitive to as possible. Some basic dimensions include sound’s temporal features, sound waves’ frequencies, and their arrival direction in space. Plotting even these essential features legibly and simultaneously on flat media is difficult. In response to this challenge, elaborate structures and systems such as the AlloSphere (Amatriain et al., 2009) have been constructed to explore high-dimensional data, but such facilities are not yet commonplace. In the meantime, there are many unexplored possibilities for data visualization that utilize commonly available technology.

Acoustic measurement technology that allows accurate assessment of the spatial properties of a sound field using multiple microphone arrays has been developed only relatively recently (Peled and Rafaely, 2012). Traditional microphone technology features omni-directional sensitivity or only basic directional filtering such as a cardioid or figure-eight pattern. The low spatial resolution of these techniques precludes associating, e.g., individual acoustic reflections with the particular reflecting surfaces from
which they originated. Now, high spatial resolution microphones and spatial processing software are becoming increasingly available to mass market consumers (Protheroe and Guillemin, 2013), and spatial analysis can reveal perceptually important spatial features of sound fields (Pätyynen et al., 2014).

One of the advantages of the early images of Sabine (1922) was how intuitively they conveyed their acoustic data; the photographs illustrated sound within the very geometry that had influenced the waves’ propagation. Unfortunately, since then, it has been relatively uncommon to relate sound and enclosing geometry so closely, as attention has shifted to analysis of the room impulse response and parameters derived therefrom. There have been several exceptions to this trend. In one of the earliest examples, Stettner and Greenberg (1989) recorded the sound energy that reached each surface of a concert hall and color coded the surfaces accordingly. Using this surface coding, in combination with icons at receiver positions, they were able to represent 12-dimensional data in their visualizations. Later, Monks et al. (2000) produced similar results as part of an optimization scheme. Rindel and Christensen (2013) used color coding to map the contribution of reflecting surfaces to listener areas. Pätyynen et al. (2013) demonstrated when and where sound was arriving at a listener by superimposing spatio-temporal response diagrams onto architectural plans and sections. Finally, Robinson et al. (2014) recently presented some additional methods for visualizing spatio-temporal and spatio-spectral responses within architectural geometry. The present work expands the theme of incorporating enclosing geometry into acoustic visualizations, with an emphasis on the relationship between architectural geometry and sound delivered to the listener.

This letter details a process for generating visualizations in which spatial impulse response levels are mapped from a listener position onto a computer model of the sound enclosure’s geometry as shading. Similar methods utilizing beam-forming of data from spherical microphone arrays and photographs have been implemented before (O’Donovan et al., 2008). The advantage of the present visualizations is the ability to interactively examine the model from any angle, not just from the measurement position, and the relative ease of measurement, using only 6 microphones, rather than the 16 or more channels required for precise and accurate spherical array beam-forming.

2. Technique
The visualization process begins with an impulse response measurement. Analysis of impulse responses from a spaced microphone array yields the direction of the sound propagation across the microphone array at each instant in the impulse response. In combination with the delay, the direction can be used to reconstruct the image sources that would have produced the incoming sound. Alternatively, the data can be interpreted as the directions of incoming rays from a ray-tracing simulation. Given the direction and delay of the incoming rays, a visualization can be constructed in which the rays are reversed, and directed as beams onto the specific architectural geometry that produced the reflection towards the listening position. For the architectural acoustic designer, this illustrates the contribution of architectural features to different temporal portions of the impulse response, but not the order of the reflection or what path the sound took to reach the surfaces before proceeding to the listener. Unlike the visualizations of Stettner and Greenberg (1989) and Monks et al. (2000) that record sound energy incident upon a surface, the present technique codes the surface based on how much sound energy actually reaches a listener from that direction. This type of display may reveal less about the overall properties of the space, but is more perceptually relevant for a given position within the space. Sound energy incident upon a surface may be reflected to another direction and never reach the listener, or it may arrive from another direction.

The sound source for measurements is a Genelec 8020 studio monitor at 1.2 m above the stage at a position 2 m behind the plaster line and 1 m stage right. The
A loudspeaker was calibrated to 87 dB at 1 m with A-level weighting using band-limited noise (200 Hz–5 kHz) in each hall. Room impulse responses were measured using a single logarithmic sine-sweep, at 48 kHz sampling frequency, for full audio bandwidth (Farina, 2000). The sweeps were recorded with an open spherical microphone array, G.R.A.S. vector intensity probe, which consists of 6 omni-directional microphones. Pairs of microphones spaced 100 mm apart were aligned along Cartesian axes, and centered at the intersection of the axes. The visualizations presented in this letter are derived from measurements in two concert halls: Herkulessaal in Munich, and Beethoven-Saal in Stuttgart. The former is a shoebox-shaped hall, and the latter is a fan-shaped amphitheater type hall.

The room impulse responses obtained from the microphone array were encoded utilizing the spatial decomposition method (Tervo et al., 2013). The method utilizes the pressure responses from the six microphones in the array to determine the direction of arrival of the sound at each sample. A short time window around the sample is used to calculate the cross correlation between the 15 possible pairs created from the 6 microphone array. The peaks of the cross correlation functions provide an estimate of the time difference of arrival of an assumed plane wave at the microphones. Consideration of the array geometry and time differences of arrival allow estimation of the direction of arrival of sound at the time sample. The direction, in combination with the time, can be used to create a reconstruction of the image sources that would have produced the response.

The spatial decomposition method operates on the assumption that there is only one acoustic event, i.e., reflection or direct sound, in each analysis window. As is well-known, the reflection density in a room impulse response, i.e., the number of acoustic events per time window, increases with respect to the square of time. It follows that the assumption is valid for a longer time with a shorter analysis time window. On the other hand, the dimensions of the microphone array impose a limitation to the shortest possible window size, since the analyzed acoustic event has to have enough time to travel through the whole array. Therefore, the analysis window length in samples \( L > \frac{2f_s d_{\text{max}}}{c} \), where \( d_{\text{max}} \) is the maximum Euclidean distance between any two microphones in the array, \( c \) is the speed of sound, and \( f_s \) is the sampling frequency. The whole spatial impulse response is processed in these short time analysis windows with \( L - 1 \) sample overlap. The current analysis utilized a window of 64 samples at a sampling frequency of 48 kHz, equivalent to 1.33 ms.

The result of the spatial decomposition method analysis is a four-component vector for each sample in time that can be used to create the visualization. The components are pressure, taken from one of the omni-directional microphones, and the three Cartesian spatial components of the arrival direction.

Given the spatial decomposition method data and a three-dimensional model of the concert hall, it is possible to determine which surfaces are reflecting the sound energy and when the reflections occur. Conceptually, the idea is to trace beams backwards from the receiver point towards the image sources. Each beam then intersects surfaces of the measured space. These surfaces are assumed to be the ones causing the reflection. This tracing was conducted with beam tracing algorithms described by Laine et al. (2009).

For visualization purposes, the surfaces of the three-dimensional model of the concert hall are texture mapped and rendered using OpenGL and the GLUT Toolkit (Kilgard, 1996). The background shade is a low intensity gray, and surfaces are initially displayed only as a slightly darker shadow. As sound energy arrives at the listener from a given direction, the surfaces in that direction are lightened. The shading map spans from the background shade to white. White is indicated in the direction of sound energy within 3 dB of the impulse response maximum, and the background shade represents a level 23 dB below the maximum.

Since it is highly unlikely that the points corresponding to the pixels coincide with the intersection points of the rays with the surfaces, we visualize probability
distributions of the reflection locations. This is done by calculating the angular distances from the points corresponding to pixels in texture map to the exact intersection points. This distance is then used as a parameter for the Gaussian distribution of lightening around the sound arrival direction. The final shade of the pixel is proportional to the sound pressure level at that time instance, and the angular distance between the pixel and the sound direction at that time instance. The result is a mapping of the probability distributions of the image source location onto the surfaces of the concert hall, weighted by the absorption of the propagation path—a complete texture map of the hall for one sample in time.

Showing all the reflections in the impulse response at the same time is not very informative, because reverberant spaces eventually approximate a diffuse field, with sound arriving from every direction and every surface. Thus, the reflections are grouped into one millisecond time windows. Since the sampling rate is 48 kHz, a single frame of the animation is the average of 48 texture maps of image sources. These frames are smoothed using a 5 ms sliding window to generate the final animation frames. This smoothing reduces the apparent spatial resolution of the visualization, but also reduces flickering in the animations. It is then possible to add reflections to the visualization 1 ms at a time, starting from the arrival of the direct sound. Similarly, it is possible to rewind the reflections. The user is also given the freedom to move the view point and direction. This makes examining the reflections easy. Time is indicated in milliseconds in the lower left corner of each frame, with \( T = 0 \) defined as the arrival time of the direct sound at the receiver.

3. Results

Four associated video files demonstrate the results of the visualization process. In addition to the maps on the surfaces of the concert hall, a semi-transparent distorted tesselated sphere is drawn around the receiver point. The radius of each tesselated point is proportional to the energy level accumulation of reflections from its direction. In a diffuse field, the tesselated points would all have the same radius from the receiver position, resulting in an exact sphere. The sound source on the stage is indicated by a yellow sphere.

Video Mm. 1, displays two halls side-by-side from a receiver position 14 m from the stage and 2 m to stage right from the centerline. Figure 1 illustrates a frame from the animation. The spatio-temporal development of the two halls’ sound fields can be compared. On the left is the Herkulessaal in Munich, and on the right is Beethoven-Saal in Stuttgart. Mm. 2 depicts the same halls, but viewed from the side rather than from the listener’s perspective.

Mm. 1. Video demonstrating animations of reflection maps at the same position in two halls simultaneously. This is a file of type “mp4” (4.5 Mb).

Mm. 2. Video demonstrating animations of reflection maps at the same position in two halls simultaneously, from the left side. This is a file of type “mp4” (2.5 Mb).

The difference between the halls is notable. In Herkulessaal, reflections from the left sidewall are evident in the response shortly after 10 ms and from the right before 20 ms. At 20 ms in Beethoven-Saal, the reflected sound energy is predominantly from the direction of the stage. It is not until 30 ms that strong reflections are evident simultaneously from the left sidewall and the ceiling. Such a reflection sequence, with overlapping ceiling and sidewall reflections is known to be detrimental to spatial responsiveness (Marshall, 1967). The ceiling reflection in Herkulessaal does not arrive at the listener until 40 ms, well after both sidewall reflections. By the end of the visualization, Herkulessaal has a well distributed and still strong response, while discrete reflections persist in Beethoven-Saal and the response is weaker overall. These observations of the visualizations are quite useful when analyzing room acoustics, but are difficult to discern from impulse responses or acoustic parameters.
Figure 2 illustrates a representative frame from Mm. 3 There are three panels that represent the same receiver point in Herkulessaal from various camera positions. Starting from the left panel and working right, the camera positions are behind, to the left, and below the receiver position. In every case, the camera is aimed at the receiver position. The animations show 100 ms of the room impulse response starting at the direct sound arrival time. Mm. 4 follows the same format for a receiver at the same position in Beethoven-Saal. These animations illustrate one drawback of the animations. Namely, there is a reflection shortly after the direct sound that originates from the seats behind the receiver. Since the seats are not included in the model, this reflection is incorrectly projected to the back wall. This effect reveals the importance of selecting an appropriate level of detail when generating the geometric model.

Mm. 3. Animation of the sound field development in Munich Herkulessaal. Three views show which surfaces correspond to sound arrival directions, animated through time. This is a file of type “mp4” (4.7 Mb).

Mm. 4. Animation of the sound field development in Beethoven-Saal, Stuttgart. Three views show which surfaces correspond to sound arrival directions, animated through time. Note that the apparent lack of energy under the balcony does not necessarily indicate that those seats would be quieter, only that no energy is reaching the selected receiver from under the balcony. A separate animation is necessary to illustrate each receiver position. This is a file of type “mp4” (4.2 Mb).
4. Conclusion
Visualizations of sound fields within enclosed spaces have been presented. The method consists of mapping shading representing sound arrivals onto a 3-D model of the enclosing geometry. Its utility has been demonstrated by comparing the spatio-temporal development of sound fields in two concert halls. This work builds on existing visualization techniques, and provides an additional way to analyze sound fields. Additional possibilities to be explored include spatio-spectral representation, and well as further incorporating perceptual relevance into visualizations.

Acknowledgment
The authors would like to thank Petri Leskinen for providing inspiration for the tessellated sphere portion of the visualization, and for the 3-D model of Beethoven-Saal.

References and links