The Acoustics Model for the Paul's Cross Project

By Matthew Azevedo

Matthew Azevedo is a consultant in acoustics at Acentech.

An auralization is generally created by convolving an impulse response of an acoustic environment with a sound recording to create a new audio stream of the sound as if it had been recorded in the space corresponding to the impulse response. An impulse response is a complex filter that can be used to describe the time and level variations of a linear, time-invariant system, such as an acoustic environment. An impulse response can either be measured directly, or it can be mathematically calculated. In the case of Paul's Cross, where the space has been gone for hundreds of years, a calculated impulse response is the only option.

Calculating an acoustic impulse response requires a 3D computer model of the space to be simulated. We commonly see visual 3D models, where a framework of planes is "painted" with the visual colors and textures of the modeled object. An acoustical model is similar, but instead of painting the model with visual characteristics, the model is painted with acoustical characteristics, such as sound absorption and sound diffusion. Sound sources, including their frequency response and directivity characteristics, are added to the model, as are sound receivers (listeners). Then, the direct and reflected sound paths between the source and receiver are calculated through a combination of geometrical and statistical processes.

There are several software packages available for simulating room impulse responses. For the Paul's Cross project, CATT-Acoustic was selected due to its flexibility, usability, and quality of results. CATT models are created from script files that define the corner locations of each plane and its acoustical characteristics. Creating the models from script files has both advantages and disadvantages. On the positive side, it means that the models can include variables in their definitions, which allows rapid redefinition or alternation of room parameters. This is extremely useful in situations where acoustical modeling is being used as part of a design exercise, and it can save significant time when a model needs multiple iterations to determine the ideal room geometry for the desired results. On the negative side, the model is built as a text editor with minimal visual feedback, resulting in a cumbersome initial modeling exercise. Thankfully, there are tools that allow models created in graphical environments to be exported as CATT scripts. We have found that the most efficient process for modeling a space is to build the initial model in a graphic environment such as Google SketchUp and then export to a CATT script, which can be edited to parameterize various aspects of the model if needed.

Modeling of Paul's Cross and St. Paul's Cathedral began with a simplified version of the high-resolution SketchUp model built by Joshua Stephens at NC State University. Due to the statistical nature of CATT's processing, a model with too high a level of detail can result in poor accuracy compared to a geometrically simpler model with the fine surface details represented by scattering coefficients applied to large, simple planes. The visual model was simplified to a level of detail appropriate for processing in CATT, with the final acoustical model consisting of 18,000 polygons compared to the visual model, which had 1.2 million polygons.

CATT then generated impulse responses for each of 12 listener positions in five versions of the model with audiences varying between no and 5,000 people. In the interest of minimizing the number of convolution channels needed to realize the final presentation, only three sources were included in the model: a source with human voice directivity at the preaching station, an omnidirectional source in the bell tower of St. Paul's Cathedral, and a catch-all "ambience" source in the center of the churchyard. The simulated impulse responses use ambisonics to represent the spatial aspects of sound in the churchyard. Ambisonics is a technique for representing not only the sound pressure at a point, as a normal microphone recording does, but also the 3D directionality of sound. The ambisonic technique has the benefit of decoupling the recording channels from the playback speakers. An ambisonic recording can be decoded to play back on any speaker arrangement, from headphones to a multi-channel speaker dome. Using ambisonics allowed us to create one auralization that could then be played back on any speaker system as required.

Source material for auralization must be recorded in a highly controlled acoustical environment, ideally in an anechoic chamber. "Anechoic" literally means "without echoes"; an anechoic chamber is a room where the walls, ceiling, and floor have been covered with very deep sound-absorbing material, resulting in a room where there are no sound reflections from any surface. This is important because any reflections in the source recording will be introduced into the auralization. An anechoic recording of speech convolved with the impulse response of a church will result in the sound of the speech in the church. If the voice was recorded in a concrete stairwell, we would instead hear the sound of the speech in a stairwell in the middle of the church. Using anechoic, or at least minimally reverberant, source material prevents this "space within a space" error.

Most of the source materials for convolution with the impulses from the CATT model were recorded specifically for the Paul's Cross project. The most critical recording was of the sermon, which was made in the anechoic chamber at Salford University, Manchester, UK. The sermon was performed by Ben Crystal, an actor who has studied 17th-century English dialects. Properly pacing the performance was a significant challenge: the sermon was performed in anechoic conditions, but would be presented with a long reverberation and the noise of a large crowd. The sermon was recorded from multiple angles, and Crystal was encouraged to visualize the crowd around him and direct his speech to various people, particularly the nobility, who would have been seated in the sheltered seats along the cathedral wall.

The less critical crowd sounds were recorded at PostPro studio in Raleigh, NC, USA. While the studio's tracking room is not anechoic, substantial additional absorption was hung in the studio to make the recordings dry enough to use in the auralization. A walla group (the traditional name for a group of actors speaking in nonsense sounds to create the din of a crowd in the background of a scene) was formed of students from NC State's linguistics department. Crowd noise was recorded in multiple long takes to allow several takes to be played back at once without giving the impression of the material being looped or repeated.

Our initial intent was to make field recordings of actual church bells to use in the auralization. However, the bell at St. Paul's was a hammered bell as opposed to the more common pulled bells. Because of this, we were unable to locate an appropriate bell. Instead, the bell was created by resynthesizing a recording of an orchestral tubular bell. The source recording was pitch-shifted, equalized, and layered to create the sound of a much larger hammered bell.

The real-time auralization itself was written in a graphical programming environment called Max/MSP. Max was originally written as a programming tool for composers who wished to create music using algorithmic techniques, but has grown into a fully-featured media programming environment that can process and generate audio and video in real time, while also creating a graphic user interface to allow user control of the processing.

Creating a unique set of responses for each of the 5,000 people in the crowd would have been a Herculean task. Instead, we developed a listening algorithm that follows the sermon and generates realistic crowd behavior automatically. A rudimentary artificial intelligence (AI) written in Max listens to the sermon and then selects samples from the walla recordings of an appropriately intense response. The full crowd is made up of 90 independent instances of this AI, which are tiled across the soundfield of the churchyard. Each AI's behavior preferences are randomized at runtime. This not only allows the impression of a very large crowd of independent listeners to be generated from a small amount of source material but also results in a simulation that is unique every time it is run. Similarly, environmental sounds such as dogs, birds, and horses are triggered and positioned in the soundfield via stochastic processes.

For a more in-depth discussion of the processes and technologies that underlie the Paul's Cross auralization, please see "Acoustical archaeology — Recreating the soundscape of John Donne's 1622 gunpowder plot sermon at Paul's Cross" by Matthew Azevedo, Benjamin Markham, and John Wall, in *Proceedings of Meetings on Acoustics*, Vol. 19, No. 1 (June 2, 2013).