

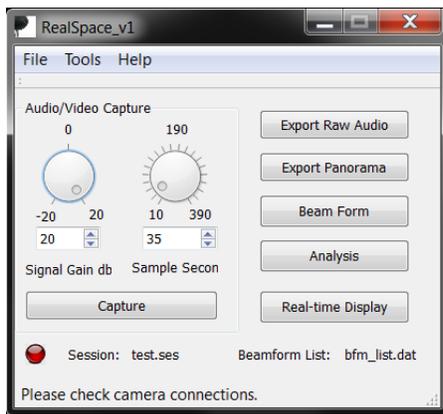
## ROOM AND CONCERT HALL ACOUSTICS

The perception of sound by human listeners in a listening space, such as a room or a concert hall is a complicated function of the type of source sound (speech, oration, concert, music and type), of the intended outcome (comprehension, comfort, enjoyment, localization), of the room (geometry, reflecting surfaces/absorbers, size) and the listener (HRTF, hearing acuity). Quite obviously, reducing this tremendous variability to a single standard is an extremely hard task. The ISO 3382 standard [1] specifies several performance space acoustic parameters (Part 1) and room acoustics parameters (Part 2) for this. These measurements include varying quantities (decay times, clarity, etc.) extracted in varying frequency bands, for both single microphone and binaural measurements. The binaural measurements attempt to account for the effects of the human head, torso and pinnae on the measurements, but do not account for inter-personal variations.

To go beyond this standard, one needs to understand the impulse response for every source-receiver position combination of interest. In the case of binaural listening, this needs to be done in a way that particular individual head-related transfer functions can be employed.

### MEASUREMENTS USING ARRAYS OF CAMERAS AND MICROPHONES

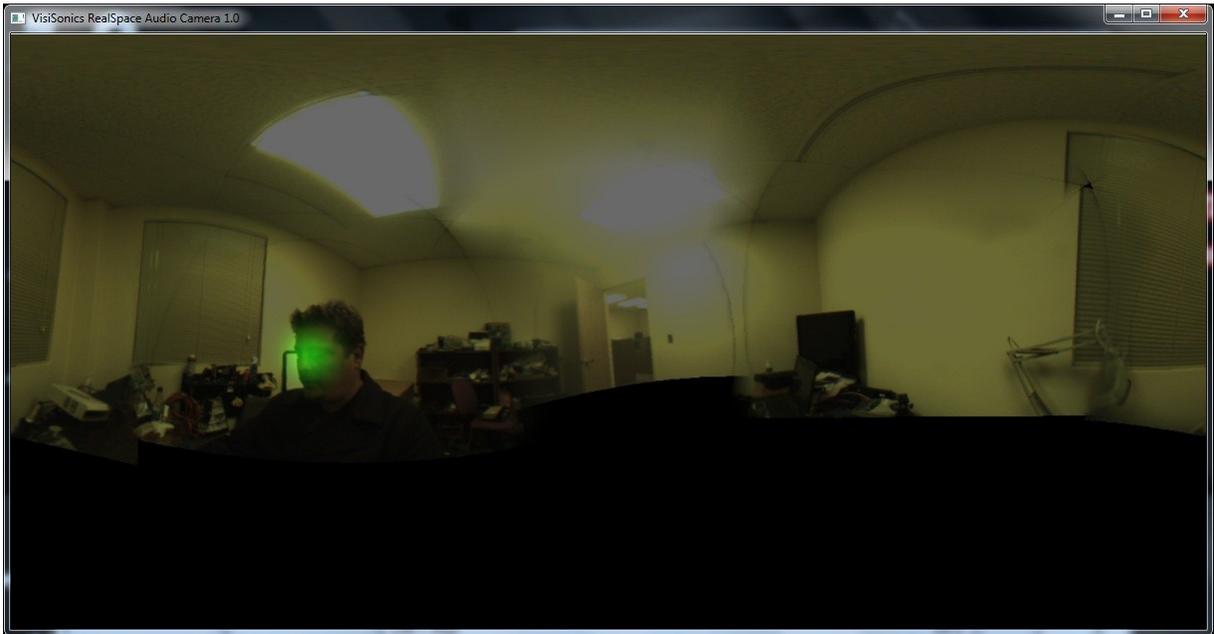
To obtain a finer understanding of the listening characteristics, individual impulse responses must be measured for each direction and their structure understood from the geometry and materials used to construct the space. In particular for the early stage, the interaction of the source sound with each scattering surface, in each frequency band can be easily characterized using the VisiSonics AudioVisual Panoramic Camera. This device provides several microphones (64 or 128), and several high definition cameras (5 or 15), which provide an ability to record sound, and to co-register the sound. The device allows 24 bit recordings, at 3 user-selectable gain levels, and 48 kHz sampling. All the cameras and microphones are mounted on a head-sized sphere (0.2 m diameter). The use of an advanced digital microphone architecture (Zotkin et al 2013) allows the audio measurements



from all cameras and microphones be sent over a combination of IEEE1394 (video) and USB 2 (audio), or over USB 3 (audio and video).

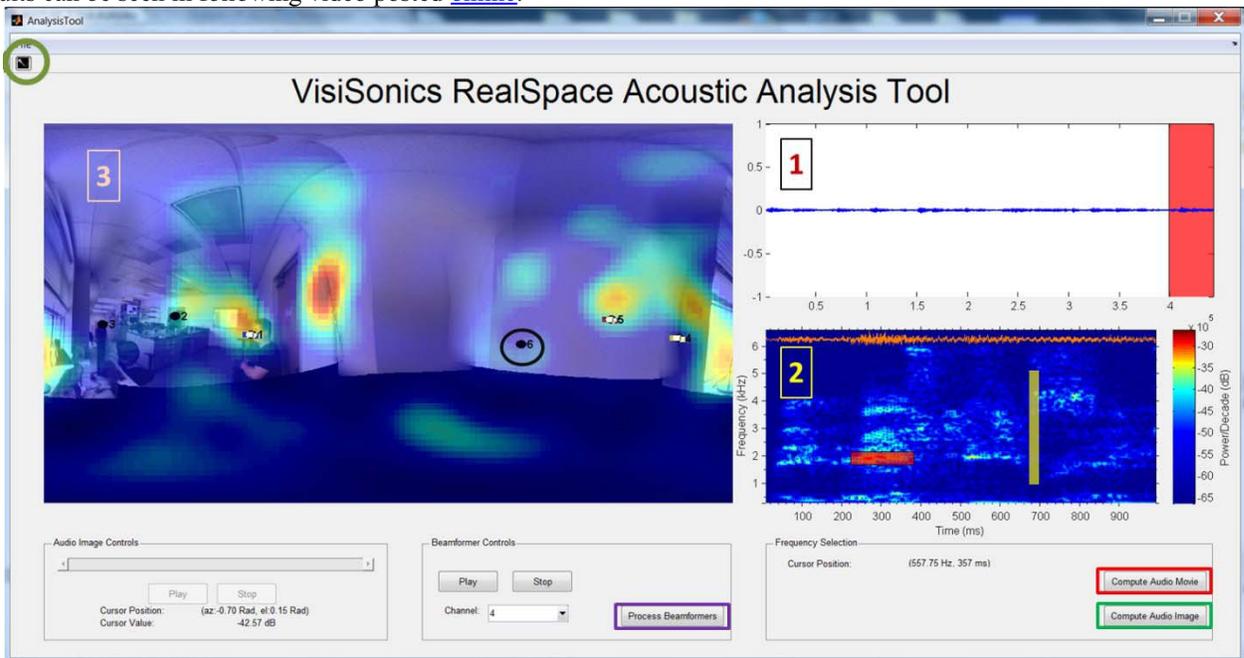
A NVIDIA GPU card equipped laptop allows both real time analysis of a scene, and of sophisticated on-scene computations. A panoramic video image is computed from the multiple video cameras by using a GPU based stitching algorithm. An overlaid audio image is computed by employing several thousand plane-wave or spherical beamformers on a grid of directions on the sphere (O'Donovan et al, 2007). The magnitude of these beamformers in particular frequency bands are mapped to false color values and overlaid via alpha-blending as a transparent texture over the panoramic video. The raw audio and video data are streamed to disk, and can be used for more careful late analysis. Real-time analysis possible via the video stream includes identification of prominent reflections, and the geometric sources of their origin. A

sample screenshot with a speaking user in a room is shown below.



## SPATIAL, TEMPORAL AND SPECTRAL ANALYSES

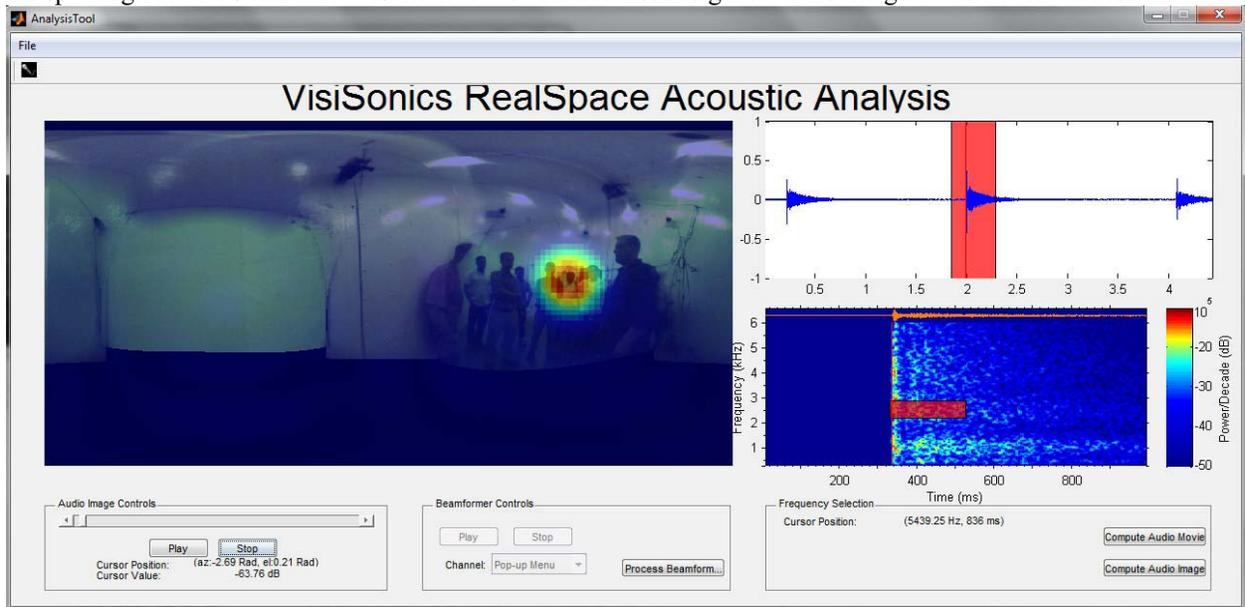
Various off-line analyses can be performed via the tool. These include the computation of directional impulse responses, the analysis of the reverberation in temporal windows, the analysis in spectral windows, the online computation of spectrograms with various weightings and frequency scales, and others. Some preliminary analysis results can be seen in following video posted [online](#).



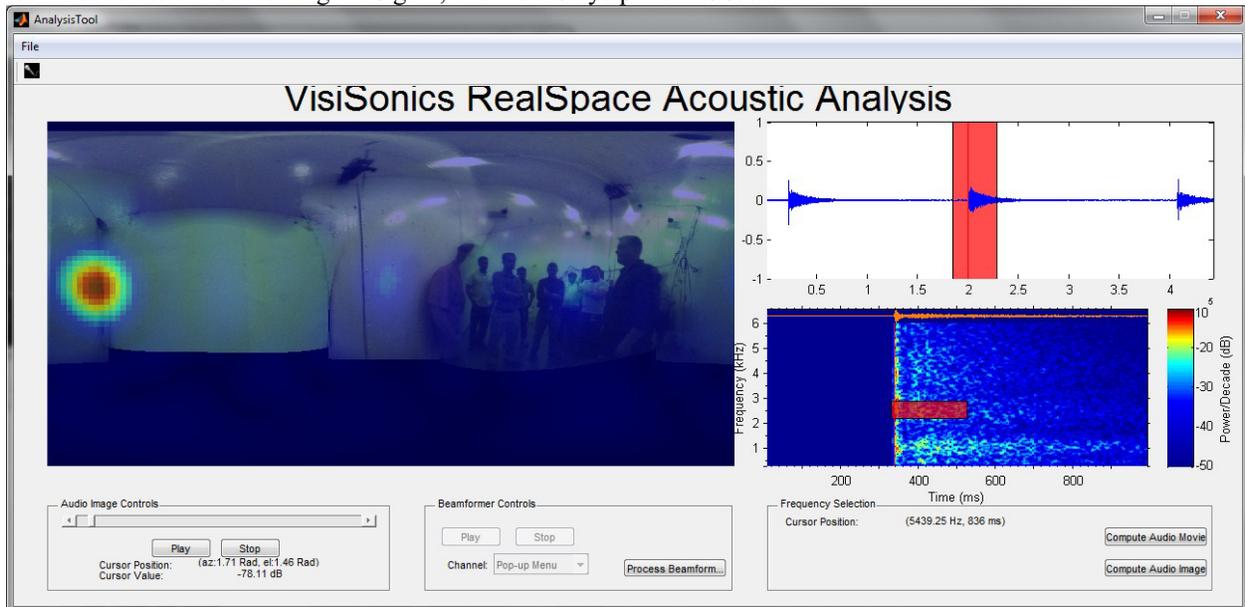
The interface for performing spectro-temporal spatial analysis of the acoustic scene is shown above. Window 1 shows the time domain plot of the signal recorded. A sliding red box allows a particular time interval of interest to be selected.

Window 2 shows a spectrogram of the selected time region. The spectrogram is a useful tool for finding structure in the recorded audio. It can be used to select a frequency and time sub region to generate an acoustic image.

In the following images, screenshots are shown from a recording made in a reverberant chamber. The image below shows the first frame of the acoustic image. It was generated by computing the acoustic images for all frequency bins within the user selected frequency range of a 256 sample frame of data. For each successive image the sample frame is advanced by 10 samples and the corresponding image is generated. This sequence can then be played to explore the dynamics of the instantaneous spatial distribution of the incident sound field. Notice the thin red line in the spectrogram. This line indicates the extent of the data used to generate the image in the current frame.

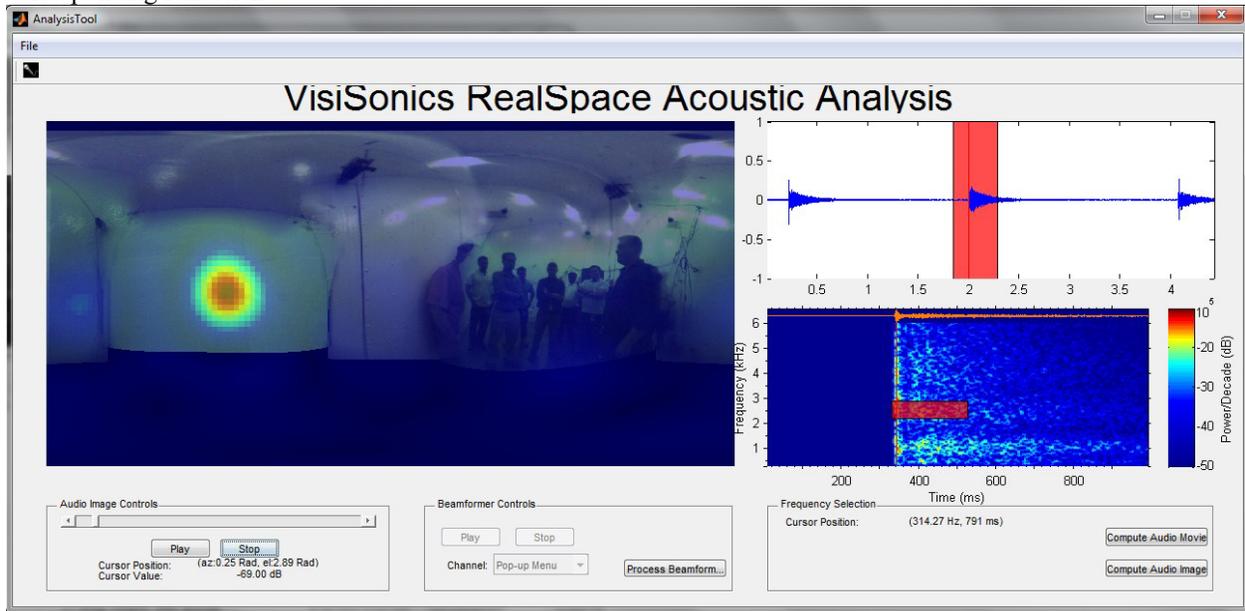


The image above shows when the impulse onset has occurred; the spatial distribution of the acoustic field is very localized to the driver creating the signal, and a bulls-eye pattern is seen.

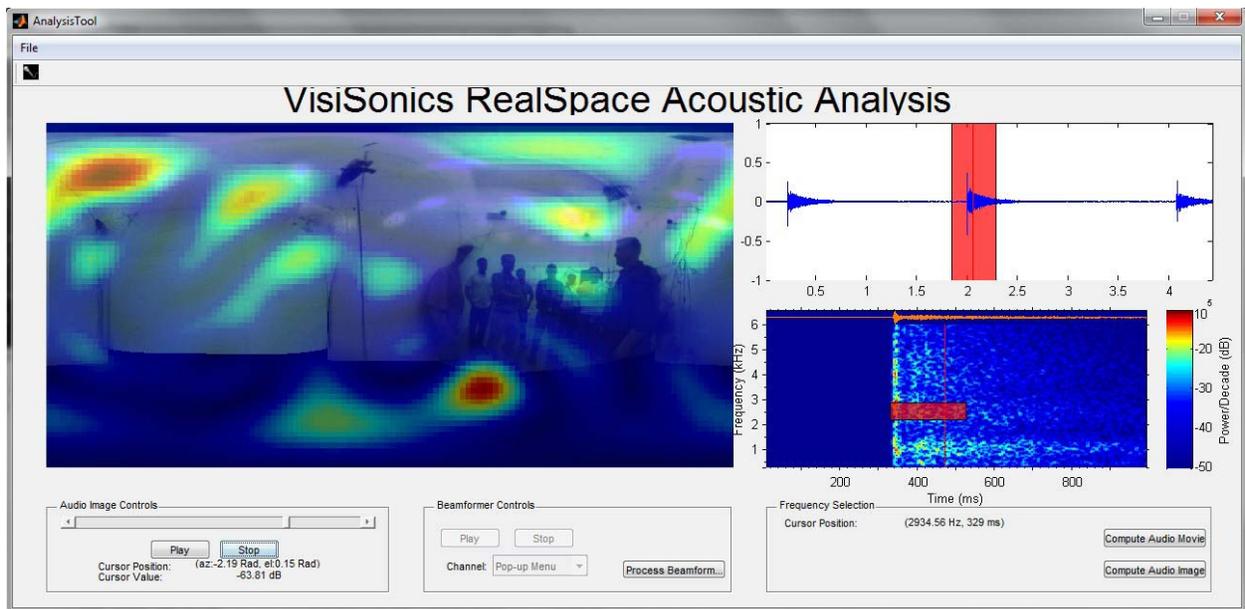


Because the analysis frame time can be selected accurately, we can advance the frames, until the first reflection is distinctly seen. In this case the image above shows that it is on the leftmost wall of the reverberant chamber. The corresponding bulls-eye pattern of the reflection is seen clearly in the image below, as is the temporal pattern of the

corresponding reflection.



In the image above, the frame has been advanced to show the second distinct reflection occurring on the center wall.



Advancing 100 ms (the image above) shows just how diffuse the field has already gotten in this reverberant chamber.

## CONCLUSIONS AND ONGOING WORK

The images above just indicate a small selection of analyses possible. Users of the device have used it to quickly identify problematic structures in rooms. It is also possible to compute integrated quantities over time and frequency. Further, using known sources we can compute impulse responses. In a version of the tool under development, we can compute impulse responses corresponding to any direction.

## REFERENCES

- ISO-3382, "Acoustics - Measurement of room acoustic parameters -- Part 1: Performance spaces" (2009)
- ISO-3382, "Acoustics - Measurement of room acoustic parameters -- Part 2: Reverberation time in ordinary rooms" (2008)
- A. O'Donovan, R. Duraiswami, Jan Neumann (2007). Microphone Arrays as Generalized Cameras for Integrated Audio Visual Processing. IEEE CVPR.
- A. O'Donovan, D. N. Zotkin and R. Duraiswami (2008). A spherical microphone array based system for immersive audio scene rendering. Proc. 14th International Conference on Auditory Displays. Paris, France.
- A. O'Donovan, R. Duraiswami and D.N. Zotkin. (2010). Automatic matched filter recovery via the audio camera. Proc. IEEE ICASSP, (pp. 2826-2829). Dallas, TX.
- A. O'Donovan, R. Duraiswami and D.N. Zotkin. (2008). Imaging Concert Hall Acoustics using Audio and Visual Cameras. IEEE ICASSP, (pp. 5284-5287). Las Vegas, NV.
- A. O'Donovan, R. Duraiswami and Nail A. Gumerov. (2007). Real Time Capture of Audio Images and Their Use with Video. IEEE WASPAA, (pp. 1-8).