BeBeC-2012 -15

SIMULATION, VISULIZATION AND PERCEPTION OF SOUND IN A VIRTUAL ENVIRONMENT USING BEAMFORMING

Mojtaba **NAVVAB**[1] , Gunnar **HEILMANN**[2] and Andy **MEYER**[2]

[1]**TCAUP ,** Bldg. Tech. Lab, The University of Michigan
2000 Bonisteel Blvd, Ann Arbor MI 48109-2069 USA, E-mai:“moji@umich.edu
[2]**Gfai** Tech GmbH, Berlin, Germany.

ABSTRACT

This paper describes a method to visualize and localize the sound that is simulated within a virtual environment using beamforming. The use of an acoustic camera along with noise image software as a short introduction to beamforming method is demonstrated. Furthermore, the transition from the three-dimensional sound recording to the three-dimensional virtual acoustic mapping, visualization and sound perception for its directionality by real subjects within the virtual environment is described. Prerequisite for this is a 3D-model which can be created quickly within this computer aided virtual environment. The results show that the subjects were able to navigate and locate a real and virtual sound source in a dynamic virtual acoustic environment. The findings from these simulations, auditory navigation experiments via visualization technique within this virtual environment demonstrate the beamforming method combined with human subject data provide opportunities to study sound localization and fine tune the current Head Related Transfer Function (HRTF) for various room acoustic design applications.

Key words: Beamforming, acoustic mapping / visualization, sound localization, Auditory Navigation, Virtual Acoustics, Spatial Hearing, Dynamic Auralization

1 INTRODUCTION

Beamforming systems for imaging analysis of complex sound sources have been made possible by Noise Image software in industrial applications for number of years. Real architectural spaces have complex deep-structured surfaces and often times require distributed sound sources within the space. These conditions require high demand in computing time during simulations and or detail measurements during space auditing for the sound system, but are not free of errors in plane approximation in the form of falsification of levels, distorted localization and aliasing effects in simulated and or measured results as a consequence. The sound auralization based on these results will not be as realistic as the measured data within these spaces. One of the objectives in this study was to conduct an experiment in such a way that the perceived sound changes according to the movements of the subject (e.g. distance & orientation) in a given simulated room for its acoustic characteristics to be evaluated. The results could be used as part of an auditory navigation experiment. Many auditory navigation tests have been conducted in the past, and an overview of these different techniques needed in auditory navigation shown in research work by **Loomis et al. [1], Rutherford [1] and Begault [3]** are directly related examples of such experimental work.

In our experiment we applied the capability of EASE, a simulation software, and FMOD auralization systems within a VR Laboratory to obtain close approximation of the simulated room architectural characteristics of an actual 3D-virtual reality laboratory space called the "CAVE". The locations of the loud speakers and their directionality were modeled based on the actual measurements using the acoustic camera recording system. **[4, 5, 6, 7, 19].**

Given the dynamic changes within today's audio system technologies; the meaning of "3D Sound" to the public at large is characterized more in terms of processing, however this characterization is not monophonic or stereo. Most human sensations involve electrochemical or chemical reaction within the brain and the "3D sound" is best understood or perceived through its localization within a given space for its volume and surface absorption characteristic. This perception is individual and totally subjective and NOT objective. Based on many technical papers published related to sound perception and sound localization in humans; the externalization and localization (directional) effect, the environmental reflections, scattering, and reverberation cues are required to produce the best 3D sound effect.

In passive "3D sound" processing, the source position as being localized does not change if the listener moves the head since the dynamic or static sound source is independent of the head position and its direction; however; in the interactive "3D sound", dynamic cues such as listener position and movement plus sound movement (if any) within the audio scene must be added. These sound localizations are achieved by the application of the HRTF through the inclusion of the cues exclusive of differences between the signals reaching each ear as Inter-aural Time Delay (ITD) and Inter-Aural Level Differences (ILD). "Relying on a variety of cues, including intensity, timing, and spectrum, our brains recreate a three-dimensional image of the acoustic landscape from the sounds we hear" **[8, 9].**

2 EXPERIMENTAL SET UPS WITHIN VIRTUAL LABORATORY

2.1 The UM3D Lab

The UM 3D virtual reality laboratory includes an immersive virtual-reality like environment, measuring 10 ft (3.048 m) in width, depth, and height. It runs on a cluster of six workstations, with one control computer, one motion-tracking computer, and four rendering computers. The renderers are Box Tech Workstations, with quad-core CPUs at 2.6 GHz, 8 GB RAM, and NVIDIA Quadro FX 5600 + GSync graphics cards. Four Christie Mirage S+4K projectors produce 3D images on the left, front, right, and floor surfaces. The resolution per surface is 1024 x 1024 pixels. The stereo mode is frame-sequential (alternating left-right) at 96 frames per second. Infrared emitters synchronize Stereo Graphics Crystal Eyes® liquid crystal shutter glasses with the projectors. A Vicon MX13 system with eight 1.3 megapixel cameras provides wireless (near infrared) motion-tracking of the shutter glasses and a Logitech Rumble Pad game controller. The sound system comprises four Klipsch speakers mounted in the upper corners, a Klipsch subwoofer on the floor a short distance away, and two amplifiers at 100 watts per channel. The software is an ongoing in-house development, named Jugular, that integrates several open-source, proprietary, and custom-developed subsystems for graphics, sound, animation, physics, motion-tracking, data management, and networking. **[9, 10].**

2.2 The Virtual Reality Modeling Language (VRML) & Interactive Audio

The VRML is representative of simulation capabilities in the Virtual Environment. VRML version 2, also known as VRML97, was adopted as an International Standard ISO/IEC 14772 in 1997. The standards specify a file format, a content model, and algorithms for its interpretation. The model is a directed acyclic graph that includes nodes for geometry, color, texture, and light, as well as sound. However, it provides for only one texture per shape and one pair of texture coordinates per vertex. VRML version 2 was amended in 2002 to add geospatial and NURBS support, but the shape, material, lighting, and sound specifications remain unchanged [11]. FMOD is a programming library and toolkit for the creation and playback of interactive audio. It supports all leading operating systems and game platforms. FMOD is a proprietary audio library made by Firelight Technologies that plays music files of diverse formats on many different operating system platforms, used in games and software applications to provide audio functionality within the 3D virtual Laboratory. The FMOD sound system has an advanced plug-in architecture, that can be used to extend the support of audio formats or to develop new output types, e.g. for streaming. FMOD sound system now contains three main parts: 1) FMOD Ex, the low-level sound engine, 2) FMOD Event System, more abstract, higher level application layer to simplify playback of content created with FMOD Designer, 3) FMOD Designer, the sound designer tool used for authoring complex sound events and music for playback [4,11,19]. Figure 1 shows the computer model and real views of the virtual laboratory

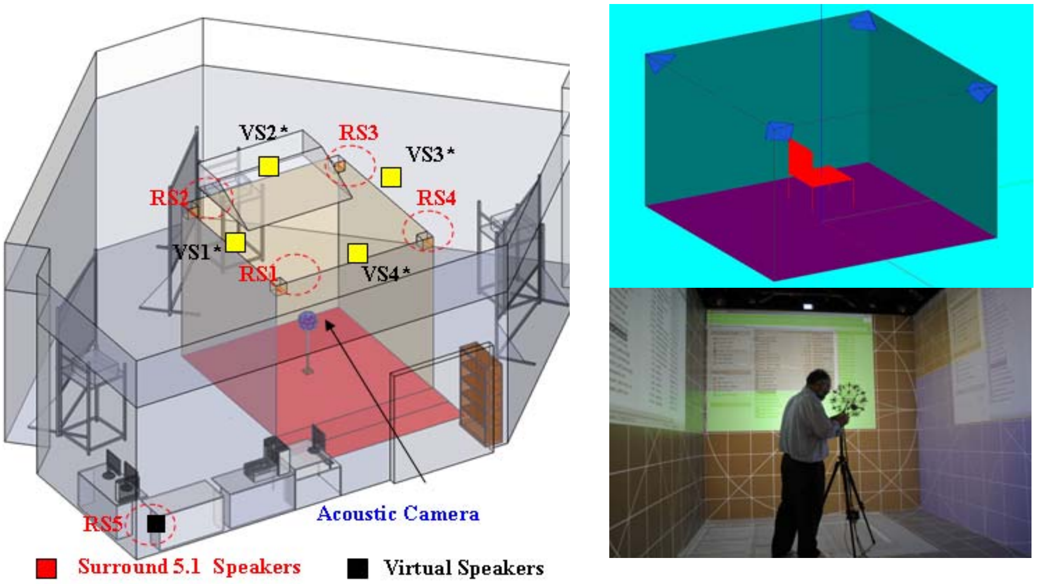


Figure 1: Schematic, computer model and real views of the virtual lab's speakers and their locations

2.3 Principle of the Delay-and-Sum Beamformer

The time domain calculation of a delay-and sum beamformer within the Noise image software is achieved by the use of the equations 1 and 2 for the reconstruction of the time function at every location seen by the acoustic camera microphones' filed of view.

$$\hat{f}(\mathrm{x},t) = \frac{1}{M}\sum_{i=1}^{M} w_i\, f_i\left(\mathrm{x},(t-\Delta_i)\right)$$ **Equation -1**

Effective value of p at location **x**:

$$\hat{p}_{eff}(\mathrm{x}) \approx \hat{p}_{eff}(\mathrm{x},n) = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1} \hat{f}^{\,2}(\mathrm{x},t_k)}$$ **Equation -2**

where x is the location of a point and t denotes the time and M is the number of the microphones in the sensor array. The **fi (t)** are the recorded time functions of the individual microphones, and the **Δi** are the appropriate relative time delays, which are calculated from the absolute run times **Δi=** by subtracting the minimum over all **Ti** . The symbol c denotes the speed of sound in air and **|ri|=|Xi -X|** is the geometrical distance between the spatial position of microphone number **i** and the actually calculated focus point **x**. Despite its extreme simplicity, the delay-and-sum method in the time domain is quite robust and powerful and has shown its practical usability in an extraordinary wide range of acoustic localization and troubleshooting applications for several years **[7, 15, 16]. Figure 2 shows** basic principle of the delay-and-sum beamformer in the time domain.

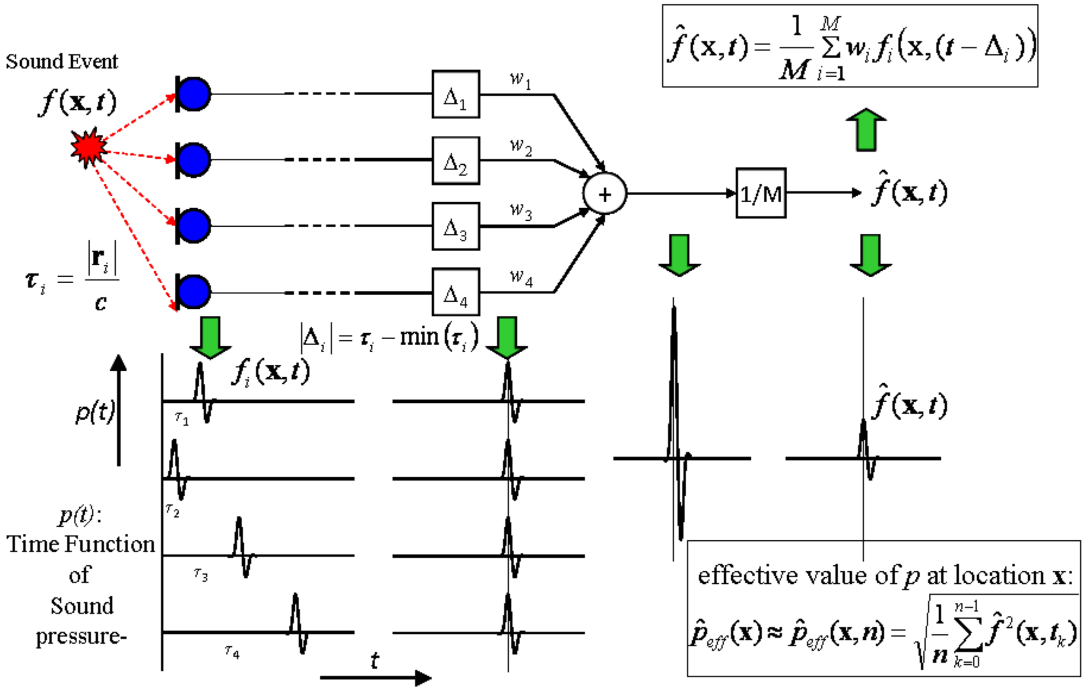
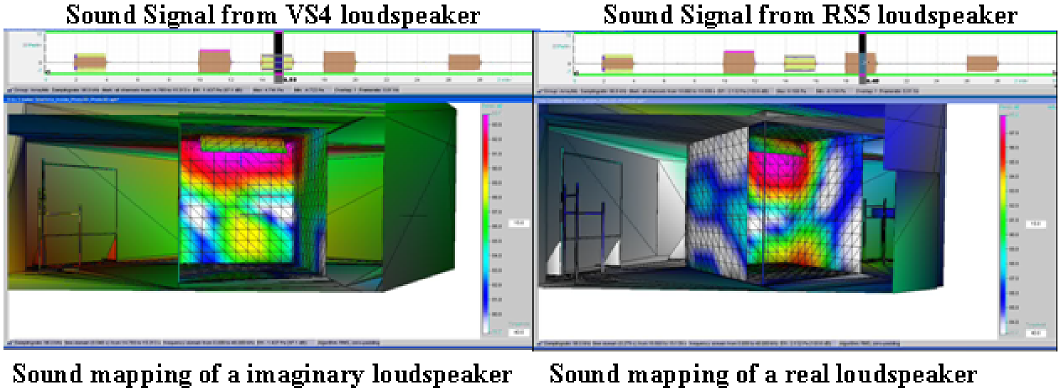

$\hat{f}(\mathbf{x},t) = \frac{1}{M}\sum_{i=1}^{M} w_i f_i(\mathbf{x},(t-\Delta_i))$
Sound Event
$f(\mathbf{x},t)$
$\tau_i = \frac{|\mathbf{r}_i|}{c}$
Δ_1
Δ_2
Δ_3
Δ_4
w_1
w_2
w_3
w_4
+
1/M
$\hat{f}(\mathbf{x},t)$
$|\Delta_i| = \tau_i - \min(\tau_i)$
$f_i(\mathbf{x},t)$
p(t)
τ_1
τ_2
τ_3
τ_4
p(t):
Time Function
of
Sound
pressure-
t
$\hat{f}(\mathbf{x},t)$
effective value of p at location x:
$\hat{p}_{eff}(\mathbf{x}) \approx \hat{p}_{eff}(\mathbf{x},n) = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}\hat{f}^2(\mathbf{x},t_k)}$

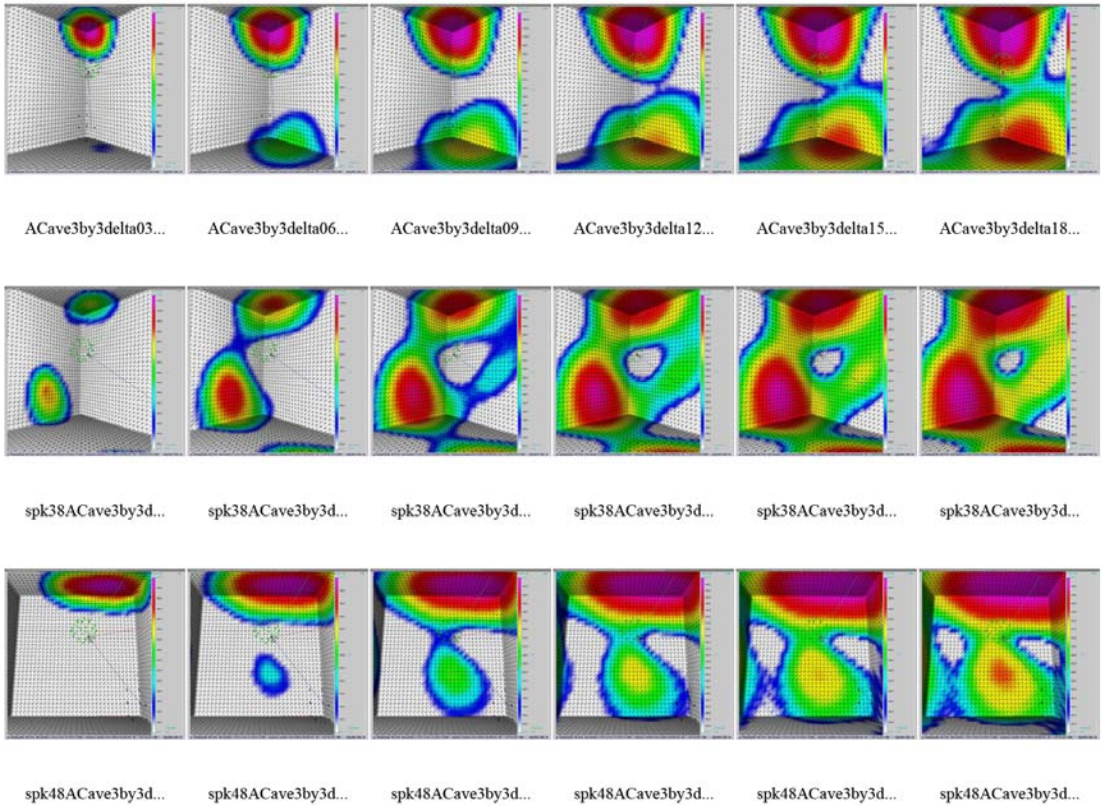
Figure 2: Basic principle of the delay-and-sum beamformer in the time domain.

The output devices such as speakers at a distance from the listener, or headphones impact the user experience with "3D Sound". Passive "3D Sound" where the results always sound the same, the sound spatialization information can be used within a sound playback file, and heard through a stereo amplifier with speakers or headphones. Interactive "3D Sound" situations where the listener's apparent position within the audio scene is required, the "3D Sound" rendering algorithms should be done on listener equipment in real time. The CAVE's sound system utilizes the FMOD and it is known that the system in its speaker mode will never be as good as the headphone mode for "3D sound" rendering. Since the use of HRTF rendering algorithm must interpolate for locations in between the given HRTF locations, it requires high real time computing **[4, 10]. Figure 3** shows sound mapping representation for two selected speakers.


Sound Signal from VS4 loudspeaker **Sound Signal from RS5 loudspeaker**

Sound mapping of a imaginary loudspeaker **Sound mapping of a real loudspeaker**

Figure 3: Sound mapping representation for two selected speakers

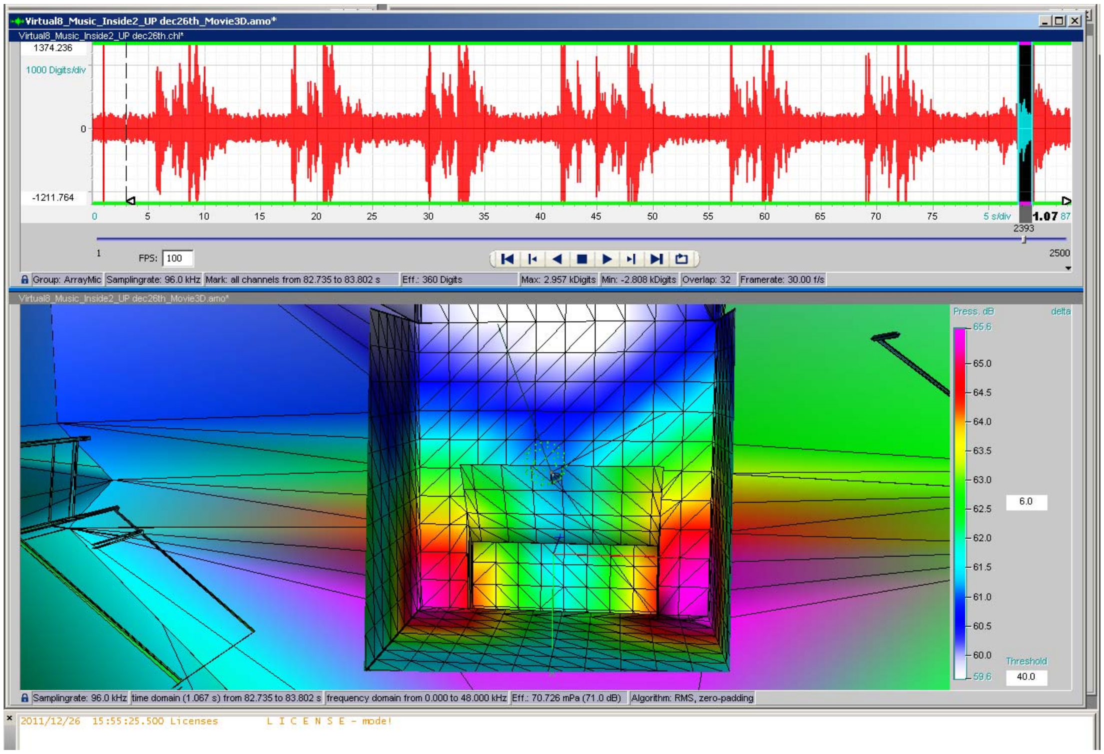


ACave3by3delta03... ACave3by3delta06... ACave3by3delta09... ACave3by3delta12... ACave3by3delta15... ACave3by3delta18...

spk38ACave3by3d... spk38ACave3by3d... spk38ACave3by3d... spk38ACave3by3d... spk38ACave3by3d... spk38ACave3by3d...

spk48ACave3by3d... spk48ACave3by3d... spk48ACave3by3d... spk48ACave3by3d... spk48ACave3by3d... spk48ACave3by3d...

Figure - 4: The sequence of Real Speaker 3(RS3) and Virtual Speaker 4(RS4)



FPS: 100

Group: ArrayMic | Samplingrate: 96.0 kHz | Mark: all channels from 82.735 to 83.802 s | Eff.: 360 Digits | Max: 2.957 kDigits | Min: -2.808 kDigits | Overlap: 32 | Framerate: 30.00 f/s

delta: 6.0

Threshold: 40.0

Samplingrate: 96.0 kHz | time domain (1.067 s) from 82.735 to 83.802 s | frequency domain from 0.000 to 48.000 kHz | Eff.: 70.726 mPa (71.0 dB) | Algorithm: RMS, zero-padding

```
2011/12/26  15:55:25.500 Licenses        L I C E N S E - mode!
```

Figure - 5: Sound output from real speaker 3 and 5 to create virtual speaker 4(RS4).

3.0 HUMAN SUBJECT EXPERIMENTAL SET UP

The subjects were free to move and turn in a virtual space to analyze the effect of various noise factors (e.g., computers, projectors and HVAC system within the space) which impact the sound stimulus and hence their directional cues. This takes place within a simulated acoustics environment while the recorded sound of a Mozart's string quartet piece is played through 4 real speakers. The FMOD system provided or simulated the additional 4 virtual speakers between the real speakers. The time spent from starting to ending position, and trajectory of the subject's motion, were recorded digitally using a digital camera equipped with an equal distance projection fisheye lens and a video recorder. Specific observations were made on each subject's ability to locate or localize the real or virtual source by looking and aiming the "crosshair" indicator closest to each speakers location.

A head-tracking calibration feature (HTCF) that projects the transverse, sagittal, and coronal planes of the user's head (centered between the eyes) was used during the experiment. The intersections of the transverse and sagittal planes with the display planes produce a "crosshair" that indicates the user's center of vision. This was used to determine the best estimate of the locations of the real or virtual speakers as identified by the subjects (e.g., **-** Left and + Right of the speakers in **centimeter**). Additionally, HTCF projects "diamond" quadrilaterals from the head center perpendicularly onto each display plane. When the user looks directly at a display, without head rotation, the crosshairs should center in the diamonds. The head-tracking system is calibrated accordingly. See **Figure 6** for the subject field of view and direction of gaze for locating the real or virtual speakers.

In the experiment, instructions were given aurally. In the beginning of the experiment there were intentionally no rehearsal tests to helped the experiment so that the subjects would not notice or recognize the locations of the real loudspeakers within the virtual simulated environment. A complete test set included 8 speakers (4 real and 4 virtual) and each source played an identical WAVE file within the virtual space. These 8 positions for all speakers are shown in **Figure 1.** as virtual speakers (VS1,VS2,VS3,VS4) and the real loudspeakers as (RS1, RS2, RS3, RS4) in the real space.

3.1 Directing the Subjects

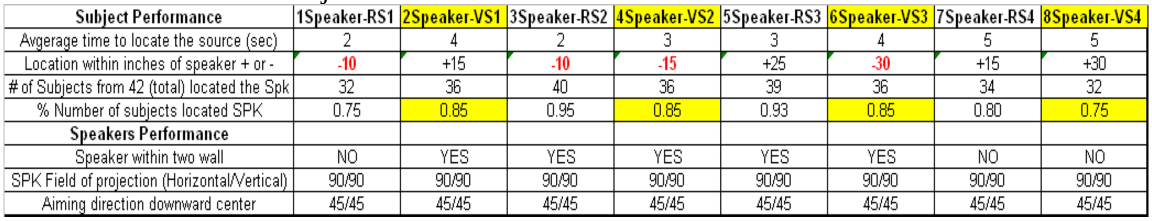
Moving in the virtual space was controlled with the use of Jugular software and a motion tracking system. The subject was able to move forward, backward, and to turn left and right at normal human speed. When subject assumed that he/she has found the sound source, he/she indicated that by stopping at the final location. This experiment was done in the horizontal and vertical plane given the 45 degree downward tilt of angle by the speakers aimed toward the center of the space at 5 feet height. The sound source was a point source. The target area was a one meter (~3.3 feet) sphere around the source. Starting positions were in order from speaker 1 to 8 and the approximate or exact locations were identified by all subjects. The entire experiment did not exceed 5 minutes per subject. WAVE files were displayed or simulated within the virtual environment using the FMOD sound system **[10]**.

4.0 RESULTS

Stimulus, panning method for localizing the sound within the acoustic environment was used to conduct this experiment. Each WAVE file had equal loudness. Each speaker cycle of time play was about 8 seconds long including one second dead band as played in a sequential loop for all 8 speaker positions. The sound source had an Omni-directional pattern. The synthesized or aurulized sections of the sound were produced by a physical-based model **[17, 18].** Panning or seeking for the sound source methods had no limitation on subject movements. The interaural time difference (ITD), was included as an auditory cue to all test conditions. The ITD was calculated from a spherical head model within EASE program **[19]** and implemented with a short delay line. When the subject asked to locate the sound , they started at the center of the space facing the center wall with full degree of freedom to move within the virtual space.

Once the music started to play, subjects were asked to seek and locate the sound and its direction, Subject movement and recognition of the source location were recorded using a camera and a video recorder. The results and sound distribution within the space were also measured using an acoustic camera in absence of the subject. The summery of the findings is shown in **Table 1.**

Table - 1: The subjective evaluation results within the simulated environment.



Subject Performance	1Speaker-RS1	2Speaker-VS1	3Speaker-RS2	4Speaker-VS2	5Speaker-RS3	6Speaker-VS3	7Speaker-RS4	8Speaker-VS4
Avgerage time to locate the source (sec)	2	4	2	3	3	4	5	5
Location within inches of speaker + or -	-10	+15	-10	-15	+25	-30	+15	+30
# of Subjects from 42 (total) located the Spk	32	36	40	36	39	36	34	32
% Number of subjects located SPK	0.75	0.85	0.95	0.85	0.93	0.85	0.80	0.75
Speakers Performance								
Speaker within two wall	NO	YES	YES	YES	YES	YES	NO	NO
SPK Field of projection (Horizontal/Vertical)	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90
Aiming direction downward center	45/45	45/45	45/45	45/45	45/45	45/45	45/45	45/45

The best results in virtual sound source localization are achieved if and only if the binaural difference-based cues, and the monaural and binaural cues that arise from the scattering process of the Head Related Transfer Function (HRTF) for the particular individual are included within the virtual scene. If the primary output device is a pair of headphones or speakers, the "3D Sound" rendering process will result in a stereo signal, e.g. one signal for the "Left ear" and one signal for the "Right ear", but the signal components tend to cancel cross talk effects between Left and Right channels. The effectiveness to which this can be done varies both with distance from the listener, as well as distance (or angle) between the speakers. The effect is best with speakers with the angle range of 27-45 degrees between the listener and speaker on either side of the listener when facing forward. **[8]**. Measured results for different viewing conditions of the surfaces within the virtual laboratory provide some clue to the incoming sound and its directions. **Figure 4** shows the sequence of real speaker 3(RS3) and virtual speaker 4(RS4). **Figure 5** shows sound output from real speaker 3 and 5 to create virtual speaker 4(RS4). The sequence of the speakers and the sound intensity distribution in terms of calculated Delta dB within each viewing scene for each real and virtual speaker provide an insight into the sound projected from each speaker. These results show the impact of the VR screen on sound distribution within the space and as to how it is perceived by the human subjects with respect to each speaker (real and or virtual) as indicated. The measured data were viewed and examined using the Noise Image software with and without the HRTF as a filter for the sound source directionality as it was perceived by the subject at the center of the space. The measurement and simulation path within real and virtual environment for all experimental procedures are shown in a flow chart diagram including the steps for HRTF filtering Noise Image software within **Figure 6.**

5.0 DISCUSSIONS

The results indicate that the inter-reflection and or the missing 4th wall (back wall) within the virtual environment has impacted the subject sound localization of the sound source due to the loss of reflections. The exact location of each speaker and the percent of the subject identified the locations of the speakers within positive or negative distance shows the impact of such condition. See 3rd and 5th rows within **Table 1.** These results in positive or negative distance (cm) from the center and or exact location of the speakers, validates the hypothesis that the wall surfaces do contribute to the interaural time difference (ITD). However, parametric tests for examining the interactions between large numbers variables require a larger number of subjects. This experiment include only 42 subjects.

6.0 CONCLUSION AND FUTURE WORK

It is important to acknowledge that error rate and large variation in times between the test subjects for locating the real and virtual source need to be fully examined. It is worth noting the percentage of the subjects and the times it took them to identify the real and virtual speakers and the correlation between the location and the missing back wall. One explanation might be the way the test subjects started and the sequence of the speakers as each subject tried to locate each target as well as possible, without caring how much time they spent. The application of the HRTF would provide the future opportunity to examine this condition more carefully for room acoustic design and architectural application. The results of this work show it is possible to navigate within a virtual environment while using even limited auditory cues. The subjects completed the navigation tasks, and had no difficulty to localize the sound given all available architectural or room acoustic characteristics for the stimulation within this virtual environment. The findings of this study were similar to other past studies using simple models of spatial hearing given enough cues for auditory navigation and equal perception of the sound for close to real space. This approach, as shown in Figure 6, allows one to experience the reverberation within the simulated space as well as validating the HRTF used within various software for sound auralizations. In the future we will make more listening tests for the analysis of interactions between two or more statistical variables for auditory navigation in virtual environments as well as testing within a true 3D spaces.

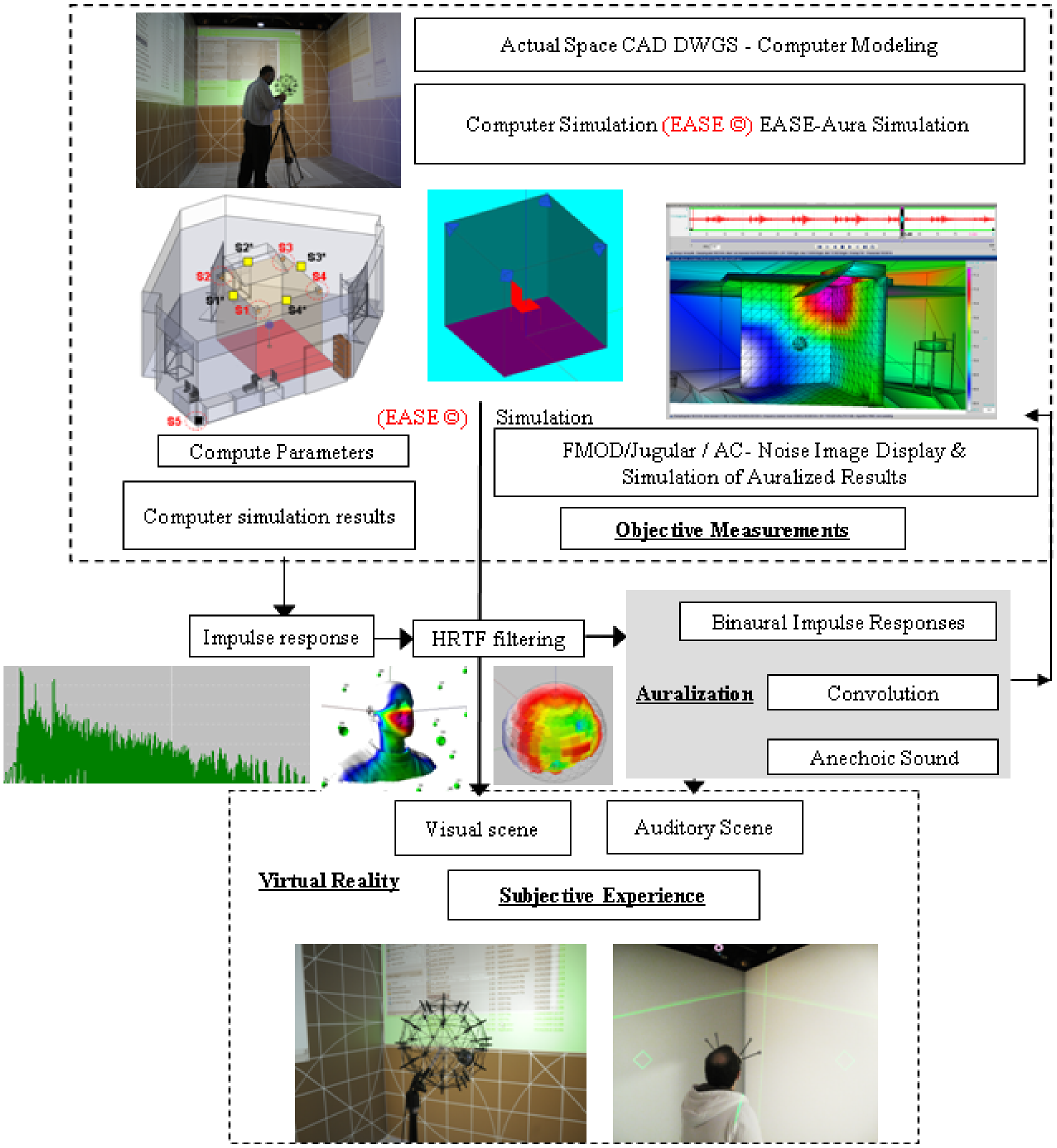

Actual Space CAD DWGS - Computer Modeling
Computer Simulation (EASE ©) EASE-Aura Simulation
S2'
S3
S3'
S2
S4
S1'
S1
S4'
S5
(EASE ©)
Simulation
Compute Parameters
FMOD/Jugular / AC- Noise Image Display & Simulation of Auralized Results
Computer simulation results
Objective Measurements
Impulse response
HRTF filtering
Binaural Impulse Responses
Auralization
Convolution
Anechoic Sound
Visual scene
Auditory Scene
Virtual Reality
Subjective Experience

Figure - 6: Measurements and simulation path within real and virtual environment.

REFERENCES

[1] Loomis, J., Golledge, R., and Klatzky, R. Navigation system for the blind: Auditory display modes and guidance. Presence: Teleoperators and Virtual Environs 7, 2 (Apr 1998), 193–203.

[2] Rutherford, P. Virtual acoustic technology: Its role in the development of an auditory navigation beacon for building evacuation. In Proc. 4th UK Virtual Reality SIG Conference (London, UK, 1997), R. Bowden, Ed., Brunel University.

[3] Begault, D. 3-D Sound for Virtual Reality. Academic Press, Cambridge, MA, 1994.

[4] Hall Ted W., Navvab, M., 2011, Ch. 11, PART III: "Virtual Reality as a Surrogate Sensory Environment. Book by Tauseef Gulrez“ Advances in Robotics and Virtual Reality", Springer's book/978-3-642-23362-3, http://www.springer.com/engineering/computational + intelligence+and+complexity/book/978-3-642-23362-3

[5] G. Heilmann, A. Meyer, D. Döbler: Time-domain beamforming using 3Dmicrophone arrays, Proceedings of the BeBeC 2008, Berlin, Germany, 2008

[6] Don H. Johnson, Dan E. Dudgeon, “Array Signal Processing,”1993 by PTR Prentice- Hall, Inc.

[7] Dirk Döbler and Gunnar Heilmann, “Perspectives of the Acoustic Camera,” The 2005 Congress and Exposition on Noise Control, August 2005.

[8] Hartmann, W. M. (1999). "How We Localize Sound," in Physics Today, pp. 24-29.

[9] Beier KP (2000) Web-Based Virtual Reality in Design and Manufacturing Apps.
In: Hansa International Maritime Jour 137/5:42–47

[10] http://www.fmod.org/index.php/fmod

[11] ISO/IEC (1997) The Virtual Reality Modeling Language (ISO/IEC 14772¬1:1997). ISO, Geneva http://www.web3d.org/x3d/specifications/vrml/ISO-IEC-14772-VRML97/

[12] Howard, D. M., and Angus, J. A. S. (2006). Acoustics and psychoacoustics (Focal Press, Amsterdam).

[13] Foley, J. D, van Dam, A. Feiner, S., Comp Graphics, Principles Practice, 2nded. (Addison-Wesley, 1992).

[14] O. Jaeckel: Strengths and Weaknesses of calculating Beamforming in the time domain, Proceedings of the BeBeC 2006, Berlin, Germany, 2006

[15] Any. Meyer, D. Döbler, "Noise source localization within a car interior using 3D microphone arrays, Proceedings of the BeBeC 2006, Berlin, Germany, 2006

[16] Andy Meyer, Dirk Döbler, Jan Hambrecht, Manuel Matern , "Acoustic Mapping on three-dimensional models", International Conference on Computer System & Technologies, 2011.

[17] Valimaki, V. Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters. Doctoral thesis, Helsinki University of Technology, Lab. of Acoustics and Audio Signal Processing, Report 37, 1995. Available at http://www.acoustics.hut../ -vpv/publications/vesa phd.html.

[18] Tapio Lokki, Matti Gr¨ohn, Lauri Savioja, and Tapio Takala, A Case Study of Auditory Navigation in Virtual Acoustic Environments, IEEE, 2000, **ISSN:** 1070-986X http://ieeexplore.ieee.org/servlet/opac?punumber=93

[19] Enhanced Acoustic Simulator for Engineers - E.A.S.E. Software- Acoustic modeling , http://afmg.eu/index.php/company.html.