BNAM2012
Joint Baltic-Nordic Acoustics Meeting
June 18th - 20th 2012 Odense, Denmark

Spatialisation accuracy of a Virtual Performance System

Iain Laird, Dr Paul Chapman,
Digital Design Studio, Glasgow School of Art, Glasgow, UK, I.Laird1@gsa.ac.uk, p.chapman@gsa.ac.uk

Dr Damian Murphy
Audio Lab, Department of Electronics University of York, York UK, dtm3@ohm.york.ac.uk

A Virtual Performance System (VPS) is a real-time 3D auralisation system which enables a musician to play in simulated acoustic environments. Such systems have been used to investigate the effect of stage acoustics on the performance technique of musicians. This article describes how a VPS can be calibrated using omnidirectional, energy-based quantities. It goes on to describe how the spatialisation accuracy of specific reflections can be assessed using time-frequency analysis. The energy-based calibration has been shown to produce a reasonably accurate simulation of a test space in terms of known acoustic metrics such as T30. However, the time-frequency analysis has shown significant differences in the spatialisation of particular early reflections in the virtual version of a test space.

1 Introduction

The acoustic design of a stage is a critical part of constructing a successful performance space. It has been demonstrated that musicians will often adapt their playing technique (sometimes unconsciously) in reaction to the acoustic response they hear on stage [1]. When studying the effect of stage acoustics on performers a primary concern is the venue in which the study takes place. It is preferable to study musicians playing in actual concert hall environments however hall hire costs and the lack of controllable variables make many experiments limited or impractical. It is therefore necessary to consider a laboratory set up which can emulate the 3D soundfield experienced by a performer playing on stage in response to the sound of their instrument in real-time.

Gade [2], Ueno [1] and Brereton [3] have previously conducted studies in virtual versions of concert halls which utilise a 3D auralisation system to recreate the sensation of playing on stage for a number of test subjects. While Gade's virtual acoustic environment was created using a combination of delay lines and loudspeaker-based reverberation, recent systems [1, 3] rely on convolution engines to apply a measured or modelled room response to the direct sound of the musical instrument which is then played back to the musician in real-time.

Systems such as this however remove the musician from their natural performing environment and place them in what is often reported to be an unnatural sounding approximation of stage acoustics. It is therefore crucial that the VPS can be verified to ensure the simulated acoustic environment is a fair representation of the target space.

This article describes how a time-frequency directional analysis was used to check the spatialisation accuracy of a calibrated VPS system. It begins by describing how the VPS is constructed and then adjusted to give an accurate level of acoustic response using energy-based quantities. It then goes on to assess how well the system is performing in relation to known acoustic metrics and also reflection spatialisation accuracy.

1.1 Stage acoustic parameters

Gade's initial studies into musician perception of stage acoustics highlighted Support and Reverberance as being two key concerns for soloists [2]. These two features of stage acoustics have since been quantified as described below.

Support (ST) quantifies the energy in an impulse response (in dB) within certain time intervals with reference to the direct sound. It is often used to determine how the energy within these time regions assists a performer's own efforts. Two main ST quantities exist which each refer to different time regions of an impulse response and are thus linked to assessing different perceptual effects.

$$ST_{early} = 10log_{10}\left(\frac{\int_{20ms}^{100ms} h^2\,(t)dt}{\int_{0ms}^{10ms} h^2\,(t)dt}\right) \tag{1}$$

$$ST_{late} = 10log_{10}\left(\frac{\int_{100ms}^{1000ms} h^2\,(t)dt}{\int_{0ms}^{10ms} h^2\,(t)dt}\right) \tag{2}$$

Where *h* denotes the sound pressure of the impulse response measured on stage.

ST_{early} (1) is generally used to describe the degree of mutual hearing in ensembles while ST_{late} (2) is indicative of the degree to which the reverberant sound supports the musician's own efforts. A standard measurement methodology is defined in ISO 3382 [4] however there remains ongoing discussion as to the reliability of ST parameters with many authors suggesting various alternatives for integration limits and source-receiver positions [5].

Reverberance can be quantified using standard reverberation time quantities such as T30 and EDT. EDT has been found to be linked to the perception of reverberance in a space [4]. It determines reverberation time (RT60) based on a linear regression of the time taken for the acoustic response to attenuate by 10dB based on the Schröeder curve obtained from a corresponding impulse response. T30 similarly defines RT60, now based on the first 30dB of attenuation and is commonly used in generic reverberation time tests.

2 SoundLab and VPS

A VPS system was constructed in the Arup/DDS SoundLab situated in Glasgow, UK. The SoundLab is a dedicated auralisation suite used for research and commercial purposes. The space is 4m (l) x 6m (w) x 2.5m (h); heavily acoustically treated and is floated on springs in order to acoustically isolate the space. The SoundLab has a measured L_{AF90} (level exceeded for 90% of measurement time) of approximately 20dBA with all equipment in the SoundLab switched on over a measurement period of 5 minutes. The average T20 of the SoundLab at 500Hz is approximately 0.15s. The SoundLab features a 12-channel periphonic loudspeaker array comprising of Yamaha MSP5A loudspeakers in addition to a Tannoy TS12 subwoofer. The speaker system has been equalised to ensure an optimum frequency response at the sweetspot and an equal level contribution measured at the sweetspot.

The VPS uses a microphone placed near the musician's instrument to pick up the direct sound of the instrument. The sound is convolved with an ambisonic [6] impulse response which can be measured using an ambisonic microphone or modelled using acoustic modelling software. The convolution is achieved using two Reaverb convolution plug-ins hosted in the Reaper Digital Audio Workstation (DAW) [7]. The resultant auralised material is then decoded using a Decopro ambisonic decoder [8] and played back in real-time to the musician sat in the centre of the speaker array.

Previous research [9] has found that it is not practical for the VPS to reproduce the direct sound or 1st reflection from the floor, as these elements are present naturally when a musician is playing their instrument in any space. Furthermore, any processing latency restricts the system's ability to accurately reproduce reflections within a certain time frame therefore these elements are discarded from the impulse responses used in auralisation. The resulting redundant propagation time is truncated by the appropriate amount in order to reduce overall system latency. It was also demonstrated in [9] that the level of acoustic response should vary in accordance with the level of the soloist's instrument and should not auralise quiet sounds to the same degree as louder sounds. Errors in the level of simulated acoustic response can contribute to a noticeable 'PA Effect' whereby the musician feels as if they are playing through a PA system in a space rather than acoustically.

3 Experiment

An experiment was devised which aimed to test the spatialisation accuracy of a VPS after it had been calibrated. First order ambisonic impulse responses were captured in a test space (described in Section 3.1) from a nearby loudspeaker using well known Swept Sine Wave techniques [10]. This data was used to create a virtual version of the space which was then calibrated using an omnidirectional energy-based quantity (described in Section 3.2). After calibration, an identical survey was undertaken in the VPS including measuring impulse responses at the SoundLab sweetspot with the VPS running. A time-frequency domain directional analysis space (described in Section 3.3) was obtained from both the measured and virtual impulse responses and then compared.

ST_{early} and ST_{late} were used in this experiment as calibration parameters rather than to assess musician perception of acoustic support. It provides a convenient quantification of the energy in particular time regions of the acoustic response in reference to the direct sound which can be easily recreated in the VPS. A number of adjustments to the standard measurement methodology were made, mainly relating to source directivity and source-receiver positions. Thus in order to distinguish these quantities from ST_{early} and ST_{late} they are renamed E_{early} and E_{late} respectively.

3.1 Impulse response capture in test room

A logarithmic sine sweep was generated in Matlab and output as a 32-bit floating-point, 44.1kHz wave file with a duration of 10s, swept from 1Hz - 22050Hz. The inverse sweep was also generated to enable extraction of the impulse responses from the recorded signal via convolution [10].

A Genelec 1029A loudspeaker was placed on a tripod in the test space at a height of 85cm above the floor. A Soundfield ST350 microphone was placed 20cm directly behind the loudspeaker at a height of 128cm from the floor. This source-receiver configuration was set to approximate a geometric simplification of a seated wind musician. The Soundfield Microphone and loudspeaker were connected to a soundcard (Behringer ADA8000 connected to an M-Audio Profire Lightbridge) and Macbook running the Reaper DAW which was used to play back the sine sweeps. The sine sweeps were played at the same level in the test room and virtual room at an L_{AFMAX} of 106.0dB (Maximum A-weighted SPL, Fast time weighting). This level was measured in the SoundLab using a B & K 2231 Sound Level Meter (SLM) positioned 10cm away from the loudspeaker.

The test room was a large empty office unit on the 3rd floor of the Hub building at Pacific Quay in Glasgow. The space is rectangular in construction and measures approximately 53m (l) x 14m (w) x 4m (h). The space has a stainless steel floor and a large number of windows on each wall; the ceiling is also of steel construction with exposed building services over the full area. Small columns are located in the centre of the space along its length. The measurement equipment was set up approximately 8m from the North wall and 4m from the nearest column. The background noise in the space was measured with the SLM over a period of 15 minutes and was found to give an L_{AF90} of 39dB.

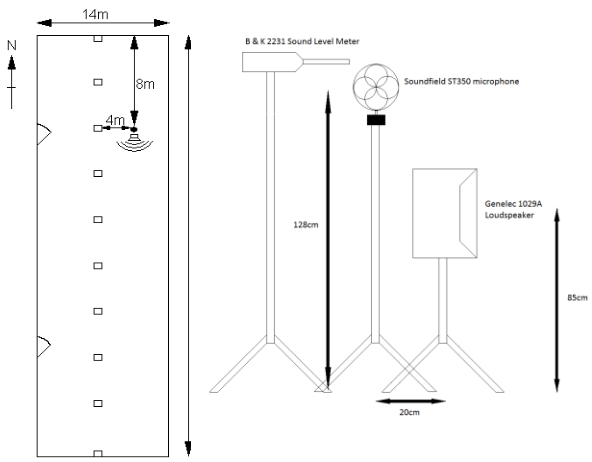

Figure 1: Plan of test room on 3rd Floor of the Hub Pacific Quay, Glasgow. Source and receiver position and orientation are also shown.

3.2 Construction and calibration of VPS

The impulse responses were edited to remove the direct sound and floor reflection and then inserted into two Reaverb convolvers hosted in Reaper. The system soundcard was set to the lowest stable I/O buffer setting which was found to be 256 samples resulting in a processing latency of approximately 12ms.

The equipment set up described in 3.1 was reproduced in the SoundLab ensuring that the Soundfield microphone was positioned in the sweetspot of the ambisonic array. A Behringer ECM8000 measurement microphone was placed 10cm in front of the loudspeaker to feed the direct sound of the loudspeaker into the VPS. The impulse response with the highest signal to noise ratio obtained in the test room was edited as described and imported into two stereo Reaverb convolvers holding W, X, Y and Z channels of the ambisonic impulse response. Sine sweeps were played through the loudspeaker at the same SPL used in the test room. The auralised response to this signal was recorded at the sweetspot by the Soundfield microphone and the impulse responses derived as before.

By comparing the time of arrival of specific reflections in the impulse responses measured in the test room with those measured in the SoundLab it was possible to determine how much of the propagation time to discard in order for the reflections to occur at the correct time. Once aligned, the gain of the decoder output was adjusted to obtain the closest possible match in E_{early} and E_{late} measured in the test space. The results for this calibration are shown in Table 1.

Table 1: E_{early} and E_{late} readings for calibrated VPS

	Test space (dB)	Virtual space (dB)
E_{early}	-7.34	-6.24
E_{late}	-8.49	-8.79

3.3 Directional Analysis

A directional analysis based on Merimaa and Pulkki's Spatial Impulse Response Rendering (SIRR) algorithm [11] was carried out on the test room and virtual room impulse responses. The directional analysis analyzes the direction of arrival and diffuseness of impulse responses measured in the time-frequency domain from an ambisonic microphone. This is obtained by estimating the active intensity (equation 3) in the time-frequency domain via the Short Time Fourier Transform (STFT). The active intensity describes the net flow of energy for a particular time-frequency element and can thus be used to indicate the direction of arrival for a particular reflection in a certain frequency band. This type of analysis can be performed on impulse responses obtained with an ambisonic microphone by using the four W, X, Y and Z channels. The results obtained in this experiment are in the horizontal plane only and therefore use the W, X and Y channels in the analysis.

$$I_\alpha(\omega) = \frac{\sqrt{2}}{z} Re\{W^*(\omega)\boldsymbol{X}'(\omega)\} \tag{3}$$

Where:

$$\boldsymbol{X}' = \left(Xe_x + Ye_y\right) \tag{4}$$

W is the sound pressure from the W channel of the ambisonic microphone and z is the characteristic acoustic impedance of air. e_x and e_y represent unit vectors in the directions of the corresponding Cartesian coordinate axes while *X* and *Y* are the sound pressure measured with the X and Y channels of the ambisonic microphone. * denotes complex conjugation.

4 Results

4.1 Reverberation time

Figure 3 compares EDT and T30 in measured and virtual rooms. 1/3rd octave results show that the virtual and test room T30 are highly correlated while EDT contains noticeable deviations particularly at 400Hz and 3.15kHz where the virtual space shows a reduced EDT. The maximum difference was observed to be approximately 2 seconds at 3.15kHz.

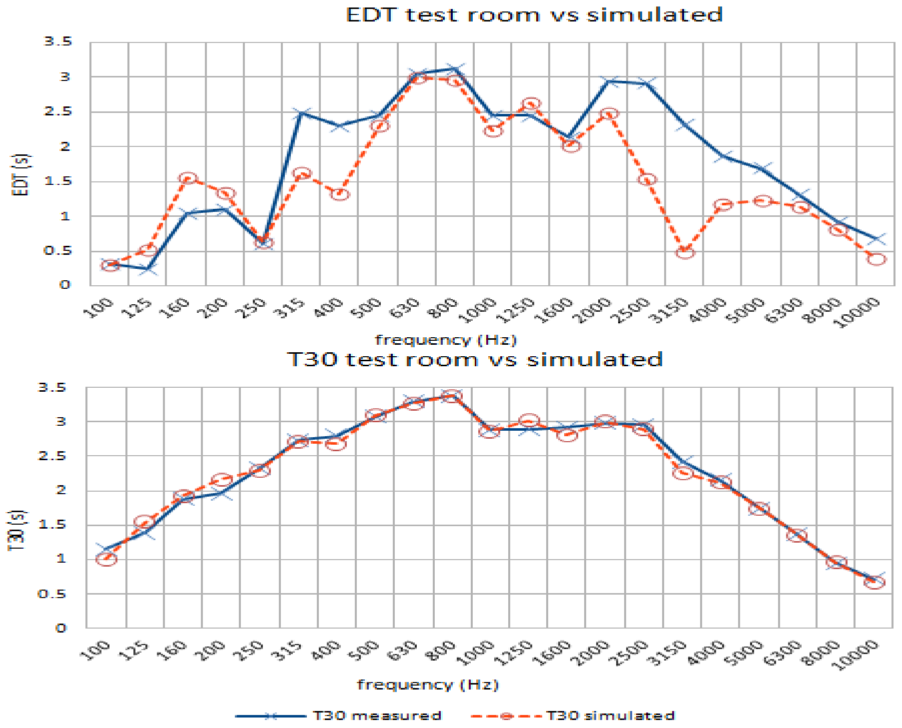

Figure 2: plots comparing (a) EDT and (b) T30 of measured test room (solid) and the virtual version (dashed)

4.2 Directional Analysis

The results in Figure 3 compare directional analyses performed on impulse response made in the test and virtual rooms for the first 100ms of the respective impulse responses. The directional analysis is represented as a time-frequency plot with overlaid active intensity vectors (horizontal plane only) for clarity, the results are shown up to 5kHz only. Accompanying amplitude envelopes are shown adjacent to these plots. The results show a 256 sample STFT with a hop size of 25%.

Overall some similarities between the two responses can be observed as well as significant differences. These have been highlighted on the plots and discussed below. Positive angles are equivalent to an azimuth in the clockwise direction.

<u>Section A</u> In the virtual response a large amplitude reflection is present where it is missing in the measured response. This reflection emanated from straight in front of the receiver and has been found to be caused by a large amplitude reflection from a TV screen which is built into the wall.

<u>Section B</u> The measured response shows a reflection which on average appears at an azimuth of 90 degrees to the receiver especially at high frequencies. The virtual response does not show a reflection at this angle of incidence but still appears to show a dominance to the front of the receiver. This may be as a result of the SoundLab room being dominant in this early time region.

<u>Section C</u> Shows a higher correlation between the measured and virtual reflections in that much of the energy is being received laterally at 90 and -90 degrees. The reflection in the virtual response is noticeably higher in amplitude than in the measured response.

Section D The virtual response shows a trend indicating a reflection emanating from -90 degrees whereas the measured response shows much of the energy approaching from the front with only some lower frequency energy appearing to arrive from -90 degrees.

Section E Shows a distinct reflection in both virtual and measured responses at 60ms. In the measured and virtual responses there is some indication of the reflection arriving from approximately -45 degrees at high frequencies. This trend is much more dominant in the measured response.

Section F shows another distinct reflection which appears to arrive from the front of the listener in the measured response at higher frequencies and behind at lower frequencies. The virtual response shows more of a trend to the rear of the musician.

Section G shows a distinct reflection in both measured and virtual responses which appears to have dominant energy to the rear of the musician in both responses

Throughout sections C to G there is a noticeable reduction in amplitude in the 2kHz – 4kHz region, confirming the differences in EDT shown in figure 2.

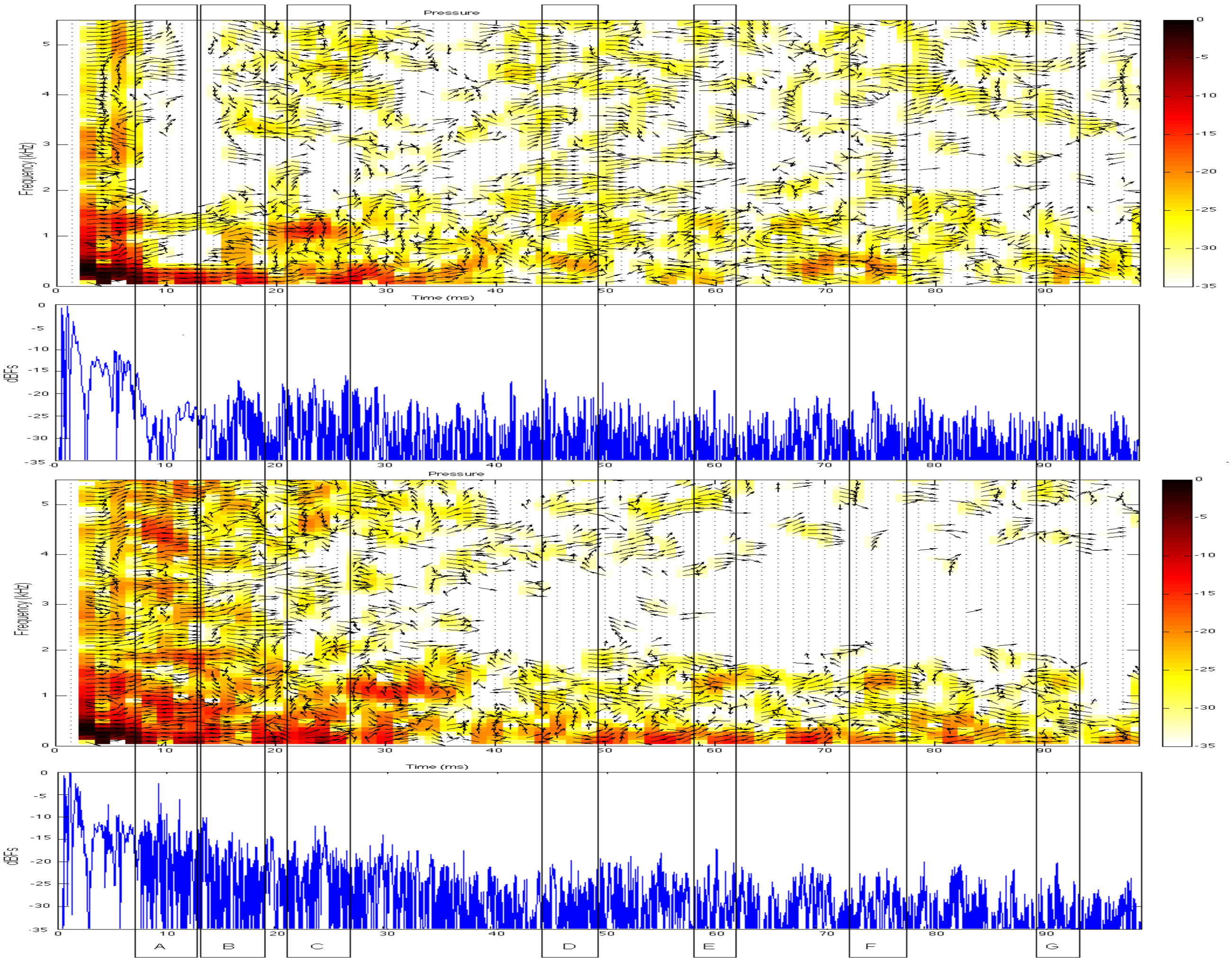

Figure 3: Directional Analysis of measured (top) and virtual (bottom) responses with accompanying amplitude envelope shown in dBFS. Directional analysis shows spectrogram of pressure with overlaid intensity vectors showing direction of arrival. Highlighted sections are discussed

5 Discussion

It can be seen that calibrating the level of the acoustic response in a VPS using energy-based parameters results in a reasonably accurate simulation according to quantities such as T30.

A visual comparison of measured and virtual impulse responses revealed additional reflections present in the first 20ms of the virtual response. This is thought to be caused by reflecting surfaces (TV screens and other equipment) in the SoundLab. It is conceivable that in order to achieve an accurate early response, the VPS should be housed within an anechoic chamber. Further work will investigate how audible the SoundLab response is to performing musicians in the presence of masking noise from their musical instrument.

A visual inspection of the azimuth of particular reflections showed significant differences between measured and virtual responses. It is well understood that 1st order ambisonic systems provide a lower level of spatial resolution in comparison to other 3D spatial audio systems such as Higher Order Ambisonics (HOA) which may have contributed to these differences. Future work will include optimizing the current ambisonic decoder and comparing 1st order ambisonic VPS systems with systems based on other spatial audio techniques.

6 Conclusions

This study has shown that calibrating a VPS using omnidirectional, energy-based quantities results in a reasonably accurate simulation in terms of known acoustic quantities such as T30. However noticeable differences are evident in the early response and EDT values in the virtual response due to the non-anechoic nature of the reproduction space. A directional analysis has further shown that the azimuth of specific reflections are not accurately reproduced using a 1st order Ambisonic System.

7 Acknowledgements

This work was supported by the Digital Design Studio and Arup. Thanks to the Hub staff for allowing access to test spaces.

8 REFERENCES

[1] K. Ueno, K, Kosuke, K Lawai (2007) Musicians' Adjustment of Performance to Room Acoustics Part 1: Experimental Performance and Interview in Simulated Soundfield. 19th International Congress on Acoustics, Madrid

[2] A. C. Gade (1989) Investigations of Musicians' Room Acoustic Conditions in Concert Halls. Part 1: Methods and Laboratory Experiments. Acustica, vol. 69, 193-203

[3] J. Brereton, D. Murphy, D. Howard (2012) A Loudspeaker-Based Room Acoustics Simulation for Real-Time Musical Performance. Proc. Of the 25th UK AES Conference in assoc. with the 4th International Symposium on Ambisonics and Spherical Acoustics

[4] International Organisation of Standardisation (ISO) (2009) BS EN ISO 3382-1:2009 'Acoustics – Measurement of room acoustic parameters'

[5] Gade, A. C. (2010) 'Acoustics for Symphony Orchestras; status after three decades of experimental research' Proceedings of the International Symposium on Room Acoustics, August 2010, Melbourne Australia.

[6] P. Fellgett, Ambisonics. Part One: General System Description, Studio Sound, August 1975, 1:20–22,40

[7] Reaper DAW, http://www.reaper.fm/, accessed Jan 2012

[8] Gerzonic Ambisonic Decoders, http://www.gerzonic.net/, accessed Jan 2012

[9] I. Laird, D. Murphy, P. Chapman, S. Jouan (2011) Development of a Virtual Performance Studio with Application of Virtual Acoustic Recording Methods. Proc. of the 130th AES Convention, London, UK

[10] A. Farina (2000) Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. 108th AES Convention. Paris, France

[11] J. Merimaa and V.Pulkki (2004) Spatial Impulse Response Rendering. Proc. of the 7th International Conference on Digital Audio Effects, Naples, Italy, October 2004.