**2006 IEEE International
Symposium on Signal Processing
and Information Technology**

Spatial Sound Rendering Using Measured Room Impulse Responses

Yan Li, Peter F. Driessen
Dept. of Elec. and Comp. Engineering, University of Victoria
Victoria, BC, Canada
yli@ece.uvic.ca

George Tzanetakis
Dept. of Comp. Science, University of Victoria
Victoria, BC, Canada

Steve Bellamy
Music & Sound, Banff Centre
Banff, Alberta, Canada

Abstract*— *Spatial sound rendering has many applications such as music production, movies, electronic gaming and teleconferencing. Each of the applications may have different quality and complexity requirements. This paper presents a new spatial sound rendering framework that aims at producing realistic multichannel audio while being flexible and scalable so that is can be extended and adopted by various applications. The proposed framework uses multi-channel measured room impulse response (MMRIR) as the basis for building a room acoustic model which is used to synthesize multi-channel audio. The proposed framework has been evaluated by informal listening tests.

Keywords*— *room acoustics, spatial sound rendering, multi-channel measured room impulse responses.

I. Introduction

The goal of spatial sound rendering is to create a virtual auditory environment that is indistinguishable from a real auditory environment. It has a wide range of applications including electronic games, movie/music production, teleconferencing, networked music performance, and audio-based navigation interfaces for the blind. Different application areas have different complexity and quality requirements. For example, music/movie production needs high quality and generally has more available computational power, while computer gaming needs lower quality with highly constrained complexity.

In the past two decades, a significant amount of research has been carried out in the areas related to spatial sound rendering. However, the focus of this research went to two extremes. Traditionally research into spatial sound has focused upon high quality renderings of the spatial environment. Spatial rendering has primarily been based upon geometrical properties of environments, physical properties of objects, and source characteristics, e.g [1]. This approach, whilst very accurate, requires powerful processing resources and is very difficult to achieve in real-time applications [2]. At the other end, a number of real-time rendering systems have been proposed, mainly for the purpose of electronic gaming [3] [4]. These real-time rendering systems are often built on (overly) simplified perceptual or physical model and in turn faintly resemble the physical reality. For example, Creative's EAX uses a Feedback Delay Network (FDN) which can be viewed as a network of multi-channel comb filters [3] and A3D only simulates the first few reflections [4]. In the context of Virtual Aural Reality, a number of projects targeted at rendering "good" quality at reasonable complexity so as to be implemented in real-time have made advances in different areas, for example, DIVA [5] utilizes a parametric RIR rendering method. However, these systems are still built on, or partially built on, the imaginary or theoretical models that may not always reflect the physical reality. Also having to control a parametric model often troubles users who do not fully understand the impact of each of the parameters.

In this paper, we present a new spatial sound rendering system based on the MMRIR with the goal of creating the acoustic impression of a specific venue using a multichannel speaker system. To achieve this goal, we use a hybrid method that models only the direct sound and early reflections individually using the image-source method and synthesizes the late reverberation using a set of filters derived from the MMRIR. Unlike the other solutions, our system is built exclusively on the MMRIR - the true reflection of the acoustic characteristics of the target venue. This paper is organized as follows. Section 2 provides a brief description of techniques we use to measure and analyze the RIR, followed by how we build the room acoustic model using the analysis results in Section 3. In Section 4, we elaborate the system design and implementation, followed by evaluation and discussions.

II. IR Measurement and Analysis

The acoustics of a reverberant space add feeling and life to music. Many concert halls are famous for their sound quality and many recording artists go to great lengths and cost to record live performances at these venues, in order that the listener can experience the concert hall surroundings in their own living room. Applying the acoustic response of a concert hall to music recorded in a studio would save the industry a

lot of money and also allow the same piece of music to be experienced at different venues [6].

The ultimate solution to this problem, from a digital signal processing perspective, is to convolve the dry musical signal recorded in a studio with the room impulse response of the target hall, given the fact that an acoustic space is by and large a linear, time-invariant system. There are several problems and difficulties with this approach. For instance, convolution is a very expensive computation and a measured impulse response corresponds to a single source-listener configuration. On the other hand, although the "artificial reverberators" can possibly run in real-time and are able to simulate arbitrary source-listening configurations, they often fail to create a faithful reproduction of the acoustic space. Our solution is aimed at bridging the gap between these two extremes, by retaining high spatial fidelity while still being flexible enough to simulate arbitrary source-listening configurations. In addition, it's scalability supports real-time implementations.

The first step towards this goal is to acquire sufficiently accurate RIR measurements. Various techniques for measuring RIR have been studied [7] [8] [9]. The three most popular excitation signals for RIR measurement are: a Maximum Length Sequence (MLS), an impulse, and a chirp signal. For analyzing a large concert hall, however, the impulse and the MLS sequence are not good choices for a number of reasons [6]. We therefore choose the chirp signal, which contains all the frequencies required, is a linear signal so is less likely to damage the equipment and also contains a large amount of energy. Using a chirp signal longer than the RIR to be measured allows the exclusion of all harmonic distortion products, practically leaving only background noise as the limitation for the achievable SNR [7].

Our measurement system works as follows. The chirp signal is generated by a laptop computer and played to a speaker. Assuming that most RIR would not exceed 3 seconds, we use a linear chirp signal with a duration of 3 seconds and frequency sweeping from 0 to 24 kHz. At the receiver end, the output signals of a microphone array with 7 microphones are recorded to the same laptop through a multichannel audio interface, together with the unaltered chirp to be used as the reference signal. The unaltered reference signal is important in that it eliminates the need to estimate the latency in the playback-record chain. To obtain the multichannel RIR, the received signals are correlated with the reference signal. Just as in a radar processing application, this function compresses the pulse and gives rise to the room impulse response that is to be analyzed [6]. We use a microphone array that consists of 5 equal-angle spaced directional microphones in the horizontal plane, plus two highly directional microphones aimed vertically up and down, spaced on a sphere of about sphere 30 cm size (0.9 milliseconds delay based on the speed of sound) [10], as shown in Fig. 1. This configuration enables us to reproduce the acoustic space faithfully on the target speaker system, and provides us with sufficient data for a robust analysis of the room responses. Sample output of of RIR measurement system

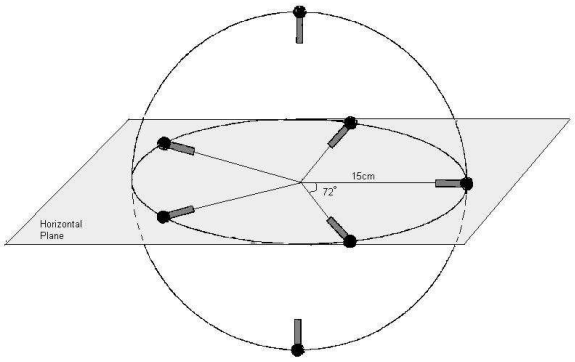

Fig. 1. Microphone Array

is shown in Fig. 2. It is worth mentioning that our method is independent of measuring techniques because what we need are the measurement results. In latter sections, we will discuss that a "good" analysis is the key to the success of our spatial sound rendering solution.

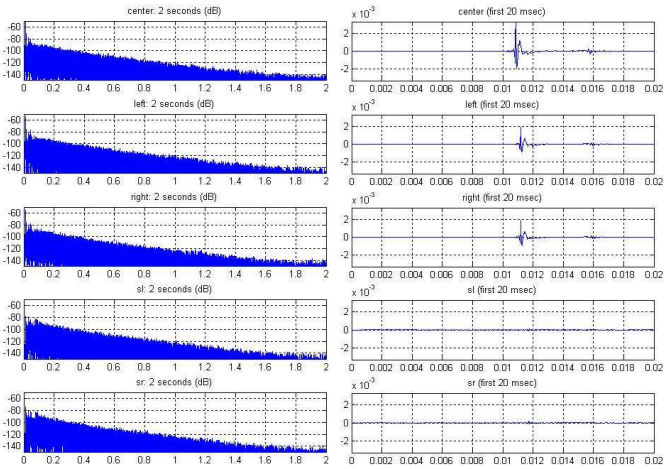

Fig. 2. Typical Concert Hall MMRIR

Analysis of the MMRIR

Various types of analysis can be performed upon the MMRIR to gain insight into the recording venue, for example, [8]. Because the purpose of our analysis is to build a image-source model, we focus on the analysis that leads to accurate estimation of wall and air absorption characteristics. Let us consider the air absorption first. The effect of air absorption is an important factor in image-source models, especially for large acoustic spaces, such as concert halls where higher order reflections can arrive considerably delayed from the direct sound. The air absorption phenomenon is observed as increased low-pass filtering as a function of distance from

the sound source. Based on the standardized equations for calculating air absorption, transfer functions for various temperature, humidity, and distance values were calculated, and second-order IIR filters were fitted to the resulting magnitude responses in [11]. In our analysis, we also use a second-order IIR filter to model the absorption of air and the filter coefficients are determined by fitting the impulse response of this filter to the tail of the direct sound (the first peak of the measured RIR). This operation is based on the assumption that the direct sound is only "filtered" by the air and attenuated by propagation. This is a problem of finding an IIR filter with a prescribed time domain impulse response and can be solved using the Steiglitz-McBride algorithm [12]. The result is shown in Fig. 3. Higher order IIR filters can be used to achieve better approximation if quality is the first priority.

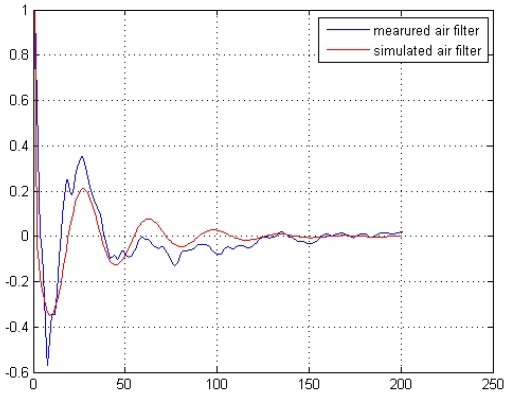

Fig. 3. 2nd-order IIR approximation of air filter

The estimation of wall absorption filter can potentially be much more complicated. The temporal or spectral behavior of reflected sound as a function of incident angle, the scattering and diffraction phenomena, etc., makes it impossible to use numerical models that are accurate in all aspects [11]. Besides, the effect of the wall filter is always coupled with the effect of the air filter in our case. For the purpose of simplicity and to avoid introducing error caused by inaccurate air filter approximation, at this stage we ignore the effect of air filter when modeling the wall absorption. This is reasonable because the air filter is effectively a low pass filter with a unit gain at low frequencies. Based on this assumption, we can then establish the frequency response of the wall filter from frequency-dependent Reverberation Time (RT60) based on the fact that the RT60 is almost solely determined by room dimension and wall material. According the the famous Sabine's formula [13], the reverberation time RT60 of an enclosure with volume V and boundary surface S, which is defined as the time it takes a signal to fall -60 dB, can be calculated by

$$RT60 = 0.163V/Sa \tag{1}$$

where a is the absorption coefficient averaged over the whole boundary. Because a is frequency dependent, RT60 is also frequency dependent. Our method is to estimate frequency dependent RT60 and then derive the wall absorption filter from it. RT60 can be estimated from measured RIR using various techniques, e.g., Schroeder's backward integration [13]. In our analysis, we decompose the RIR into a number of subband components using Short-Time Fourier Transform (STFT) with an FFT size of 2048, which gives us the frequency resolution of 23.44Hz (given the sampling frequency of 96 kHz) in each band. Then in each subband, an individual RT60 is estimated as the time it takes to decay to -60 dB of the direct arrival.

Having obtained the RT60, the next task is to estimate the frequency dependent wall absorption factors. Since the dimension of the room is know at the time of measurement and RT60 indicates the time, and in turn the approximate distance d_{RT60} , that the sound has traveled before it reaches -60dB, we can estimate roughly how many times the sound hits the wall as

$$n = d_{RT60}/dim_{average} \tag{2}$$

The total wall attenuation in each frequency band is

$$w_{total} = -60dB/(1/d_{RT60}) \tag{3}$$

where $1/d_{RT60}$ is the propagation loss by $1/r$-law. Then the wall absorption in that band is $w_{single} = w_{total}^{1/n}$. The frequency-dependend absorption factors composite the frequency response of the wall filter. Similar to the air filter, wall absorption can also be approximated by a second-order IIR filter [11]. Now the problem becomes designing a IIR filter from its frequency response and can be solved using standard filter design techniques. We use the Modified Yule-Walker Method [14] and the frequency response of the designed filter is shown in Fig. 4.

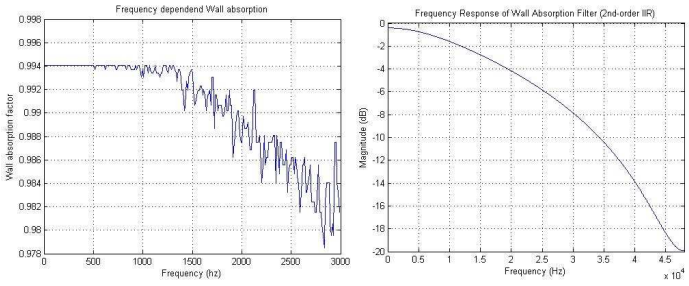

Fig. 4. 2nd-order IIR approximation of wall filter

The analysis can also be performed in a non-uniform frequency band, e.g. an auditory filterbank such as the gammatone filter bank, in order to make the analysis consistent with the human auditory system. Note that because we have multichannel RIR measurements, the air filter and wall filter are estimated using all the channels. The target impulse response of the air filter is taken from the average of the

normalized direct arrival tails. Similarly, the target frequency response of the wall filter is the average over all channels. The filters estimated using above methods may not be very accurate in some cases. In the Section 4, we will describe how this approximation can be refined using our ray tracing model.

III. Spatial Sound Rendering

One of the most important tasks in building our spatial sound rendering system is to select an appropriate room acoustic model. This model needs to be scalable, easily controllable and able to render high quality at a reasonable complexity. Another important task is to customize this model using the analytic results from the previous section. In our system where the real-time requirement imposes a limit on the computation complexity, we use a hybrid method that models only the direct sound and early reflections individually using the image-source method and simulates the late reverberation using a set of filters derived from the MMRIR.

Computational room acoustic modeling has been studied and used for more than three decades and a number of modeling schemes have been proposed. They can be largely catalogued into wave-based methods, ray-based methods and statistical models [15]. Based on geometrical room acoustics, the ray-based methods are the most often used modeling techniques, while the other two types of methods do not fit into a real-time sound rendering system due to a number of reasons [15]. One of the most commonly used ray-based methods is the image-source method. The basic principle of the image-source method is to replace the reflected paths from the real source by direct paths from reflected mirror images of the source. In the image-source method the sound source is reflected at each surface to produce image-sources which represent the corresponding reflection paths. In our system, only a small number of early reflections are calculated with the image-source method due to its accuracy in finding reflection paths. Unlike the image-source models used in other auralization systems which need the user to specify the air and wall characteristics [5], our image-source model is derived from the MMRIR.

For each sound source, these early reflections are modeled as a FIR filter which is called the early reflection filter $h_e(n)$ or $H_e(z)$ in this paper. If the air and wall absorption is ignored, this filter has a series of discrete peaks and each peak corresponds to the signal arrived from an image-source. When the effects of the air and wall are taken into account, each peak becomes a filter itself that is called the image-source filter h_{is} or $H_{is}(z)$. Using the analysis result from the previous section, this filter can be expressed as

$$H_{is}(z) = H_{p,is}(z)H_{a,is}(z)(H_w(z))^{n_{is}} \tag{4}$$

where $H_{a,is}(z)$ and $H_w(z)$ are the air and wall filters obtained from MMRIR analyzes, n_{is} denotes the number of times this image-source hits the wall (the order the reflections), and $H_{p,is}(z)$ denotes the delay and attenuation from propagation. Because the signals arriving at the receiver (microphone array) are the superposition of direct arrival and all the reflected copies, we can express the early reflection filter as

$$H_e(z) = \sum_{is} H_{is}(z) \tag{5}$$

One of the disadvantages of ray-tracing based methods is that the wall is often supposed to be perfectly flat and have constant absorption characteristics everywhere. In order to add the impression of diffused reflection, we impose randomness on the reflection angles by adding a small gaussianly distributed random number to the calculated position of image-sources,

$$\vec{x}' = (1 + \beta_x)^n \vec{x} \tag{6}$$

where $\beta_x \sim n(0, v_x)$ and n is the order of reflection. The roughness of the reflecting surface can be easily controlled by the variance v_x. We do the same to the wall absorption factors based on the assumption that the material on the reflecting surface is uneven to a certain degree. For simplicity, we use a universal random factor for the entire frequency range. The modified image-source filter becomes

$$H'_w(z) = (1 + \beta_w)H_w(z) \tag{7}$$

where $\beta_w \sim n(0, v_w)$. Similarly the unevenness can be controlled by the variance v_w.

However, the above mentioned method alone doesn't fit into a real-time framework because the number of image-sources grows exponentially as a function of order of reflections, and it is computationally inefficient to use the image-source method to find the higher order reflections. In other words, the image-source method is not a good choice for simulating the late reverberation.

The late reverberation in a room is often considered nearly diffuse and the corresponding impulse response exponentially decaying random noise [16]. Under this assumption, the late reverberation does not have to be modeled individually for each source or listener location because it does not contain information for critical direction perception. To optimize computation in late reverberation modeling, a number of artificial reverberation algorithms have been proposed, e.g., [3] [5]. However, these algorithms are derived from (often overly) simplified physical model or perceptual models that are not fully established. Therefore they are often incapable of creating the acoustic impression of a sound space faithfully. Additionally, all these artificial reverberators contain multichannel feedback network so that the stability is not always guaranteed, especially when tuning the parameters. With the RIR measurement at hand, we have the power of re-creating the actual acoustic impression of the recording venue. One straightforward way of generating multichannel late reverberation is to convolve the dry signal with the tails of MMRIR directly. This method has the advantage of preserving

the exact acoustic field at locations where MMRIR is made. Together with the early reflections generated by the model built upon the actual impulse response of the same recording venue, the consistency between the early reflections and the late reverberation, and the consistency between the synthesized impulse responses and the real ones are guaranteed.

We have introduced the methods we used to measure and analyze room impulse responses, as well as the process to build an acoustic model based on the measurement and analysis. In next section, we will describe the development of a system that uses our acoustic model to render multichannel audio in such a way that the acoustic impression of the recording venue is faithfully re-created.

IV. Implementation and Evaluation

Our spatial sound rendering system is implemented as in Fig. 5. It consists of two main components, namely, online processing unit on the left and offline processing unit on the right. There is also a control unit that controls the analysis and rendering process.

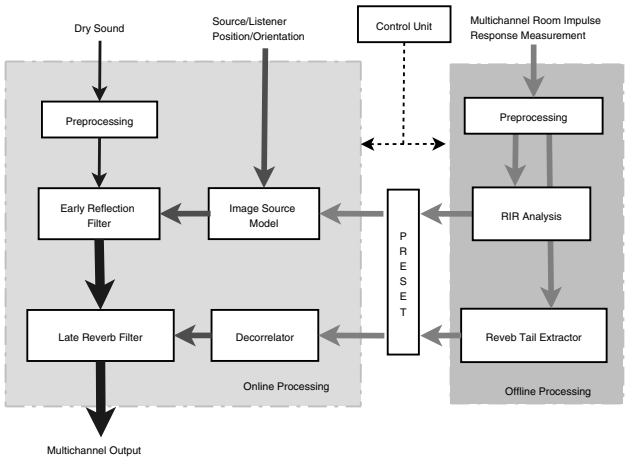

Fig. 5. System Diagram

The offline processing unit handles the tasks that do not need to be done in real-time, including preprocessing and analysis of RIR and extracting the reverb tails. The offline unit needs to run only once when a new set of RIR measurement is fed in or critical control parameters such as the order of the air/wall filters need to be modified. The analysis results from a certain set of RIR, including wall/air filters, reverb tail and optionally room dimension and geometry, are grouped together and called a "PRESET". A preset may contain several "profiles" that are targeted for different complexity and real-time requirements. Creating profiles is fairly straightforward by controlling, for example, the order of the air/wall filter and the length of the reverb tail. The offline unit may also contain an optional preprocessing block that is responsible for outputting "nice and clean" RIR's by, e.g., normalizing, removing the distortion caused by the playback-recording chain and an inappropriate source signal, and/or removing noises. In order to refine the estimated air and wall filters, a high order image-source model can be used to synthesize the MMRIR and the air/wall filter can be adjusted iteratively so that the synthetic MMRIR matches with the measurements statistically or perceptually.

Offline unit:

MMRIR	7-channel, sampled at 96kHz, effective length of 2.5 second, recorded at Rolston Hall of Banff Center
Preprocessing	Normalization
RIR analysis	2nd-order IIR for both air and wall filter
Tail Extractor	retain reverb tail from 0.1 sec to 2 sec, resampled to the working frequency of online unit, 44.1 kHz

Online unit:

Source listener	a number of source positions were test, listener (microphone array) sits on the middle point of the room. Source and listener are assumed omnidirectional. Single static source.
Image-source model	Shoebox geometry was used with the estimated dimension of Rolston Hall (22m x 17m x 6m); order of reflection was 4; variance of air and wall randomization factor were both 0.1.
Dry Sound	clarinet recording at 44.1 kHz
Reflection filter	Fast convolution using FFT
Reverb filter	Fast convolution using FFT

Table 1. System Configuration

The online processing unit is responsible for rendering the "dry signal" to a multichannel speaker system in such a way that the perceived sound source is located at the user determined position in the recording venue. The online unit contains an image-source model for generating the early reflection filter based on the room geometry and the user-defined source location in real-time, a filter chain including the early reflection filters and reverb tail filters, and optionally, a decorrelator and a preprocessor. The decorrelator may become necessary when using the same set of reverb tail to render multiple sources.

After experimenting with various parameters, we selected the options in Table.1 to build a testing system that is capable of offering good quality at a reasonable complexity to run in real-time. This testing system is an example of a "music production" profile that is targeted for studio production where the top priority is the quality.

Informal listening tests have been carried out. The synthesized multichannel sound successfully created the acoustic impression of Rolston Hall at Banff Center, as confirmed by the recording engineers who work there.

V. Conclusions and Future Work

In this paper, we proposed a new spatial sound rendering system derived from multichannel room impulse response measurements. The new system uses a hybrid model that models only the direct sound and early reflections individually using the image-source method and synthesizes the late reverberation using a set of filters derived from the MMRIR. The image-source model is built upon the parameters estimated

from MMRIR. Randomization can be applied to these parameters to simulate diffraction. The multichannel reverberation tails are created by filtering the input signal with (optionally) decorrelated MMRIR tails. Compared with existing solutions, the proposed system is capable of offering the following key features: 1) The model is built upon RIR measurement which is a true reflection of physical acoustics in the measured room; 2) It is can be easily extended to produce a new spatial impression - only multichannel RIR measurements are needed; 3) it is scalable and flexible in that its quality and complexity can be controlled easily; 4) it is able to simulate arbitrary source-listener configurations. Informal listening tests indicate that the proposed system is effective.

There are several areas where our system may be improved. Because the tail of RIR is normally very long, fast convolution using FFT may still exceed the available computational capacity in some cases such as rendering multiple moving sources. More efficient methods, e.g. IIR approximation [17] or the Common-Acoustical-Pole Zero model [18], are being investigated. The current system focuses on the 5-channel multichannel reproduction, but it may be extended to any speaker or headphone configuration, with or without the corresponding MMRIR. In order to gain a better understanding of the limitations and benefits of our proposed solution, formal listening evaluations and tests are being performed.

ACKNOWLEDGMENTS

The authors would like to thank Theresa Leonard of the Banff Centre for access to the recital hall, equipment and staff to do the recordings. We would also like to thank Kirk McNally and Paul Geluso for critical listening with their very experienced ears.

REFERENCES

[1] H. Kuttruff, "Sound field prediction in rooms," in *Proc. 15th Int. Congr. Acoust.*, 1995.

[2] D. A. Burgess, "Techniques for low cost spatial audio," in *ACM Symposium on User Interface Software and Technology*, 1992, pp. 53–59.

[3] J. Jot, "Efficient models for reverberation and distance rendering in computer music and virtual audio reality," in *Proc. 1997 Int. Computer Music Conf.*, 1997.

[4] Aureal Corporation, "3-d audio primer," 1998. [Online]. Available: http://www.headwize.com/tech/aureal1_tech.htm

[5] L. Savioja, J. Huopaniemi, T. Lokki, and R. Vaananen, "Virtual environment simulation - advances in the DIVA project," in *Proc. Int. Conf. Auditory Display*, 1997.

[6] J. Edwards, "Acoustic room response analysis," TechOnLine Publication, 1997.

[7] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. 108th AES Convention*, 2000.

[8] P. Fausti, A. Farina, and R. Pompoli, "Measurements in opera houses: comparison between different techniques and equipment," in *Proc. of ICA98 - Int. Conf. on Acoustics*, 1998.

[9] I. Mateljan, "Signal selection for the room acoustics measurement," in *Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1999.

[10] J. D. Johnston and Y. H. V. Lam, "Perceptual soundfield reconstruction," in *Proc. 109th AES Convention*, 2000.

[11] J. Huopaniemi, L. Savioja, and M. Karjalainen, "Modeling of reflections and air absorption in acoustical spaces a digital filter design approach," in *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics*, 1997.

[12] K. Steiglitz and L. McBride, "A technique for the identification of linear systems,," *IEEE Trans. Automatic Control, Vol. AC-10*, 1965.

[13] H. Kuttruff, *Room Acoustics*. Elsevier Applied Science, London, UK, 1991.

[14] B. Friedlander and B. Porat, "The modified Yule-Walker method of ARMA spectral estimation," *IEEE Trans. on Aerospace Electronic Systems*, vol. AES-20, no. 3, 1984.

[15] L. Savioja, J. Huopaniemi, T. Lokki, and R. Vaananen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc*, vol. 47, no. 9, pp. 675–705, 1999.

[16] M. R. Schroeder, "Natural-sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, 1962.

[17] A. Mouchtaris and C. Kyriakakis, "Time-frequency methods for virtual microphone signal synthesis," in *Proc. 111th AES Convention*, 2001.

[18] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. on Speech and Audio Processing*, vol. 2, no. 2, pp. 320 – 328, 1994.