Adaptive Room Acoustic Response Simulation: a Virtual 3D Application

GIOVANNI COSTANTINI, DANIELE CASALI
Dept. of Electronic Engineering
University of Rome “Tor Vergata”
Via del Politecnico 1, 00173 – Rome
ITALY

Abstract: - In this paper we propose a method to simulate a 3D acoustical environment in which sound sources are positioned in a well defined side. The spatial position that human brain assigns to a sound source is influenced by two main elements: the reverberation, which is related to many factors, including the distance of the source and the type of the environment, and the differences between the sound signals that reach the listener’s ears, related to the sound source angulation with respect to the listener’s head. All this elements have to be simulated in order to give the illusion that the instrument sound comes from a particular position in a particular environment. To obtain this result, the proposed method requires to get the stereo Impulse Response (IR) of the environment for each position that have to be simulated. Then, we approximate each IR by a corresponding adaptive IIR filter, that performs real-time operations.

Key-Words: - Environmental Acoustics, Space Acoustics, Adaptive Filters

1 Introduction

It’s common knowledge that different types of music need different acoustical characteristics related on requirements of the particular kind of music. Every closed space, in fact, induces some reflections of the signal generated inside and, with other phenomena, influences on the auditory feeling of a listener, and gives a particular personality to the sound, like spatial impression, spectral coloration, time length of the signal, clarity, etc.

For example, listening a sacred music choir needs a very reverberant place, that stirs the complex of vocal sounds, and works like a powerful case of resonance; a piano concert, at the contrary, needs a drier ambient, that allows a care distinction of every single note, but adds that right spectral coloration which renders the sound more pleasant. So an attractive problem is rendering a listening place adaptable to different types of music, artificially reproducing acoustical characteristics of a particular ambient, to create the illusion to be in that ambient.

On the other side, the reverberation depends also by the position of each instrument: for example, the ratio between dry and reverberated sounds is related to the distance. But also the early reflections, which are the reverberation components mainly related to the shape of the environment, can be influenced by the position of the source. For that reason, if we want to simulate accurately the reverberation of an environment, we have to consider that reverberation could, in general, be different for each position of sound sources. So we need to get the impulse response of every position that we want to simulate.

In addiction, we want to give a 3D sound effect, which means a stereo output for headphones, to simulate also the exact direction where each sound comes from. To achieve this results, we need a stereo impulse response.

This paper is organized as follows: Sect.2 presents a brief overview on the physical phenomenon of reverberation and describes how the reverberation can be modeled by IIR filters; Sect.3 and Sect.4 illustrate the filtering structure that we employed, and the used learning algorithm; finally, Sect.5 presents an application of the proposed method in which we can virtually reproduce the sound environment of a theatre.

2 Reverberation

The reverberation phenomenon consists in a persistence of sound gradually attenuated during a certain temporal interval after the sound source has stopped. This is due to the multiple wall reflections of acoustic spherical waves in the listening room, which reflects them with an absorption coefficient $\alpha$, dependent on the constitutive materials [1]. Reflected waves can be divided into two class. First, there are the early reflections, i.e. waves reflected only one time before reaching listener; this part is very important to build in the listener brain the idea of the room spatial dimensions. Then, many other wave-rays are reflected more times by the walls before reaching the listener, thus creating a very dense whole of echoes that arrive at the listener in random times and constitute the reverberating tail.

To quantify the reverberation entity in a room, the reverberation time $T_{60}$ is defined, as the time for the impulse response decaying from $-5$ dB to $-65$ dB of its maximum level. An ambient is much reverberant if its $T_{60}$ $> 2$s, while it is very dry if its $T_{60} < 1$s [2].

Schroeder was the pioneer who first attempted to make digital reverberation. The first prototype he tried, called comb filter (or plain reverberator), consisted of a single...
delay line of $m$ samples with a feedback loop containing an attenuation gain $g$. To simulate the frequency selective absorption of air and walls, Schroeder modified the plain reverberator, inserting a low-pass filter in the feedback loop; in order to increase the time echoes density and to decreasing the metallic effect produced by comb filters, he cascaded multiple all-pass filters. Schroeder’s structure is composed of four parallel comb filters, followed by two cascaded all-pass ones [3].

The Schroeder’s frequency response results to have a good quality, but it is rather “anonymous”. An evolution of Schroeder’s model was developed by J. Moorer [4]. He observed that Schroeder’s model can modelize well only the reverberation tail, while a room is characterized by its early reflections, that have to be reproduced by a FIR filter. Thus Moorer added a FIR filter before the Schroeder’s reverberator for the aim to efficaciously reproduce the early reflections. Moorer chose the length of this FIR filter to reproduce the first 60÷80 ms of the real impulse response.

3 Employed Adaptive Filter

The Moorer’s and Schroeder’s models can be used to provide a well-sounding effect, setting by hand the coefficients of the basic blocks (comb and all-pass) that compose them.

Fig. 1. Diagram of reverberator

Our objective is quite different: we don’t want to obtain a general purpose effect, we want to simulate artificially, as faithfully as possible, the acoustic response of a particular room. Starting by a defined filtering structure and a desired IR, we have to find a procedure to right identify the coefficients of the filter, so that it can satisfy the following requests:

• the reverberated signal sounds well, i.e. it come out pleasure for the subjective hearing.
• the dry signal filtered by the identified reverberator is as similar as possible to the signal really produced inside the room whose the IR was measured.
• the structure is purposed for any Real-Time implementation.

The structure chosen for the artificial reverberator is shown in Fig.1. It is a generalization of the filtering structure developed by J. M. Jot in 1992 [5] and was already presented by the authors in [6].

A structure that can efficaciously simulate a desired IR must have this properties:

• Elasticity: the filter must have quite a lot of free parameters and must allow free variations of its characteristics;
• Robustness: the filter must be stable and able to produce every kind of IR, with different $T_{60}$;
• Enough complexity: the structure must generate an high echo density.

A structure that satisfy this properties is an “adaptive filter”.

The novelty that Jot introduced, in comparison to the Moorer’s model, is the presence, in the IIR part, of a feedback coefficients matrix $G$. That is for two main reasons:

• remixing in time the echoes, for the aim to cancel the periodicity effect introduced by the comb filters and also by the FIR filter;
• increasing the number of free parameters for a better approximation of the real IR.

For the $G$ matrix we chose $g_{ii} = 0 (i=1,...,6)$, so every comb is feedback on all the others, except on itself. The $b_i$ and $c$ coefficients are employed for scaling every comb and all-pass contribute, thus increasing the elasticity of the filter.

Thus the employed structure is characterized by the following parameters:

• $g_{ij}$, $g_{i0}$, $D_i$ for the six comb filters (18 coefficients);
• the $G$ matrix (30 coefficients);
• $b_i$ and $c$ (7 coefficients);
• $g_i$ and $R_i$ for the two all-pass filters (4 coefficients);

for a total number of 59 free coefficients. We can indicate the transfer function of the whole structure with $F(z, \hat{w})$, where $\hat{w}$ indicates the free coefficients vector.

4 Learning Algorithm

The learning algorithm must identify the values of the 65 coefficients that make the IR of the filter as similar as possible to the desired IR. Accomplishing this procedure setting the parameters by hand is impossible; we have to employ, on the contrary, an automatic procedure, in which the coefficients of the filter are iteratively corrected by a suitable adaptation algorithm, based on the difference $e(n)$ between the desired response $d(n)$ and the filter output response $\hat{d}(n)$, to make the error $e(n)$ as low as possible.

The optimisation algorithm used to the reverberator filter identification is the Simultaneous Perturbation Stochastic Approximation (SPSA) one; it was developed by J.C.Spall in 1992 [7,8]. Let’s consider the problem of minimizing a
scalar differentiable loss function \( J(\hat{w}) \), where \( \hat{w} \) is the \( p \)-dimensional vector of parameters: the optimisation problem can be translated into finding the minimizing \( \hat{w} \) such that \( \frac{\partial J}{\partial \hat{w}} = 0 \).

The choice of the loss function \( J(\cdot) \) represents the crucial point of the entire identification procedure: the final result depends on the capacity of the loss function to express the filter reverberation quality, also regarding the sentence “it sounds well”. \( J(\cdot) \) must quantify the similarity between the artificial response and the real response, as well as must guarantee a good quality of the artificial reverberation.

The loss function we propose in this work carries out:

1. A comparison between the power \( p(t) \) of simulated IR and real IR, to recreate the envelope and the echoes distribution; we windowed the temporal axis in \( N \) windows, we calculated the power in each window and we minimized the maximum power. It allows us to simulate rather well the temporal behaviour of the real IR, reproducing it with the same \( T_{60} \) and echo distribution.

2. A comparison between the real and the artificial frequency responses, minimizing the mean square error, to obtain a similar frequency coloration between the real and artificial entire frequency response.

3. A further frequency test, minimizing the maximum shifting, to reduce the defects introduced by the employed structure. We windowed the frequency axis in \( M \) windows; then, we minimized the maximum error between the frequency responses in each window.

5 Proposed Application

We present an application of the proposed method. For this aim, we consider a situation in which three instruments play together in a theatre: a violin, a cello, and a piano. We consider the listener is positioned quite near to the performers, so that he can perceive the direction of every instrument; in addiction, we note that the sound of every instrument presents his own reverberation, depending on its position in the theater.

A preliminary operation is to record the impulse response of the environment. This needs to be done only one time for each environment and for each position. Two microphones are positioned in the environment at the position where should be the listener’s ears, or, for a better simulation, a dummy head is used (see Fig.2). An impulse source (for example a blank shot gun) is positioned where should be the first instrument, so the first stereo Impulse Response is recorded. Then we repeat this operation positioning the impulse source where we want to simulate the presence of the other instruments, recording a stereo IR for each position. We will have in this way a total of six IR (three for the left channel and three for the right one).

A second phase consists on the learning process. We use one adaptive filter to simulate each IR. The learning process will return us the set of coefficients for every one of the six adaptive filters. Thus we have two adaptive filters for each instrument, which means three filters for the right channel and three for the left channel.

The last phase is the musical performance. This can be done both in real-time or in delayed-time: in real-time, the instruments can play in an anechoic room, and their sound is captured by microphones, one for each instrument, as shown in Fig.3; in delayed-time, the performance of each instrument can be recorded on a separate audio track. Then, the sound coming from every instrument is filtered.
by the corresponding left and right adaptive filters, and go
to the listener’s headphone.
The listener will have the illusion to be in a real theatre
where the instruments are positioned as shown in Fig.4.

The sound sent to the listener by the headphone presents the
same properties, such as reverberation and interaural time
difference, of sounds coming from well defined positions in
the 3D environment that we want to simulate.

References:
[1] W. C. Sabine, “Reverberation”, Lindsay, R.B., editor,
Acoustics: Historical and Philosophical Development,
Originally published in 1900.
reverberation,” Journal of the Audio Engineering Society,
artificial reverberation,” Proc. IEEE Int. Conf.
Acoustics, Speech and Signal Proc., vol. 2, pp 221-224,
response simulation by an IIR adaptive filter”,
330-332.
a simultaneous perturbation gradient approximation,”
simultaneous perturbation method,” IEEE Transactions

6 Conclusion
In this paper we presented a 3D application of room
acoustic response simulation: a virtual 3D sound space with
piano, violin, and cello was reproduced using two channels
headphone. The proposed method uses a set of adaptive
filters, two for each instrument, to simulate the left and the
right IR of each musical source. The adaptive filters are
based on a flexible structure derived by the Jot’s model of
reverberator and the SPSA algorithm is used for the
parameters identification.

Fig. 4. Environment as perceived by the listener