TECHNIQUES FOR THE EXTRACTION OF THE IMPULSE RESPONSE OF A LINEAR AND TIME-INVARIANT SYSTEM

Lorenzo Picinali[1]

Music, Technology and Innovation Research Centre
De Montfort University
Leicester, UK

ABSTRACT

After a brief overview of the simulation of a linear and time-invariant system through the digital convolution, the paper will start with the description of the various kinds of techniques for the calculation of the impulse response (IR) of the system that has to be simulated.

For each technique, and for each signal used for the extraction, we will analyze the positive and negative aspects, then the problems and the advantages that can help the choice of one signal, instead of another, for the simulation of certain kinds of systems.

Starting with the IR extraction through the reproduction and recording of the *Dirac δ* (the impulse function), we will analyze the advantages of this simple technique, and the disadvantages connected with the impossibility of a correct reproduction of the impulse function.

The second technique discussed in the paper will be the white and pink noise one: we will reflect on the computational advantages of the FFT algorithm and on the phase problems of pseudo-random noise signals.

Then, we will move on to describing the *Minimum Length Sequence* signal (MLS), the shift register and the XOR for its generation, the extraction of the *Dirac δ* through the auto-correlation between the original MLS and the one passed through the system, and the problems of this technique, which are strictly linked to the linearity of the system used to measure the IR.

At the end, we will talk about the sweep signal: a simple sinusoid, modulated in frequency by an exponential function, seems to be the best method for the extraction of IR from various kinds of systems. The simplicity of the inversion of the sweep signal and its independence from the non-linearities of the measuring system, make this technique the most suitable for the IR calculation of various kinds of systems. A brief example of an IR extraction from a dummy head system (Head Related Impulse Response), should then give the idea of how this technique can be used for the simulation of all kinds of systems, from the old style compressors and equalizers, to the best sounding rooms.

Keywords – Convolution, impulse response, sweep, MLS, white noise, pink noise, system simulation.

1. INTRODUCTION

The exponential increase of the calculation power of CPU and DSP for the consumer market, and the consequent decrease of the prices of fast personal computers, have allowed the development of increasingly "computationally heavy" algorithms for the processing of signals. For this reason, the real time digital convolution technique, which just five years ago could be carried out only using dedicated hardware, is becoming really popular as the engine of audio plugins for the simulation of systems, such as environments (convolution reverb plugins), frequency equalizers and dynamic compressors.

In this scenario, the importance of the techniques for the extraction of the impulse responses of these systems, has become primary. In the digital domain, the use of a simple *Dirac δ* is the best way to extract the transfer function of a system. The problems start when we try to pass from the digital to the analogue domain: it is impossible (in reality) to generate a *Dirac δ* from any kind of transducer. For this reason, in previous years various techniques for the extraction of the characteristic of an analogue system have been developed: of course each of these techniques use a particular kind of signal.

In this paper we will have a brief overview of four different techniques: the *Dirac δ*, the white or pink noise (FFT algorithm), the MLS (autocorrelation algorithm) and the seno-logaritmic sweep (convolution algorithm).

2. THE SYSTEMS AND THE TRANSFER FUNCTIONS

Given two families of signals, F1 and F2, a system is an apparatus that can transform each F1 signal into an F2 signal. A system can be seen as a "black box", the behaviour of which is described by the transform law S: F1 → F2.

In environmental acoustics, a system is a room or a hall; in a recording studio, a system is an outboard effect; in an orchestra, a system is a musical instrument…all these can be schematized as follows:
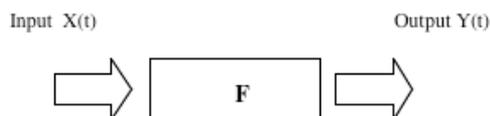
Figure 1. Schematization of a system

The mathematical expression to describe a system is:

$$Y(t) = F[X(t)] \tag{1}$$

We now assume that we have a system that is:

---
[1] Email: lpicinali@dmu.ac.uk

- **Linear:** the "overlap property" needs to be valid; if the input consists of a weighted sum of different signals, the output of the system is an overlap (that is a weighted sum) of the replies of the system to all the single signals in input, as shown in the formulas numbers 2 and 3 (obviously, this is an approximate definition, but it is enough in this case).

$$X(t) + Z(t) \rightarrow Y(t) \qquad (2)$$

$$Y(t) = F[X(t) + Z(t)] = F[X(t)] + F[Z(t)] \qquad (3)$$

- **Time-Invariant:** it needs to be independent from the time; if the input is $X(t)$ and the output is $Y(t)$, for $X(t-t0)$ the system has to give in output $Y(t-t0)$, as shown in the formula number 4.

$$X(t) \rightarrow Y(t) \quad \Rightarrow \quad X(t-t0) \rightarrow Y(t-t0) \qquad (4)$$

A really important notion about the linear spaces it is "basis": a basis is a set of arrays $\{a1,\ldots\ldots an\}$, and each array x can be obtained as a linear combination (weighted sum) of $\sum_i \alpha_i a_i$ elements of the basis, and at the same time no element of the basis can be obtained as a linear combination of the others.
A linear system S can be univocally defined knowing the responses of the system on the elements of a basis. In fact, for each input x, x is obtained as a linear combination $x = \sum_i \alpha_i a_i$, therefore:

$$S(x) = S\left(\sum_i \alpha_i a_i\right) = \sum_i \alpha_i S(a_i) \qquad (5)$$

Knowing $S(a_i)$, for each i, we are able to know $S(x)$ for all the x signals of the space.

2.1. The Impulse Function, or *Dirac δ*

The function $\delta(t)$ can be thought as a rectangle with an "infinitesimal" base Δ and an infinite height $1/\Delta$ (see figure 2), so that:

$$\int_{-\infty}^{+\infty} p_\Delta(x)dx = 1. \qquad (6)$$
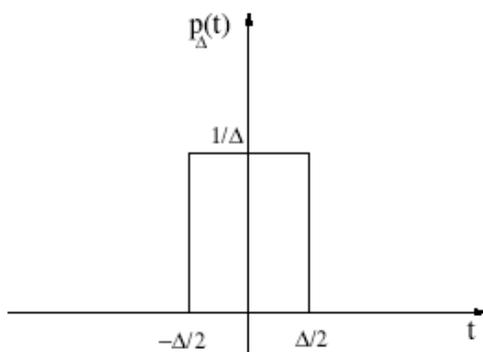
Figure 2. The Impulse Function, or *Dirac δ*

The frequency analysis of the Impulse Function is an horizontal line, parallel to the x axes: therefore, it is possible to deduce that the $\delta(t)$ impulse contains all the frequencies at the same intensity.
It is then possible to say that a linear and time-invariant system can be described by its response to a specific signal: the Dirac Delta $\delta(t)$. In fact, referring to the response of the system at the $\delta(t)$ impulse, it is possible to use as a characterizing element the function $h(t)$:

$$h(t) = F[\delta(t)] \qquad (7)$$

In a linear and time-invariant system, is it possible to describe the output $y(t)$ with the following expression:

$$y(t) = x(t) \otimes h(t) = \int_{-\infty}^{+\infty} h(\tau) \cdot x(t-\tau)d\tau \qquad (8)$$

This formula is the so-called "analogue convolution".

2.2. The digital convolution

When we are talking about audio plugins for the real time convolution (like a convolution reverb), we are not in the analogue domain, but in the digital one, where, fortunately, the formulation of the convolution theory is particularly easy.
In the digital domain, the signals are represented dividing their variability interval into 2^n "sub-intervals" (this operation is called "quantization", where n is the number of bits used for the digital representation). The analogue signal is periodically measured (an operation called "sampling") and, depending on the value of the signal in that time gap, the sample takes on a value expressed in the number of n bits.
The signal gets into the system as an array of numbers, and goes out as another array of numbers, with the same sample rate and the same bitrate. Calling X the array in input and Y the one in output, they can be represented as shown in figure 3.

x_1		y_1
x_2		y_2
x_3		y_3
x_4		y_4
x_5		y_5
x_6		y_6
x_7		y_7
x_8		y_8

Figure 3. The arrays in input and in output from a system

It is important to notice that the numbers in output are directly dependent on the numbers in input: having

inputted a sequence of zeros (silence) followed by non-null numbers and by zeros again, in output we would have a sequence similar to the one in input, but with a different number of zeros before and after the signal. This imparity is due to the fact that the response of the system is not immediate, nor when the system is excited (attack), nor when the system goes back to its beginning state (release). In mathematical terms, it is possible to say that Xn is not just in function of Yn, but of a certain number of samples in input, starting from the n one and going backwards. In the digital domain, this is expressed by the following equation:

$$yn = xn*h1+xn\text{-}1*h2+xn\text{-}2*h3+...+xn\text{-}m*hm \qquad (9)$$

where m is the last sample in the memory. This operation is the digital convolution, and is indicated by the following expression:

$$y = x \otimes h \qquad (10)$$

Therefore, the h coefficients are "characteristic" of the system: looking at them as a waveform, it represents the impulse response of the system.

3. MEASURING THE IMPULSE RESPONSE

After having defined the impulse response of a system, it is essential to establish how to "extract" it: I have outlined below an overview of four techniques for the extraction of the IR from a linear and time-invariant system.

3.1. The deconvolution

Having the system S, knowing the signal x in input and the signal y measured in output, we have to determine h. To do this it is necessary to find a signal x which has an inverse x^{-1} so that:

$$x \otimes x^{-1} = \delta \qquad (11)$$

In this case, we could then obtain:

$$y \otimes x^{-1} = h \otimes x \otimes x^{-1} = h \otimes \delta = h \qquad (12)$$

This means that, knowing x^{-1} (starting from x) and the measured response y, it is possible to obtain h.

It theory, as much as in the digital domain, the extraction of an IR from an analogue system is rather simple: it is enough to put into the system a *Dirac* δ (a signal made by an one followed by a sequence of zeros), and record the output signal. As said before, the *Dirac* δ contains all the frequencies at the same intensity: in this way is it then possible to "test" and have the response of the system for all the frequencies.

On a theoretical level, this operation seems to be simple, however, in practice, there are numerous difficulties: first of all, it is essential to reproduce an impulsive noise sufficiently intense (at least with 60 dB of Signal to Noise Ratio) and short (for example, working at a sample rate of 96 kHz, the impulse should last for 1/96000 seconds).

It could be possible to think about the shot of a gun, but unfortunately the shot doesn't generate a signal with the duration just one sample, but of a few tens of cycles. To get round this problem, it is possible to convolve the signal with itself reversed on the time axes, so that the first sample would be the last one. This technique, known as *Time Reversal Mirror*, helps to arrive quite close to the *Dirac* δ, but it is impossible to reach it, just because it is really complicated to arrive at x^{-1} with enough precision.

Synthesizing a digital impulse, it could be possible to reproduce it through a loudspeaker, but with an elevated intensity for such a short time, the reproduction of the IR signal through a transducer (in this case a woofer and a tweeter) is almost impossible without frequency and phase distortions, and this could create big problems on the calculation of h.

3.2. The pink and the white noise

Looking at the problems showed above, it could be convenient to pass from the time to the frequency domain using the Fourier transform: the convolution between two signals in the time domain becomes a simple multiplication in the frequency one:

$$y = x \otimes h \qquad H = X * H \qquad (13)$$

Each harmonic is multiplied with an m coefficient, for a result of m operations (in the time domain they were m^2), considering then that the transforms and the anti-transforms have a limited computational cost (referring to the *Fast Fourier Transform*, which is much more efficient if compared with the *Discrete Fourier Transform*).

With this technique, the extraction of the m coefficients is quite easy, because they represent the quotients between X and Y: once obtained the coefficients, it is sufficient to carry out an anti-transform to obtain $h(t)$. In this particular case, H is defined as the "transfer function", while h is the "impulse response". Nevertheless, there is an instability problem: having a frequency with a null m coefficient, its relating H coefficient diverges.

To get round this problem, is it possible to refer to the white and the pink noise signals: these are two kinds of signals that have the same energy on all the frequencies. The white noise has a flat spectrum if displayed on a linear frequency scale, while the pink noise has a flat spectrum if displayed on a logarithmic (or exponential) frequency scale. The rich frequency content of these signals, make them really useful for a lot of applications. The main problem is that the samples are generated randomly: therefore, the spectrum is rather discontinuous if visualized with a short windowing, and the phase is not known at all. Increasing the dimension of the FFT windowing, besides the enhancement of the computation time, it introduces problems on the time resolution.

These frequency and phase problems make this technique unsuitable for the majority of cases, where the phase and frequency resolution are essential.

3.3. The MLS signal

A really "clever" signal to use, instead of the white and pink noise, is the MLS (*Maximum Length Sequence*). It is a binary sequence generated by a shift register that follows the following scheme:
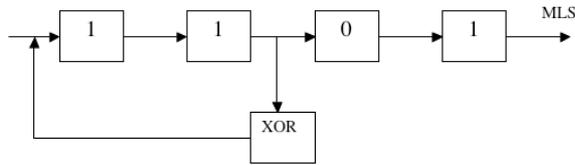
Figure 4. The MLS generation scheme
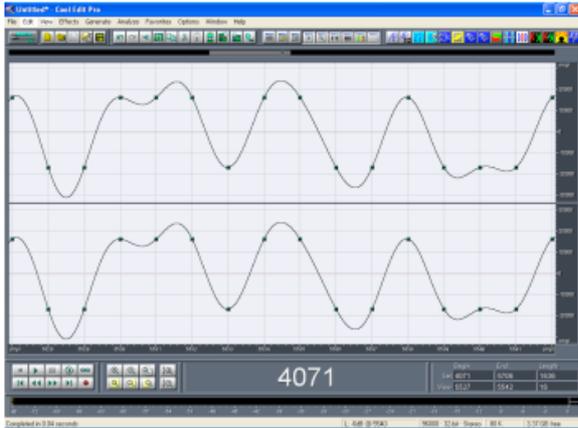
Figure 5. The MLS signal generated with Cool Edit

With a correct positioning of the XOR, is it possible to obtain different kinds of MLS signal. A really important property of this signal is that, by auto-correlating itself, it is possible to obtain a *Dirac δ* without using the FFT algorithm. In fact, it would be enough to generate a MLS x in input, sample the y output signal and cross-correlate x with y; this operation, in the time domain, will generate the impulse response h.

$$\text{If } y = h \otimes x \text{ and } x \subset\subset x = \delta$$
$$\Rightarrow y \subset\subset x = h \otimes x \subset\subset x = h \otimes \delta = h \quad (14)$$

Unfortunately, the principal disadvantage of this technique is the strong dependence on the linearity of the system. The MLS technique requires a perfectly linear and time-invariant system: inexistent echoes and phase problems can appear even with small non-linearities. These problems make this really simple technique unusable for the IR extraction when it is impossible to have a really precise measurement system (in the analogue domain, this happens frequently).

3.4. The sinus-logarithmic sweep signal

Right now, it seems that the most effective and efficient technique for the IR extraction is the one that uses a signal made by a sinusoidal function which goes from the low frequencies to the high ones; a pure tone that increases its frequency with time.

The advantages of this technique is that generating the sweep signal x, its inverse x^{-1} is just the x signal reversed on the time axes. Having x^{-1} and measuring y, it is possible to calculate the IR h with a deconvolution operation:

$$x \otimes x^{-1} = h \otimes x \otimes x^{-1} = h \otimes \delta = h \quad (15)$$

The only problem of this calculation is that the convolution operation is not streamlined, and the computational efficiency of the sweep technique is lower, if compared to the ones described above.
The sweep signal can be linear or logarithmic, depending on the frequency-increasing curve. The most used is the logarithmic one, because with it, it is possible to give more energy on the lower frequencies (critical zone), and to go faster on the higher ones.
The sinus-logarithmic sweep formula is the following:

$$x(t) = \text{sen}\left[\frac{2\pi f_{\text{inf}} \cdot T}{\ln\left(\frac{2\pi f_{\text{sup}}}{2\pi f_{\text{inf}}}\right)} \cdot \left(e^{\frac{t}{T} \cdot \ln\left(\frac{2\pi f_{\text{sup}}}{2\pi f_{\text{inf}}}\right)} - 1 \right) \right]$$

$$(16)$$

where f_{inf} is the starting frequency, f_{sup} is the arrival frequency and T is the time duration.
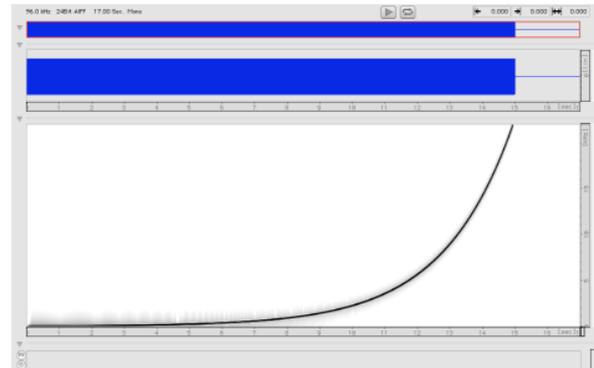
Figure 6. The sonogram of a sinus-logarithmic sweep signal

4. A BRIEF EXAMPLE: THE HRTF CALCULATION

As an example, I will shortly describe how these issues are embedded in my research work on binaural spatialization: the main objective of the research project is to realize an algorithm for the binaural spatialization of audio signals based on a convolution engine.

4.1. The dummy head, the HRTF and the HRIR

Considering a "dummy head" (an artificial head mannequin with two microphones placed at the beginning of the ear canals) as a system to simulate, it is possible to extract the HRTF (*Head Related Transfer Function*) calculating all the HRIR (*Head Related Impulse Response*). In this case it would not be enough to calculate only one HRIR, just because the response of

the dummy head system is different depending on the direction of the sound source. In fact, the DDF, *Direction Depending Filtering*, is one of the localization cues used by our hearing system to establish the position of the sound source in a three dimensional soundscape.

4.2. The HRIR extraction

To extract the impulse responses of the dummy head, it is necessary to sample the azimuth and the elevation parameters in spheres around the head, at different distances (different diameters of the spheres). In all the sampled positions, we have then to reproduce the signal chosen for the IR extraction, and record the output of the system through the microphones placed on the dummy head.

Using the sweep signal (which seems to be the most suitable for this kind of system), we just have to reproduce the signal in all the sampled position, then to convolve the inverse of the input signal with the output one (the one recorded through the dummy head microphones), and we will have in output the HRIR of that specific position. The database of all the HRIR at all the sampled positions will constitute the HRTF.

4.3. The convolution with the HRIR

At this point, the algorithm will work quite easily: giving in input the signal that has to be spatialized and its virtual position in a three dimensional space (azimuth, elevation and distance), the algorithm has just to extract the suitable IR from the HRIR database, and to convolve the input signal with it, giving in output a stereo signal with the binaurally spatialized sound source.

5. CONCLUSIONS

After this short overview, is it possible to say that each of these techniques have their advantages and disadvantages. For this reason, each kind of signal (*Dirac δ*, white and pink noise, MLS and sweep) and each kind of algorithm (deconvolution, FFT, autocorrelation and convolution) are suitable for different kinds of systems and purposes.

For example, to extract acoustical parameters from a theatre hall, the reproduction of an impulse from a gun shot would be the most suitable technique, while for the digital simulation of the same environment, the sweep technique would be much more precise.

At the end, it is possible to say that the most important thing is the careful choice of the technique that seems most suitable for the specific system that has to be simulated. It is important to take into consideration all the aspects, from the computational efficiency of the algorithm, to the dependence on the non-linearities of the measuring system.

6. REFERENCES

[1] V. Ralph Algazi, Richard O. Duda, Dennis M. Thompson, C. Avendano, *"The CIPIC HRTF data base"*, paper from the IEEE Workshop on Application of Signal Processing to Audio and Acoustics, New Paltz, New York, 21-24 Oct, 2001

[2] A. Bertoni, G. Grossi, *"Dispensa del corso di Elaborazione Numerica dei Segnali"*, Università degli Studi di Milano, 2001

[3] A. Capra, *"Caratterizzazione quantitativa e valutazione attraverso test soggettivi di teatri e sale da concerto"*, tesi di laurea Ingegneria Elettronica, rel. Prof. A. Farina, corr. Ing. F. Bozzoli, Università degli Studi di Parma, anno accademico 2002- 2003

[4] S. Muller, P. Massarani, *"Transfer Function Measurement with Sweeps"*, Journal of the Audio Engineering Society, Vol. 49, No. 6, June, 2001

[5] Douglas D. Rife, J. Vanderkooy, *"Transfer-Function Measurement with Maximum Length Sequences"*, Journal of the Audio Engineering Society, Vol. 37, No. 6, June,1989

[6] L. Tronchin, A. Farina, M. Pontillo, V. Tarabusi, *"The Calculation of the Impulse Response in the Binaural Technique"*, 7th ICSV7, Garmish-Partenkirchen Germany, July 4-7, 2000