A Numerical Investigation of the Representation of Room Transfer Functions for Artificial Reverberation.

Davide Rocchesso
Centro di Sonologia Computazionale
Dipartimento di Elettronica e Informatica
Università degli Studi di Padova
Email: roc@csc.unipd.it

William Putnam
Center for Computer Research
in Music and Acoustics
Stanford University
Email: putnam@ccrma.stanford.edu

Abstract

In this paper we try to establish the system order for the correct representation of a set of room transfer functions (RTF), in order to partition memory between the common recursive part and the non-recursive part specialized for each RTF. To this end, we apply a few system theory concepts on a set of simulated rectangular rooms, whose impulse responses were generated using the image method [1]. A further validation of our results is provided by an analysis of the frequency density of a comb-filter modeling of the recursive part.

1 Introduction

Several models have been proposed in the literature for the representation of room transfer functions. FIR filters have been used for echo cancellation, especially using adaptive filtering techniques [2].

Recursive filters are often used for artificial reverberation. These usually take the form of comb and allpass filters, and are often cascaded with tapped delay lines [9, 6, 4, 8].

More recently, it has been proposed to model the RTFs of a given room by a set of common poles, provided by an IIR filter, and an FIR part which varies according to source and receiver positions within the room [3]. A similar approach is taken in [8] for building artificial reverberators, where the recursive part is implemented by means of a feedback delay network (FDN), and the FIR part is a tapped delay line. In this latter model, the FIR part can be associated with the early reflections of the room, while the IIR part can be interpreted as a representation of normal modes and diffusion. This physical correspondence allows one to control the filter parameters in a natural way, and physical consistency is automatically achieved.

An open question that we address in this paper is that of the memory requirements necessary for the correct representation of a set of RTFs of a room. Furthermore, we address the issue of partitioning this memory between the common recursive part and the non-recursive part, which is specialized for each RTF.

In our investigation, we apply system theoretic concepts on a set of simulations we have conducted using the image method [1]. Since the image method can be shown to converge to a correct representation of normal modes [1], we believe that it constitutes a useful tool for studying the room response, even in the steady state. This is an approach which guarantees a large and controllable dynamic range. This allows for a straightforward and repeatable construction of reference rooms, as opposed to actual measurements in actual rooms. In another project, we used the same set of simulations for determining numerical conditions for invertibility of RTFs [7].

2 Methodology

In this section we briefly introduce the system theory background which is needed for our investigation. A thorough treatment of this material can be found in [5].

The Hankel matrix of a single-input-single-output system can be constructed from m samples of a simulated impulse response $h(\cdot)$:

$$\mathbf{H}_h = \begin{bmatrix} h(0) & h(1) & \dots & h(\frac{m}{2}-1) \\ h(1) & h(2) & \dots & h(\frac{m}{2}) \\ \vdots & \vdots & \dots & \vdots \\ h(\frac{m}{2}-1) & h(\frac{m}{2}) & \dots & h(m-1) \end{bmatrix} \quad (1)$$

Given a linear system in its matrix representation

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad (2)$$

$\mathbf{H}_h$ is the product of the Observability and Reachability matrices [5]. Thus, a singular value decomposition (SVD) of H gives information about the dimensionality of a minimal realization of $\mathbf{\Sigma}$, which is equivalent to its reachable and observable part, i.e., it gives the memory requirement of the system. In particular, the matrix $\mathbf{S}$ of the singular values has a form

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \dots & 0 & \mathbf{0} \\ 0 & s_2 & \ddots & 0 & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & \vdots & \dots & s_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \vdots & \mathbf{0} \end{bmatrix} \tag{3}$$

where n is the order of the reachable and observable part, i.e. of the minimal realization.

In practice, using measured or simulated impulse responses, the singular values s_i will never go exactly to zero, corresponding to the fact that an infinite state space would be necessary to represent three-dimensional wave propagation in discrete time. Nevertheless, we can plot the singular values on a log scale (see fig. 1) and consider n to be the abscissa corresponding to a reduction of, say, $60dB$ in the singular values.

3 The Experiment

In figure 1 we have depicted the singular values of a simulated impulse response for a room with edges $8m \times 10m \times 6m$, with a sample rate of $1KHz$. A line has been drawn to interpolate the data from the sample 20 to 480. We cut the spike in very-low frequency, since we believe it is due to the fact that the simulated impulse response has a strong DC component. We see that the singular values go down with a slope of about $0.26dB/sample$, thus indicating that about 231 memory cells are needed to represent the RTF with an "accuracy of $60dB$", and 369 memory cells are needed for $96dB$ of accuracy.

It has to be noticed that, throughout our discussion, the accuracy of representation is measured directly in the space of singular values of the impulse response to be modeled, rather than actually modeling the room and measuring a signal-to-noise ratio. For instance, on the basis of the example of fig. 1, we can say that, in principle, there exists a discrete-time linear system having 369 state variables, which can reproduce the impulse response $h(\cdot)$ within the 16 bits of accuracy.

Now, let's consider more than one impulse response simultaneously, namely a system with r inputs and one output. The product of the controllability and reachability matrices now takes the

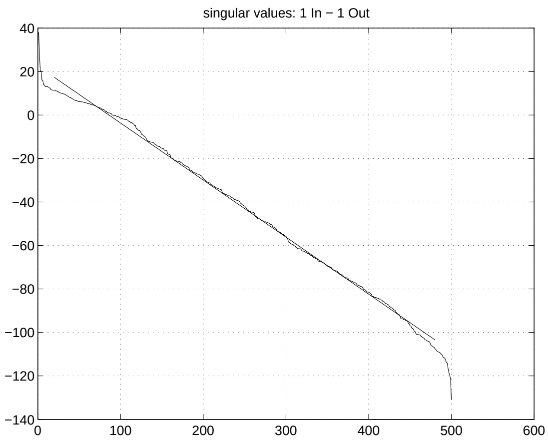

Figure 1: Singular values for 1 input and 1 output

form of a block Hankel matrix:

$$\mathbf{H}_{bh} = \begin{bmatrix} \mathbf{h}(0) & \mathbf{h}(1) & \dots & \mathbf{h}(\frac{m}{2}-1) \\ \mathbf{h}(1) & \mathbf{h}(2) & \dots & \mathbf{h}(\frac{m}{2}) \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{h}(\frac{m}{2}-1) & \mathbf{h}(\frac{m}{2}) & \dots & \mathbf{h}(m-1) \end{bmatrix} \tag{4}$$

where any $\mathbf{h}(\cdot)$ is a $1 \times r$ vector. We can still perform a SVD on this matrix and compute the memory requirements for a correct simultaneous representation of the different RTFs.

When a second input is added to the previous one-input-one-output impulse response, we notice that the slope of the singular values reduces to about $0.2dB/sample$ (fig. 1), thus indicating that about 300 memory cells are needed to represent the RTFs with an "accuracy of $60dB$", and about 480 memory cells are needed for $96dB$ of accuracy.

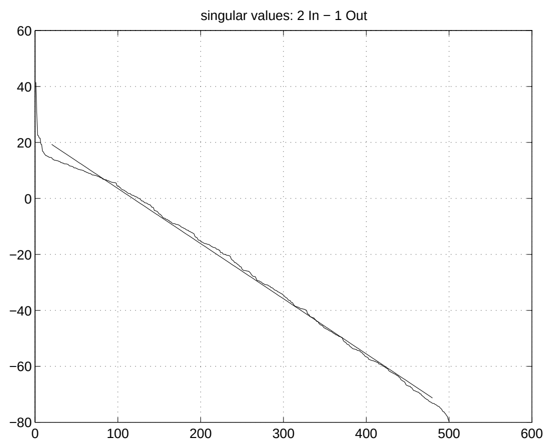

Figure 2: Singular values for 2 inputs and 1 output

Given an accuracy of $60dB$ we have, for one input and one output

$$M_1 = 231 \tag{5}$$

and, for two inputs and one output

$$M_2 = 300 \tag{6}$$

Let us assume that there is a common recursive part M_r. Let's also assume that the non-recursive parts are totally independent for different inputs or outputs. This means that, if in the one-input-one-output case we have a non-recursive memory requirement M_f, in the two-input-one-output case the non-recursive memory requirement has to be $2M_f$.

We come up with the equations

$$\begin{array}{l} M_1 = M_f + M_r = 231 \\ M_2 = 2M_f + M_r = 300 \end{array} \tag{7}$$

The system (7) can be solved and it provides

$$\begin{array}{l} M_f = 69 \\ M_r = 162 \end{array} \tag{8}$$

In other words, in a single RTF representation, about 70% of the memory should be put on the recursive part.

This simple memory allocation has been done on the basis of a somewhat arbitrary assumption. The same assumption has been taken in the past with no formal justification. We now try to justify the presence of a common IIR part and of independent FIR sections by considering a third RTF. When taking an additional input and computing the SVD on the respective Hankel matrix, we get a slope of $0.17 dB/sample$ in the singular values (fig. 3), thus indicating that about 352 memory cells are needed to represent the RTFs with an "accuracy of $60dB$", and about 564 memory cells are needed for $96dB$ of accuracy.

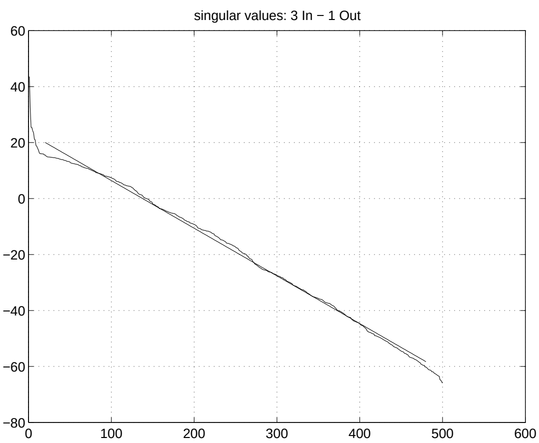

Figure 3: Singular values for 3 inputs and 1 output

If our model is correct, the system

$$\begin{array}{l} M_1 = M_f + M_r = 231 \\ M_2 = 3M_f + M_r = 352 \end{array} \tag{9}$$

shouldn't give a solution very different from (8). In fact, we get

$$\begin{array}{l} M_f = 61 \\ M_r = 170 \end{array} \tag{10}$$

so that about 73% of the memory should be put on the recursive part. This is different from the previous 70%, but the difference is compatible with the tolerance of the line fitting and with the somehow arbitrary choice of the lower and higher cutoff frequencies for the linear interpolation.

Our calculations are very sensitive to the actual placement of sources and receiver. In particular, symmetric positions should be avoided, since they break the assumption of independent non-recursive transfer functions.

4 Physical considerations

The experimental evidence fits well with the memory splitting as determined by strictly physical considerations of the model [8].

In the cited model, the digital representation of the recursive part of a RTF is essentially a superimposition of recursive comb filters. If g is the attenuation coefficient in the feedback loop of a comb filter, then the approximate bandwidth of one of the resonances is

$$\Delta f = \frac{\left(1 - g^{1/n}\right) F_s}{2\pi} \tag{11}$$

where F_s is the sampling rate. The coefficient g is related with the reflectivity of the walls. Each comb filter in the model is associated with a harmonic series of normal modes, and with a direction in space. Progressively shorter delay lines are used for representing higher order modes so that, asuming that the reflectivity is kept constant, the bandwidth (11) also increases.

The richness of the model increases as long as an increasing number of directions in space is considered. It is easy to compute the number of memory cells needed for the case at hand, for an increasing number of space directions (fig. 4). It makes sense to stop the increase in the number of directions when the bandwith of a resonance exceeds the mean distance among two adjacent resonances. It is well known that in an actual three-dimensional enclosure, this distance decreases quadratically with frequency in a rectangular environment, hence at some point it will be less than Δf. From a simulation perspective, it makes sense to try to reproduce the resonances which are clearly separable, thus limiting the number of comb filters to be used. As stated in [8], this corresponds to finding a limit to the number of directions in space where we are considering plane-wave propagation.

If we approximate our room by sampling 8 directions in space, we find that the mean number of resonances per Hz (i.e. the frequency density) is about

$$f_d = 0.27 \tag{12}$$

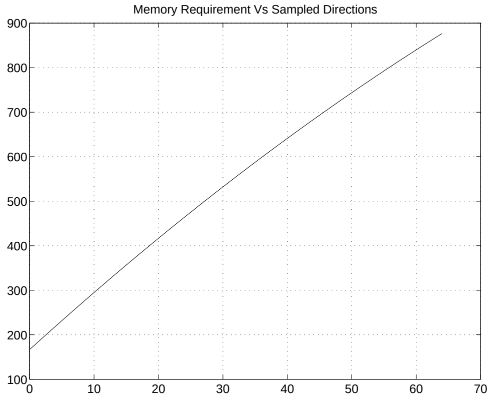

Figure 4: Memory requirement for an increasing number of sampled directions

The resonance bandwidth in low frequency ranges from $1.8Hz$ to $4.3Hz$. These values are very close to the reciprocal of the frequency density (i.e. the mean peak-to-peak distance) which is $3.7Hz$.

The conclusion we can draw from these speculations is that 8 is a good choice for the number of sample directions, and that it is not worth going much higher, since the resonance peaks would not be resolvable. Now, from fig. 4 we can deduce the memory occupation of these 8 comb filters, which is 270 memory cells. Somewhat surprisingly, this number is close to the memory requirement for the recursive part, as computed by means of the singular value decomposition of the Hankel matrix.

5 Conclusion

We have outlined two criteria for estimating the memory requirements for a discrete-time modeling of room transfer functions. The first method is based on the singular value decomposition of Hankel and block Hankel matrices, while the second method relies on the discrimination of resonance peaks in the frequency domain. Even though these two criteria do not show an evident relationship, we found that they provide similar estimates in a practical case. This result is a clear evidence of the fact that any RTF can be split in an FIR part and an IIR part, and that this latter part is independent on the source or receiver position, being given by the common resonances of the room. While this split of RTFs was assumed in previous works, this investigation shows that it is justified by numerical results and it indicates a couple of ways of setting the system dimensions.

References

[1] J.B. Allen and D.A. Berkley, Image Method for Efficiently Simulating Small-Room Acoustics, J. Acoust. Soc. Am., vol. 65, 1979.

[2] S. Haykin, Adaptive Filter Theory, Prentice Hall, 1991.

[3] Y. Haneda, S. Makino, and Y. Kaneda, Common Acoustical Pole and Zero Modeling of Room Transfer Functions, IEEE Trans. on Speech and Audio Proc., vol. 2, n. 2, apr. 1994.

[4] J.M. Jot, An Analysis/Synthesis Approach to Real-Time Artificial Reverberation, proc. Int. Conf. Ac. Speech and Sig. Proc., San Francisco, 1992.

[5] T. Kailath, Linear Systems, Prentice Hall, 1980.

[6] J.A. Moorer, About this Reverberation Business, Comp. Music J., vol. 3, n. 2, 1979.

[7] W. Putnam, D. Rocchesso, and J.O. Smith, A numerical Investigation of the Invertibility of Room Transfer Functions, proc. IEEE ASSP workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk, NY 1995.

[8] D. Rocchesso, The Ball within the Box: a Sound Processing Metaphor, Comp. Music Journal, vol. 19, n. 4, 1995

[9] M.R. Schroeder, Natural Sounding Artificial Reverberation, J. Audio Eng. Soc., vol. 10, n. 3, 1962.