ACOUSTIC TEST OF A BASIC AUDIO DEREVERBERATION PROCESS

Davide Bonsi

The Musical and Architectural Acoustics Lab.
Fondazione Scuola di San Giorgio-CNR
Isola di S. Giorgio, 30124 Venezia, Italy
davide.bonsi@cini.ve.cnr.it

Domenico Stanzial[1]

Institute for Earth-Moving Machinery
and Off-Road Vehicles-CNR
Via Canal Bianco 28, 44044 Ferrara, Italy
stanzial@cemoter.bo.cnr.it

ABSTRACT

Removing reverberation effects from music and speech produced in a sound reflecting environment is a useful process which allows to recover the audio signals due to the acoustic source only. In particular, this may be essential for performing the subsequent acoustical optimization during a restoration procedure.

The process can be considered also from the point of view of the control system theory as a special case of active control of the reverberation time in real rooms as suggested by one of the authors [1].

The present paper discusses the implementation of a simple sound dereverberation method by means of a basic algorithm applied to minimum-phase excitation signals. The procedure will be used for performing some sample deconvolutions using impulse responses measured in the same boundary conditions as those actually holding during the sound recording: through this method we will test the acoustic effectiveness of the algorithm.

1. INTRODUCTION

The basic dereverberation techniques developed for linear time-invariant systems rely on the *convolution* relationship

$$y(t) = x(t) * h(t) \tag{1}$$

where $x(t)$ is the excitation signal, $h(t)$ the impulse response and $y(t)$ the response due to $x(t)$. Whenever h can be measured or taken as a known quantity, Eq. 1 makes it possible to transform a given *anechoic* signal x into a signal y having the reverberation pattern typical of the environment under consideration. This usually happens in room acoustics, where Eq. 1 is the basic equation underlying the *auralization* techniques. The inverse problem, that of *de-reverberation*, can be accomplished finding a function $g(t)$ such that $h(t) * g(t) = \delta(t)$. If inserted into Eq. 1, g gives the *deconvolution* relationship $x(t) = y(t) * g(t)$, so that in principle, from a recorded signal y it is possible to get back to its anechoic source by determining the inverse impulse response and executing a convolution between this function and the waveform y itself. However, some important constraints have to be taken into account for putting the procedure into practice. Some are of acoustic nature: firstly, the signal source should be point-like compared to the typical wavelength of the recorded sound, otherwise the signal y would not be expressed by a simple time-convolution, for a spatial integral on the effective extension of the source should be performed as well; moreover, the impulse response of an enclosure is a temporal function depending on both the positions of the source and the receiver [2]. This means that for a correct recovery of x these positions must be the same when measuring y and h.

Other constraints are due to the signal processing rules and regard mainly the restrictions of stability and causality imposed on the inverse function. In fact, as shown by the wave theory of room acoustics, the impulse response may be expressed as a sum of complex exponential functions of time; in the discrete domain, this is written as a finite sequence whose z-transform gives the following "all-zero" transfer function

$$H(z) = \prod_i (1 - a_i z^{-1}) \prod_i (1 - b_i z) \tag{2}$$

where $|a_i|$, $|b_i|$ are less than unity. It is now easy to see that the poles of $G = 1/H$ correspond to the zeros of H. As a consequence, G is stable if and only if all the coefficients b are zero, meaning that all the zeros of H are located inside the unit circle in the complex z-plane, that is H is a minimum-phase function.

Unfortunately the minimum-phase condition is seldom fulfilled by ordinary impulse responses, thus in general it is necessary to devise special algorithms for finding approximate transfer functions. Nonetheless, in the present work we will test experimentally the deconvolution procedure in a case where the general algorithm may be implemented for arbitrary impulse responses thanks to the carefully chosen characteristics of the excitation [3]. To see how this may be achieved, we remind that any transfer function may be written as a product of a minimum-phase term H_m and an "all-pass" H_{ap} term, such that $|H_m(z)| = |H(z)|$ and $|H_{ap}(z)| = 1$. If the excitation is a minimum-phase waveform $X_m(z)$, the convolution relationship becomes

$$Y(z) = X_m(z)H_m(z)H_{ap}(z) = Y_m(z)H_{ap}(z) \tag{3}$$

where the right-hand term holds since a product of two minimum-phase signals gives a minimum-phase signal. Therefore, for the recovery of X_m, it suffices to determine the minimum-phase component of the output signal and subsequently operate on it with the inverse of the minimum-phase component of the transfer function. The most important step in the procedure is the determination of the minimum-phase component of the impulse response and the signal y, which is accomplished by the "homomorphic filtering" technique. The algorithm follows from the property according to which the logarithmic magnitude and phase of the Fourier transform of a minimum-phase signal satisfy a Hilbert transform relation. This allows to obtain for any mixed-phase sequence s the complex cepstrum $\widehat{s}_m$ of the minimum-phase sequence s_m having

[1] Also with Fondazione Scuola di San Giorgio c/o Fondazione G. Cini, Venezia, Italy.

the same Fourier transform magnitude as s, the basic relationship being

$$\widehat{s}_m(n) = c_s(n)w(n) \tag{4}$$

where c_s denotes the real cepstrum of s and w the "frequency-invariant" linear filter: $w(n) = 2u(n) - \delta(n)$ [4]. The sequence s_m may then be recovered from the inverse complex cepstrum of $\widehat{s}_m$.

2. EXPERIMENTAL METHOD

The setup was developed so as to meet the above-mentioned physical restrictions which should be fulfilled for treating signals by means of simple convolutions: an approximation valid when the source can be considered as point-like and omnidirectional. The source and the receiver were respectively a B&K loudspeaker mod. 4295 (*Omnisource*), which is characterized by an almost isotropic response and a small radiating surface, and an electrostatic microphone. They were placed $3\,m$ far from each other in a small empty room of 3.3 $\times$ 5 $\times$ 3.4 m^3, having a reverberation time (T_{60}) of about 3 s. We simultaneously generated and recorded some minimum-phase sequences and then, keeping the system in the same geometrical configuration (that is with the positions of the source and the receiver unchanged), we measured the corresponding room impulse response which served us for executing the deconvolution. The unit of input/output was a PC equipped with a full-duplex PCMCIA sound board (Digigram VXpocket©). As regards the signals, the impulse response measurements were performed with the aid a software employing the MLS technique (WinMLS©), while for the playback and recording of the test sequences the SpectraLAB© software was used. The processing, from the minimum-phase decomposition to the implementation of the deconvolution algorithm (the latter executed in the frequency domain via FFT), was accomplished by MATLAB© routines.

For building the excitation signals we used weighted sine waves

$$x(n) = A^n \sin(n\phi) \qquad -1 < A < 1 \tag{5}$$

($\phi = 2\pi f/f_s$) whose minimum-phase property is clearly highlighted by the unilateral z-transform

$$X(z) = \frac{(A\sin\phi)\, z^{-1}}{1 - (2A\cos\phi)\, z^{-1} + A^2 z^{-2}} \tag{6}$$

In order to avoid the non-linear effect which an abrupt start of the excitation could cause to the response of the loudspeaker we used a smoothed version of the sequences of Eq. 5: the start of signal was put equal to $x(n) = (n/n_0)A^n \sin(n\phi)$ in the range corresponding to n less than a given n_0. Though not directly evident from the calculation of the z-transform, it can be shown by the algorithm of Eq. 4 that the smoothed $x(n)$ is still a minimum-phase sequence and therefore can be employed in our experiment. In practice, the frequencies used were around 250 Hz with an attenuation factor $A = 0.9998$; the chosen smoothing interval was $n_0 = 400$ corresponding, for the sampling frequency used ($f_s = 16\,kHz$), to a time interval $n_0/f_s = 0.025\,s$, which proved to be sufficient for preventing the distortion.

3. RESULTS

The typical pattern of the excitation signal x is shown in Fig. 1, corresponding to $f = 246\ Hz$. Figg. 2 and 3 respectively show

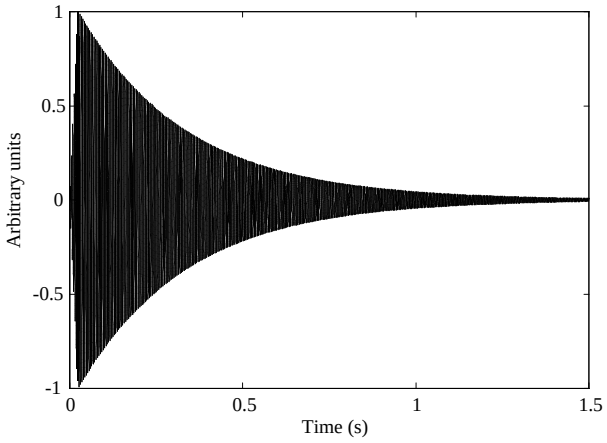

Figure 1: *Typical pattern of the excitation signal.*

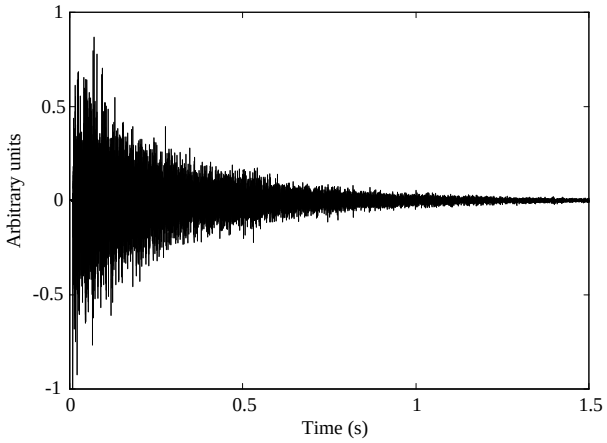

Figure 2: *The room impulse response.*

the impulse response h and the response y due to the excitation, while the result of the deconvolution, which will be called x' in the following, is reported in Fig. 4. A clear resemblance between x' and x can be grasped from a first glance. The first property to be investigated is the relative time-scaling of the two signals, which can be evaluated through the comparison of their envelope functions. We remind that, given a signal $s(t)$, this quantity is defined as the modulus of the corresponding analytic signal $z(t) = s(t) + j\,\widetilde{s}(t)$, $\widetilde{s}(t)$ being the Hilbert transform of $s(t)$. Figg. 5 and 6, reporting the two envelopes in the ranges: $0 \div 0.1\,s$ and $0.1 \div 0.5\,s$, show that the deconvolved signal matches the excitation apart from local fluctuations, which seem to be related to spurious low frequency components. In fact, the effect is far less evident if the deconvolution is performed on a response y filtered below 100 Hz, as shown by the result reported by Fig. 7 and the initial envelope in Fig. 8. An intuitive explanation of this fact might be given by considering the different processing of the two quantities y and h: the first one is obtained by simply averaging the response due to the excitation x reproduced several times in the environment (about 12), while the impulse response is the result of a cross-correlation between

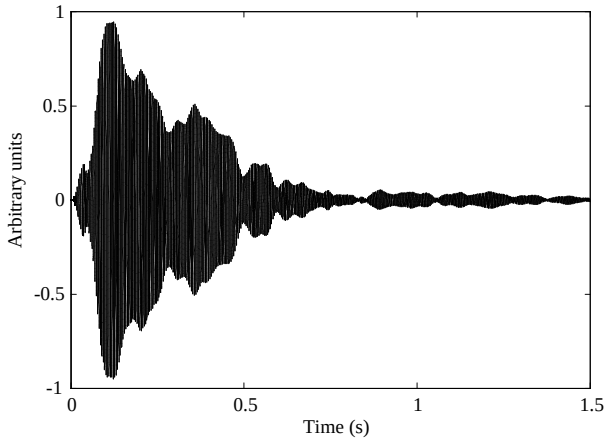

Figure 3: *Response to the excitation signal shown in Fig. 1.*

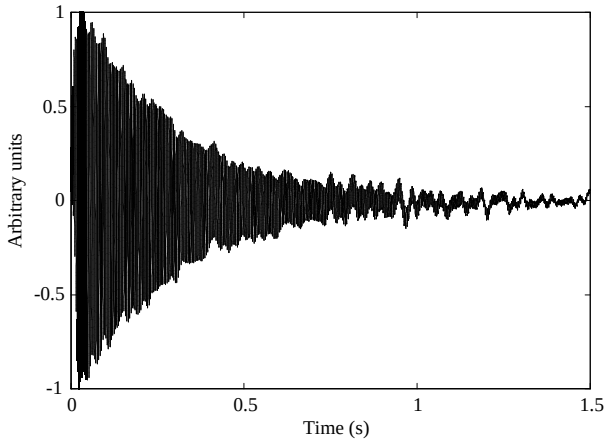

Figure 4: *Deconvolution between the measured signal (Fig. 3) and the impulse response.*

an excitation (a maximum-length sequence) and the corresponding averaged response. In the second case the noise suppression is more effective, especially at low frequencies, so that these components become relatively high in the inverse frequency response obtained from h, and as a consequence they tend to be amplified in the deconvolved signal. A confirmation of this phenomenon may be obtained by looking at the degree of correlation of the recovered time history with respect to the excitation, as expressed by the normalized cross-correlation function

$$R_{xx'}(\tau) = \frac{\langle x(t)x'(t+\tau)\rangle}{\sqrt{\langle x^2(t)\rangle \langle x'^2(t)\rangle}} \qquad (7)$$

The function has been calculated for x' obtained from both unfiltered and filtered y. Its pattern, exhibited by the envelope functions reported in Fig. 9 ($-0.2\,s < \tau < 0.2\,s$), shows that the cross-correlation for unfiltered y is almost perfectly symmetrical around its maximum but systematically lower than the other one. As an absolute reference, the $x' \div x$ similarity may be evaluated from the indicator $\xi = \max(R_{xx'})$, located approximately at $\tau = 0$: the

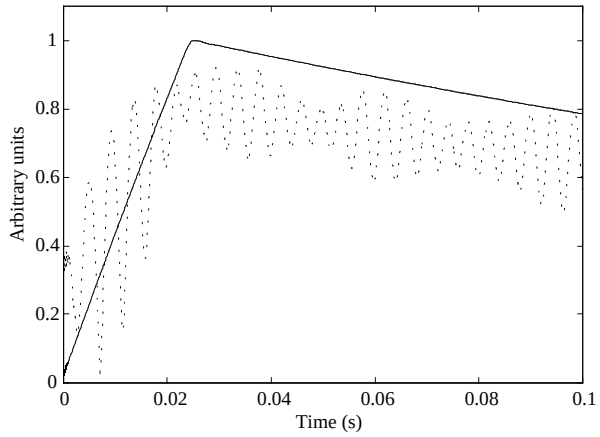

Figure 5: *Envelope functions of excitation (solid line) and deconvolved response (dotted line) in the range 0÷0.1 s.*

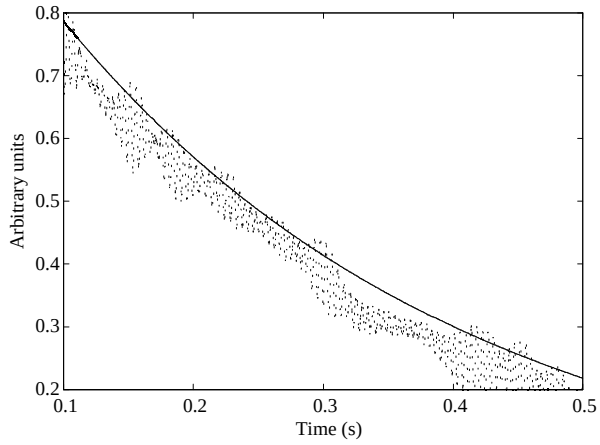

Figure 6: *Envelope functions of excitation (solid line) and deconvolved response (dotted line) in the range 0.1÷0.5 s.*

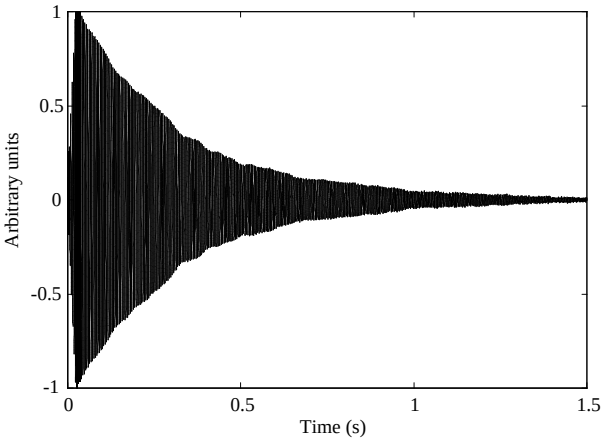

Figure 7: *Deconvolution of the filtered signal.*

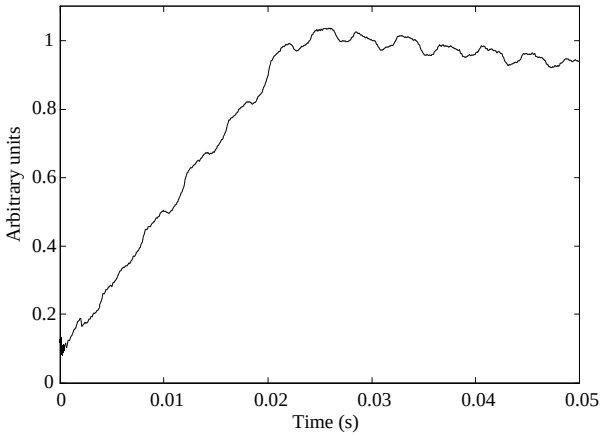

Figure 8: *Envelope function of the signal obtained from the deconvolution of the filtered response.*

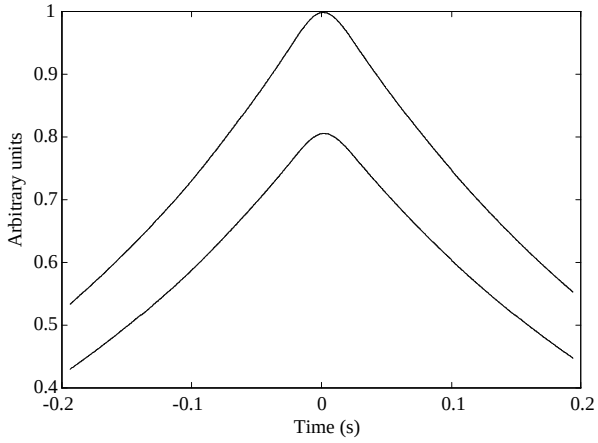

Figure 9: *Cross correlation of excitation and recovered signal (envelopes). Upper line: y filtered. Lower line: y unfiltered.*

two obtained values are $\xi \simeq 0.805$ and $\xi \simeq 0.997$, the second one denoting a very good correlation of x' to x. Anyway, a time correspondence is present in both cases as indicated by the function shape, which is almost perfectly symmetrical around the time origin: the noise is just responsible for the observed scale relationship $R_{xx'}(\tau) = kR_{xx',fil}(\tau)$ where $k \simeq 1.2$.

Besides the analysis of the deconvolution from the processing viewpoint it is important to show the importance of the above mentioned physical restrictions related to the source-receiver position. To this purpose, we made impulse response measurements with the receiver displaced from the position occupied during the recording: a subsequent deconvolution yielded waveforms definitely dissimilar to the original stimulus. As an example, Fig. 10 reports a deconvolution executed with an impulse response taken about $1\,m$ from signal y.

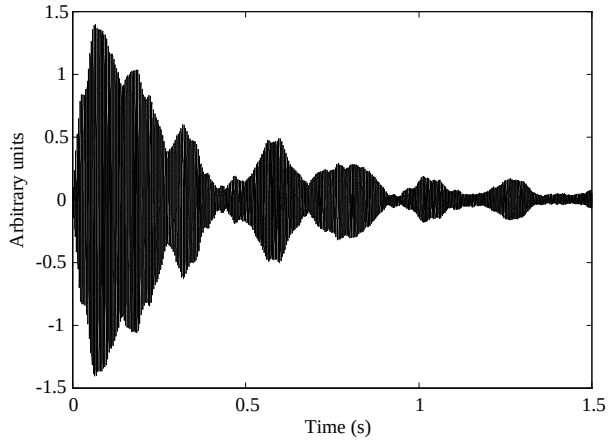

Figure 10: *Deconvolution of signals at different locations.*

4. CONCLUSIONS

A test for evaluating the effectiveness of the basic deconvolution procedure for the recovery of minimum-phase signals has been discussed. Results similar to those here presented have been obtained for several examples of signals, though the accuracy of the procedure proved to be somewhat dependent on the frequency content of the excitation. This behavior is extremely important for investigating possible limitations inherent in the linear model assumption: a rigorous study will be undertaken in our following research, which will be extended to mixed-phase signals as well.

A quantitative characterization of the results has been achieved by the comparison of the envelopes and by examining the time history behavior through the cross-correlation function: this has allowed us to appreciate the way how random noise is amplified by the raw deconvolution. Finally, the critical role played by the source-receiver placement has been highlighted. In this context, an issue of great interest is the development of a deconvolution procedure less sensitive to spatial constraints and therefore more useful in practical situations.

5. REFERENCES

[1] Stanzial, D, Il controllo attivo del tempo di riverberazione: impostazione del problema, Proc. XIX AIA meeting , Napoli, Italy, pp. 241-244, 1991.

[2] Morse, P.M. and Ingard, K.U., Theoretical Acoustics, Princeton University Press, Princeton, pp. 319-322, 1986.

[3] Tohyama, M. and Tsunehiko, K., Fundamentals of Acoustic Signal Processing, Academic Press, pp. 276-278,1998.

[4] Oppenheim, A.V. and Schafer, R.W., Discrete-Time Signal Processing, Prentice Hall, pp. 781-785, 1989.

ACKNOWLEDGEMENTS

This work has been fully supported by Fondazione Scuola di San Giorgio c/o Fondazione G. Cini, Venezia, Italy.