Finding the onset of a room impulse response: Straightforward?

Guillaume Defrance, Laurent Daudet, and Jean-Dominique Polack
UPMC University Paris 06, Institut Jean Le Rond d'Alembert, LAM CNRS UMR 7190, Ministère de la Culture et de la Communication, 11 rue de Lourmel, 75015, Paris, France
defrance@lam.jussieu.fr, daudet@lam.jussieu.fr, polack@ccr.jussieu.fr

Abstract: This letter deals with precision issues in the determination of the timing of the room impulse responses (RIRs) onset. First, it is shown that while errors of onset timing estimation do not have that much effect on temporal indices, an erroneous onset estimation leads to significant differences in energetic and statistic acoustical indices. Twelve automatic onset detection methods are compared, in terms of precision, robustness, and complexity. Experimental validation made on a set of 100 RIRs provides good evidence in favor of spectral and/or energetic methods, according to the type of sound source.

PACS numbers: 43.55.Br, 43.55.Mc, 43.58.Gn [NX]
Date Received: April 22, 2008 **Date Accepted:** June 13, 2008

1. Introduction

A standard way to document the acoustics of a room is to measure a set of room impulse responses (RIRs). A RIR should ideally be recorded within absolute silence, a condition which of course is never met in practice. It then belongs to the acoustician to identify the edges of the RIR. Acousticians have for a long time designed various methods, as attested in the ISO 3382[1] standard, for determining the last point of the reverberation tail, which is mixed with background noise. However, nothing is said about the onset of the RIR, implicitly assuming that this is a straightforward task: Each acoustician, or software, can have its own method. In order to find the onset time, the simplest ways that come first to mind would be (1) to determine it manually, i.e., visually on the waveform; or (2) to consider that the RIR always starts by direct sound, which often has the largest amplitude of the signal. Measurements in a concert hall typically lead to 100 RIRs, so that the first method becomes easily cumbersome. The second method, as we shall see, can in some cases provide poor estimates of the onset time. For instance, when strong scattering attenuates the direct sound, it is the first reflection that presents the largest amplitude, and the RIR maximum and its beginning can differ from less than 0.1 ms up to more than 50 ms [Fig. 3(b)]. Therefore, direct comparison between different authors and/or methodologies may become unreliable.

The first goal of this letter is to show that, for some of the most commonly used room acoustical indices, the above methods lead to large errors. The ISO 3382 standard lists some of these indices. Energetic and statistic, such as Clarity (C_{80} in dB, at 80 ms) and Central Time (T_C in ms), are derived from a ratio of integration of the energy of the RIR; while temporal indices, such as Early Decay Time (EDT_{10}) and the Reverberation Times (RT_{20} and RT_{30}) are obtained from a linear regression made on the integration of the energy of the RIR. We do not attempt to discuss the relevance of such indices, be they orthogonal or not. This has already been done, for instance, by Pelorson *et al.*[2] We do attempt, however, to propose and evaluate a number of different automatic onset detection methods for their precision, robustness, and complexity. This is the second goal of this letter.

The letter is constructed as follows. In Sec. 2, we investigate how the precision for the localization of RIR onsets influences the result for five commonly used acoustic indices. Section 3 describes 12 methods for the automatic determination of RIR onsets. The performance of

Table 1. Variations of acoustical indices as a function of errors in the onset time Δt_0: $\Delta EDT_{10}(\%)$, $\Delta RT_{20}(\%)$, $\Delta RT_{30}(\%)$, $\Delta C_{80}(dB)$, $\Delta T_C(\%)$

Δt_0 (ms)	$\Delta EDT_{10}(\%)$	$\Delta RT_{20}(\%)$	$\Delta RT_{30}(\%)$	$\Delta C_{80}(dB)$	$\Delta T_C(\%)$
0.05	0.002	0.002	0.002	0.25	3.6
0.1	0.005	0.006	0.006	0.31	4.5
0.2	0.01	0.01	0.01	0.37	5.3
1.0	0.05	0.06	0.06	1.29	17.6
2.0	0.11	0.11	0.12	1.42	18.7
4.0	0.23	0.23	0.024	1.35	16.5
10.0	0.57	0.59	0.6	0.92	9.9

these methods is empirically compared in Sec. 4. The last section (Sec. 5) discusses guidelines for choosing the right method according to the application at hand, and concludes on the need to present reliable results with documented uncertainty.

2. Precision of onset timing versus precision of the indices

Assuming that t_0 is the reference time index corresponding to the maximum of the RIR, we compute on one full-band typical RIR (from Salle Pleyel, Paris[3]) the EDT_{10}, RTs, C_{80}, and T_C indices using different onset times $t' = t_0 - \Delta t_0$, with Δt_0 varying from 0.05 to 10 ms. Table 1 shows how acoustic indices can be affected differently by differences on onset timing. Even with $\Delta t_0 = 10$ ms, errors on EDT_{10} and RTs still remain weak (around 0.6%). The authors would like to point out that calculating the EDT according to Jordan's definition[4] leads automatically to large errors, since it looks for the time that the total energy has decayed by 10 dB.

On the contrary, $\Delta t_0 > 2$ ms leads to a difference in clarity of 1.45 dB (i.e., 55%). Variations of T_C remain inferior to 20% for the RIR tested here, with similar values for other RIRs. This way, even if the method used to determine the onset index is the maximum of the RIR, the user should prefer to calculate the central time than the clarity, because they both refer to the density repartition of the energy in the signal. This simple example shows the importance of exact onset detection of impulse responses in room acoustics. As errors on temporal indices are low, and would be considered as insignificant by most experts, as documented in Ref. 5, they will not be discussed in this letter. Nevertheless, attention is paid to energetic and statistic indices, since they have a strong dependence on the precision of the onset timing.

3. Proposed onset detection methods

This letter presents 12 different methods for onset detection. As the study of the acoustics of a hall can lead to typically hundreds of measurements (hence 100 RIRs), this article is focused on methods that can be computed automatically. The methods presented below rely on the idea that the onset is linked to abrupt changes in one or more properties of the audio signal,[6] and can therefore be detected by detecting the changes.

3.1 Temporally based methods

When observing the temporal structure of a RIR, it is noticeable that the occurrence of an onset is accompanied by a sudden increase of amplitude. The first methods of onset detection were based on this property by using a detection function which follows the envelope of the signal.[6]

Four simple functions are presented here: (1) Maximum (M): As seen in introduction, the first idea is to consider the onset t_0 to be the maximum of the absolute value of the RIR. (2) Maximum minus 5 ms (M_5): A few softwares, such as the MIDAS package,[7] consider that the onset can be defined 5 ms before the maximum of the RIR. (3) Mean over time (D_E): An envelope follower can easily be constructed by low-pass filtering the local energy.[6] The maximum of $E(n)$ is detected, and the signal analyzed from its beginning to its maximum. The ratio of two

successive windows is calculated; the index is found when the ratio is maximum [Eq. (1)]. This can be written as:

$$E(n) = \sum_{m=-N/2}^{N/2} w(m)x^2(n\,.\,h+m), \tag{1}$$

$$t_0 = h \times \arg\max_n(E(n+1)/E(n)), \tag{2}$$

where w is a smooth windowing function, and h the time step between two windows.

(4) Threshold (E): This method works on the energy of the RIR, which is windowed by rectangular windows $w(m)$; t_0 is here defined as the first time index where this local energy is K (typically $K=3$) times higher than its median running on past windows.

3.2 Spectrally based method

Since the spectra of the direct sound and the first reflections of an impulse response are very different from the background noise, a function based on spectra comparison is expected to give to good results, as seen in Ref. 8. One can expect an increase of low frequency components when looking at the direct sound.

(5) Mean over spectra (D_S): The idea is almost identical to D_E method, but ratios are calculated over spectra [Eq. (3)].

$$\tilde{E}(n) = \sum_k |X(n\,.\,h,k)|, \tag{3}$$

$$t_0 = h \times \arg\max_n(\tilde{E}(n+1)/\tilde{E}(n)), \tag{4}$$

where $X(n,k)$ is the short time Fourier transform of the signal $x(n)$, and h the time step between two windows.

3.3 Time frequency method

(6) Wavelet Transform (W): This method, often used for denoising,[9] is a natural tool for analyzing transient signals, since its time frequency resolution provides an increasingly finer time resolution at smaller scales. Let Ψ be a zero-mean real Gaussian wavelet. The wavelet transform of $x(t)$ is defined as

$$Wx(u,s) = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{s}}\Psi^*\left(\frac{t-u}{s}\right)dt, \tag{5}$$

where u is the translation parameter and s the scale factor.

In other words, $Wx(u,s)$ measures variations of $x(t)$ near u within an equivalent window of size s (Eq. (5)). When $s \to 0$, the decrease of the wavelet coefficients characterize the regularity of x around u. The onset of the transient signal is estimated by applying a threshold to the wavelet coefficients,[9] such as

$$Tm = \sigma_m\sqrt{2\log_e(N)}, \tag{6}$$

where σ_m is the standard deviation of the noise, and N the length of the signal. After inverse wavelet transform, the onset is estimated as the first nonzero sample.

3.4 Refinement step

The six methods presented above can be refined by a method based on linear prediction (L_P). Linear prediction analysis finds the coefficients of a finite impulse response linear filter that predicts the current value of the real-valued time series based on past samples, minimizing the prediction error in the least squares sense. L_P residual is computed on adjacent windows [t

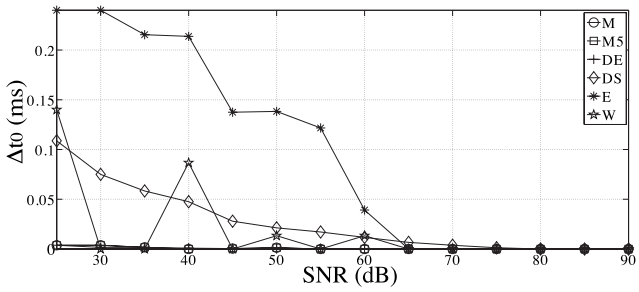

Fig. 1. Variations of onset times estimation as a function of *SNR*.

$-m:t]$ and $[t:t+m]$, with $m=0.8$ ms; t_0 is the time index that maximizes the likelihood of having a stationary Gaussian residual in both backward and forward windows, with a χ^2 goodness-of-fit test.[10] In other words, t_0 corresponds to a change point in the behavior of the system. This method is potentially very precise, but cannot be used as such, since it detects any spurious event regardless of its size. Instead, it is used only as a refinement step locally around the times provided by the previous methods. We call the six improved methods: ML_P, $M5L_P$, D_EL_P, D_SL_P, EL_P, and WL_P.

3.5 Reference method

An important issue regarding the evaluation of these methods is that the exact onset timing is not known (no absolute truth). These reference points could be determined manually by experts, at the cost of a tedious hand labeling and a potential lack of consistency between experts. Here, we decide to choose method D_SL_P as reference. The rationale for choosing this method is the following. First, on all the RIRs that we have at hand, D_E and D_S are the only methods which always return onset indices validated at hand by the authors (this is not always the case for M, M_5, E, W). As differences of onset estimation between methods D_E and D_S are inferior to 0.1 ms, we need to look at the enhanced methods. Indeed, method D_EL_P estimates the onset at the beginning of a small ripple that we call *precursor* (Figs. 3(c) and 3(d)), while method D_SL_P detects the onset at the end of the precursor. It is assumed that the precursor does not belong to the RIR itself, but is either an artifact of the pistol shot (maybe due to the cylinder rotation), or more probably an artifact of an anti-aliasing filter in the A/D conversion. Second, we assume that the most precise method should return the best onset index. L_P being a refinement step, this leaves D_SL_P as reference.

4. Comparison of the 12 automatic onset detection methods

4.1 Evaluation method

The 12 detection methods are tested over 100 audio wavfiles. These RIRs have been measured in Salle Pleyel in Paris,[3] according to the international standard,[1] with pistol shots as sound sources. For each RIR, the onset time is estimated by each of the 12 methods. From this, acoustical indices (C_{80}, T_C) are calculated and compared to the reference value, returned by method D_SL_P. Onset times that are not roughly consistent with the reference onset (i.e., not within 100 ms) are not taken into account for the statistics.

4.2 Robustness to noise

A way to assess the robustness of the presented estimators is to vary the background noise level of the RIRs. Figure 1 shows the mean variations of the onset time estimated by the first six methods on 100 experimental RIRs, with a signal to noise ratio (SNR) varying from 25 to 90 dB, by step of 5 dB. For each method, the reference onset time is the one obtained without adding noise to the signal. As expected from Sec. 3.3.3, the most robust method is W, followed by M and M_5; the worst method is E, followed by D_S.

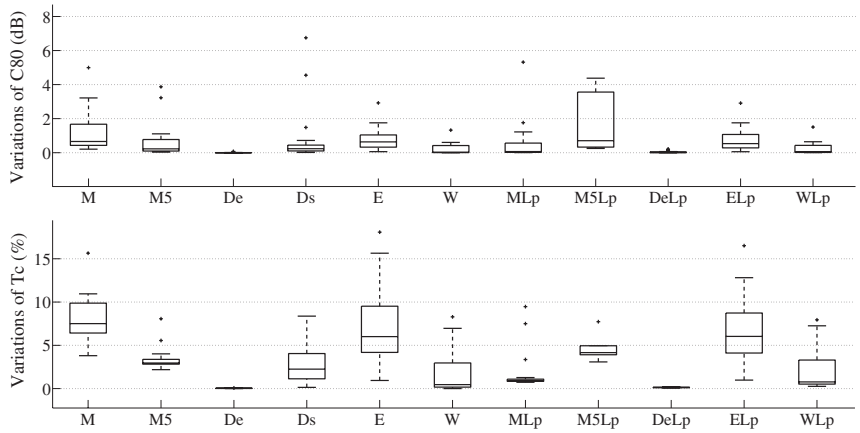

Fig. 2. Variations of C_{80} (top) and T_C (bottom) for the 11 methods, $D_S L_P$ being taken as reference.

4.3 Results and discussion

Except for large variations, which are discussed later, results (Fig. 2) show that adding the method L_P to any other method considerably improves the accuracy of onset detection, and hence, decreases errors, except for method M_5.

Methods M and M_5 are extremely robust to artifacts, but present important variations, since they do not account for a potential scattering effect. Figure 3 presents three different RIRs. The first one (a) starts very near its maximum. One can expect that detection results should not vary from a method to another. The second RIR (b) presents a long scattering effect ($\approx$30 ms), caused by a balcony. Methods M and M_5 provide bad estimations in that case. In cases (c) and (d), the onset is not the maximum, but a *precursor*, as introduced in Sec. 3.3.5. These small variations cannot be detected by M and M_5, but also by method E, since its threshold is not always adapted to the RIR's precursor amplitude. Moreover, because of its threshold, that the user has to set differently for each RIR, E is not suited to an automated analysis.

Methods D_E, D_S, and W seem to be particularly indicated for a quick and precise onset index determination, method W being more robust to noise than the other methods, as seen in Sec. 4.4.2. The resulting estimated C_{80} is always within ±0.01 dB (i.e., ±3%) of our reference method given by $D_S L_P$. Indeed, even if method L_P improves on other methods, the computation time is significantly increased for a negligible gain in precision (typically below 0.1%).

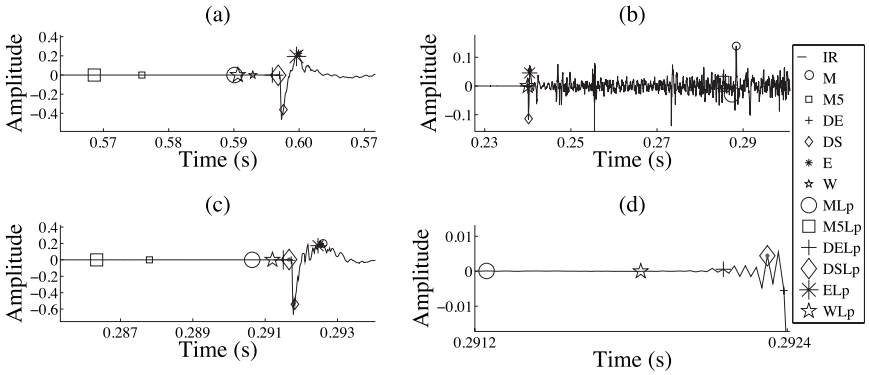

Fig. 3. Example of three different RIRs (note the different time scales). a) Simple case onset; b) Scattering effect; c) Precursor; d)Details of a precursor presented in (c).

4.4 Measurements with balloon bursts

The same experiment has been carried out on 100 RIRs measured with balloon bursts in Salle Pleyel (for the same source and receiver positions). Method $D_S L_P$ being also taken as reference, results slightly differ from those obtained with the pistol shots. Although method L_P improves all methods, except M_5, method D_S offers the best estimate ($\bar{\Delta}_{C_{80}}$=0.02 dB, $\sigma_{C_{80}}$=0.02 dB; $\bar{\Delta}_{T_C}$=0.3%, σ_{T_C}=0.01%), instead of D_E for pistol shots. This can be explained as follows. The pistol shot has a much sharper increase of energy than balloon burst,[11] typically 1.5 times faster. Thus, differences between pistol shots and balloon bursts are both spectral and temporal, and also related to their directivity, as explored in Refs. 11 and 12. Further studies are required to refine these claims. It should be noticed that nonlinear effects that may affect pistol shots are not detected by these methods since they intrinsically are parts of the response.

5. Conclusion

The main goal of this letter is to raise awareness on a loophole in the ISO 3382 standard for the computation of room acoustics indices. If one is interested in statistic and energetic indices such as T_C and C_{80}, a robust and precise method for determining the onset time is necessary. For instance, a variation of only 2 ms can generate high variations of clarity (around 1.5 dB, i.e., 55%). Furthermore, it is shown that for some of RIRs, there are large differences in the results given by obvious used onset detection methods, inducing significant differences in the acoustic indices.

Our experimental tests also show that methods based on energetical differences (method D_E), for pistols shots, on spectral differences (D_S) for balloon bursts, and on time-frequency analysis, such as wavelet transform (W), seem to provide reliable estimates, with a precision that is appropriate for most uses. Nevertheless, the computation time needed by method W, and the slight difference with results obtained with D_S, do not justify its use. These results highlight the inner spectral and temporal differences of frequently used sound sources. A statistically based refinement method is also a viable approach, but the slight gain in precision does not seem to justify the additional computational complexity. Further studies should test these methods on other sets of RIRs, including Ambisonics measurements of Salle Pleyel, and adjust the different important parameters such as window lengths, for an extended set of indices.

One may question the expected outcome of such detection algorithms. In the case of complex RIRs such as those with a precursor, what is the most relevant onset time from a perceptual point of view? Finding the beginning of the precursor may not be the best choice, since the precursor could be inaudible due to temporal masking effects. Such studies that are definitely beyond the scope of this letter, would require extensive listening tests. However, they remind us that these automatically generated signal processing indices are only meaningful if they provide information that has a *perceptual* relevance.

Acknowledgment

This work was partly supported by grants from Region Ile-de-France.

References and links

[1]ISO 3382, Acoustics-measurements of the reverberation time of rooms with reference to other acoustical parameters (1997).
[2]X. Pelorson, J.-P. Vian, and J.-D. Polack, "On the variability of room acoustical parameters: Reproducibility and statistical validity," Appl. Acoust. **37**, 175–198 (1992).
[3]G. Defrance, J.-D. Polack, and B.-F. Katz, "Measurements in the new Salle Pleyel," in *Proc. Int. Symp. Room Ac.* (Sevilla) (2007).
[4]V. L. Jordan, "A comprehensive musical criterion: The inversion index," JAES **23**, (2) 131–135 (1975).
[5]X. Meynial, J.-D. Polack, and G. Dodd, "Comparison between full-scale and 1:50 scale model measurements in theatre municipal, Le Mans," Acta Acust. **1**, 199–212 (1993).
[6]J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in musical signals," IEEE Trans. Speech Audio Process. **13**, 1035–1047 (2005).
[7]X. Meynial, G. Dodd, J.-D. Polack, and A.-H. Marshall, "All-scale model measurements: The MIDAS system,"

in *121st ASA Meeting, special session on auditorium measurements*, Baltimore (1991).
[8]D. J. Hermes, "Vowel-onset detection," J. Acoust. Soc. Am. **87**, 866–873 (1990).
[9]S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1999).
[10]M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes—Theory and Application* (Prentice–Hall, Englewood Cliffs. NJ, 1993).
[11]A. Nash, "On the acoustical characteristics of a balloon," in *Proc. Int. Symp. Room Acoustics (Sevilla)* (2007).
[12]D. Griesinger, "Beyond MLS-occupied hall measurement with fft techniques," 2004, URL http://world.std.com/~griengr/sweep.pdf.