

An improved method to determine the onset timings of reflections in an acoustic impulse response

John Usher

Barcelona Media – Innovation Centre, Avenida Diagonal 177, 08018 Barcelona, Spain
john.usher@barcelonamedia.org

Abstract: Determining the absolute onset time of reflections in an acoustic impulse response (IR) has applications for both subjective and physical acoustics problems. Although computationally simple, a first-order energetic analysis of the IR can lead to false-positive identification of reflections. This letter reports on a method to determine reflection onset timings using a modified running local kurtosis analysis to identify regions in the IR where the distribution is non-normal. IRs from both real and virtual rooms are used to validate the method and to find optimum analysis window sizes.

© 2010 Acoustical Society of America

PACS numbers: 43.55.Br, 43.55.Mc, 43.58.Gn, 43.60.Cg [NX]

Date Received: November 4, 2009 **Date Accepted:** February 18, 2010

1. Introduction

The absolute timings of early reflections in a room impulse response (RIR) can be used to solve a variety of problems relating to both subjective and physical acoustical phenomena in the corresponding environment (be it real or virtual). Although slight imprecision in the timing detection will not dramatically change coloration nor spatial impression, since both are evaluated at a more global level, the initial time-delay gap (ITDG) can affect perceptual integration of the direct and reflected sound (i.e., the precedence effect) and may affect speech intelligibility or musical timbre. Physical characteristics of a space can also be determined from precise timings of reflections in an acoustic IR; a process commonly used in SONAR, seismic exploration, and medical diagnosis.

The most computationally simple method to identify the reflection onset timings (ROTs) in a RIR is from an analysis of the energy envelope. There are a variety of approaches to this first-order energy profile analysis for onset detection within an audio signal, such as using tuned edge-detection wavelet filters or spectral-domain energy fluctuations (for instance, this has been applied to finding the onset of transient events in music audio signals¹). It is the goal of this short letter to first describe the short-comings of using an energy profile analysis to determine ROTs in a RIR, and to second introduce an improved method based on a local running statistical analysis of the RIR. The analysis is particularly directed toward identifying low-order early reflections after the direct-sound.

2. Characterizing a room impulse response by ROTs

2.1 Energy profile analysis vs statistical energy analysis

To demonstrate the short-coming of using a first-order energetic analysis to determine ROTs, we refer to the RIR described in Fig. 1 that was measured in an open-ended room of dimensions $1.10 \times 2.73 \times 4.9$ m (w, h, d). By analysis of an image-source model constructed using the same acoustic measurement parameters, we can verify that the “true” first reflection actually corresponds to the peak marked with the box symbol in Fig. 1. However, using a simple analysis of the energy envelope of the RIR, the “false” reflection peak marked with the cross symbol shortly after the direct-sound peak would be mistaken as the first reflection. Such false reflections in the energy profile may be a result of distortion from anti-aliasing filters in the analog to digital converter system or from nonlinearities in the excitation and measurement transducer systems. Precise reflection onset

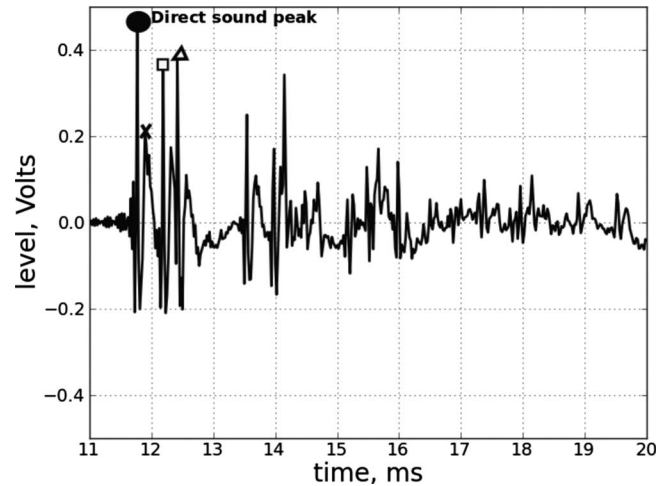


Fig. 1. RIR measured in an open-ended room of dimensions $1.10 \times 2.73 \times 4.9$ m (w, h, d). The true first and second arriving reflections were determined using a room model and are marked with the square and the triangle. The false reflection peak is marked with a cross, clearly indicating the problem of using an energy profile analysis to determine the timing onset of reflections in a RIR.

detection in a RIR is a problem further compounded by scattering effects from room surfaces, which can introduce a Poisson-like distribution for early reflections from averaged acquisition methods (such as maximum length sequence or swept-sine techniques).²

Statistical approaches to identifying reflection timings in an acoustic impulse response can overcome the problems with an energy analysis by utilizing the second-order statistics of a RIR. Reverberation is generally defined as a physical phenomenon where the sound reflections impinging on a point are such that the pressure can be modeled as a stochastic function with a normal distribution and a mean of zero, and that part of a RIR that has a normal distribution is called the reverberant part.³ The degree to which a part of a RIR has a normal distribution (the “degree of normality”) has been used before to determine the onset of the reverberation. Two such measures of normality are “average rms response fluctuation” and kurtosis. Schroeder’s frequency-domain stochastic model³ predicts that 68.26% of the response curve will lie within about 10 dB of the average level. Similarly, the Abel and Huang criterion⁴ for the reverberant part of a RIR is based on the well-known phenomenon that 68.26% of the observations of a normal population will be found within 1 standard deviation of its mean.⁵

Another well established and understood measure of the degree of normality is *kurtosis*. Kurtosis can be considered as a measure of the degree of peakedness of a distribution, or more correctly as the degree of bi-modality. The kurtosis k of a data series x is defined as the ratio of its fourth central moment to its standard deviation raised to the power of 4 (hence kurtosis is always positive) as follows:

$$k = \frac{E\{x - \mu\}^4}{\sigma^4}, \quad (1)$$

where $E\{\cdot\}$ is the statistical expectation operator, μ is the mean, and σ is the standard deviation of the data series x .

To develop a running kurtosis analysis of data x , consider a block of samples centered about discrete time sample n as follows:

$$\mathbf{x} = [x_{n-m/2}, x_{n-m/2+1}, \dots, x_{n+m/2-1}].$$

The running kurtosis of \mathbf{x}_n is now defined as

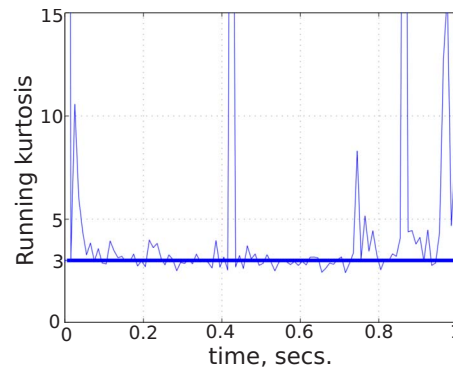


Fig. 2. (Color online) Sliding kurtosis echogram of a RIR from a 3000 m³ concert hall. The reverberant component is defined when the kurtosis is equal to 3. Strong late individual reflections can be discerned, but fine-structure in the early part cannot. A sliding 10 ms rectangular window was used to calculate kurtosis according to Eq. (2).

$$k_n = \frac{(x_n - \mu_n)^4}{\sigma_n^4}, \quad (2)$$

where μ_n and σ_n are the mean and standard deviation of the m -length vector \mathbf{x}_n . If sample x_n is large compared with the local samples \mathbf{x}_n then the kurtosis will be high (*leptokurtic*); a flat distribution in \mathbf{x}_n —a low peakedness—will have a low kurtosis (*platykurtic*); and a normal distribution has a kurtosis of 3. However, a bi-modal distribution will also give a low kurtosis; so therefore low kurtosis values are not necessarily an indication of low peakedness.⁶ Kurtosis has been used before for acoustic IR analysis: for determining the optimum length of adaptive filters for applications such as blind audio upmixing⁷ and for determining the onset of musical instruments in an audio signal.¹ In a RIR, spurious large-value samples from strong individual reflections will clearly have a leptokurtic distribution, giving a kurtosis greater than 3. It is this principle that forms the basis of the proposed method for identifying the onset of early reflections in a RIR.

2.2 Modified kurtosis method

Similar to the “echogram” or “reflection diagram” of Kuttruff,⁸ which shows the absolute strength (i.e., acoustic power) of reflections impinging on a point over time, we can show a plot of the RIR kurtosis as a function of time (though the magnitude of the kurtosis is not necessarily related directly to the strength of the reflections). Such a “kurtosis echogram” for a RIR measured in a concert hall is shown in Fig. 2, where the onset of the strong first-order reflection from the rear of the concert hall can be discerned about 0.4 s after the direct-sound. However, fine-structure detail of the early impulse response is unclear and individual early reflections cannot be discerned.

The temporal resolution of the running kurtosis echogram is affected by two parameters: the range over which the local kurtosis is calculated (i.e., the window size m) and the overlap between successive kurtosis frames. As we saw in Fig. 2, a long temporal window smears the fine-structure of the early impulse response. This can be understood analytically by considering the reflection density of the RIR. The reflection density (reflections/s) can be approximated by

$$\frac{dN}{dT} = 4\pi \frac{c^3 t^2}{V}, \quad (3)$$

where c is the speed of sound, V is the room volume, and t is the time since the RIR onset. It has been reported that this relationship holds not just for rectangular rooms but also for any room shape.⁸ Using Eq. (3), we can determine that for the first 6 ms of our 15 m³ test room the

echo-density is approximately 1 arrival/ms, corresponding to 1 reflection arriving every 5 samples for a 48 kHz sample-rate system. If the kurtosis analysis window is less than the inter-reflection time then the expected difference between the instant RIR sample value x and the local mean μ may be lower than the RIR mean (or the RIR mean calculated with a larger local window) and likewise the kurtosis will be low even when a reflection is within the analysis window. Alternatively, if the kurtosis window is too large then the temporal resolution for kurtosis will be smeared, and a bi-modal distribution from double peaks within the analysis window can also give a low kurtosis.⁶ We therefore introduce a modified kurtosis metric that is a function of two analysis windows: a first small window to estimate a “fast” local estimate of the RIR level [this replaces the single sample value x in Eq. (1)], and a second longer local window over which a “slower” mean and standard deviation are calculated. The modified kurtosis estimate is now defined as

$$k_{n,l,m} = \frac{(\mu_{n,l} - \mu_{n,m})^4}{\sigma_{n,m}^4}, \quad (4)$$

where l and m are the lengths of a first and second data window, both centered about sample n of the data series \mathbf{x} ; $\mu_{n,l}$ and $\mu_{n,m}$ are the means of the l -length and m -length local data windows ($m > l$); and $\sigma_{n,m}$ is the standard deviation of the m -length window. An empirical investigation was conducted to find optimum window lengths l and m .

3. Results and discussion

The short window size for the estimate of the fast running-average, i.e., the length l , was four samples. This gave sufficient temporal resolution to capture a single reflection; and yet the time of arrival between two consecutive reflections in the early part of a RIR is generally at least four samples for small rooms, as discussed previously with reference to the Kuttruff echo-density trend in Eq. (3). The window size for the longer local IR window, i.e., the length m , was varied from being 2 to 16 times the length of the four-sample window. Figure 3 show that as the ratio of the long to small window (i.e., m/l) is increased, then the resulting “kurtosis echogram” becomes less noisy and the kurtosis peaks can be more easily discerned. However, for m/l ratios greater than 16, the width of these peaks increases, making it difficult to precisely identify the timing of peak maxima.

The criterion for selecting the optimum m/l ratio was by a difference between the ROTs of the measured RIR using the new two-window running kurtosis method, compared with the true ROTs as determined using the analytic image-source room model. The empirical ROTs were determined by analysis of the kurtosis echograms shown in Fig. 3 using a peak-picking inspection of the kurtosis data for peaks greater than a value of 3. The measurement was repeated with six different source-receiver locations (in the same room), and the real-world ROTs were compared with the virtual ROTs determined from the corresponding room model. For the first six ROTs, there was a 100% correspondence between the real and virtual ROTs for six different source-receiver test configurations when the m/l ratio was equal to 16. However, these exploratory findings used source-receiver locations that were chosen to be non-symmetrical about the room axes. For source-receiver locations centered near the central room axes, we would find that ROTs nearly coincided and would therefore decrease the inter-reflection time to a value less than the smallest analysis window yielding a bi-modal distribution with a low kurtosis; hence the reflection onsets may not be detected. Accordingly, we conclude that the kurtosis analysis method is effective for RIR echo-densities less than approximately 1-per-ms, which corresponds to approximately the first 80 ms of the RIR for a 3000 m³ concert-hall sized room [predicted using Eq. (3)]. Although the analysis was conducted on full-band audio signals, the method for identifying ROTs is applicable for filtered IRs, such as octave of third-octave filtered time signals.

4. Conclusion

A new method has been proposed to determine the onset of individual reflections in the early part of a RIR. The method is based on a running local kurtosis analysis with two analysis win-

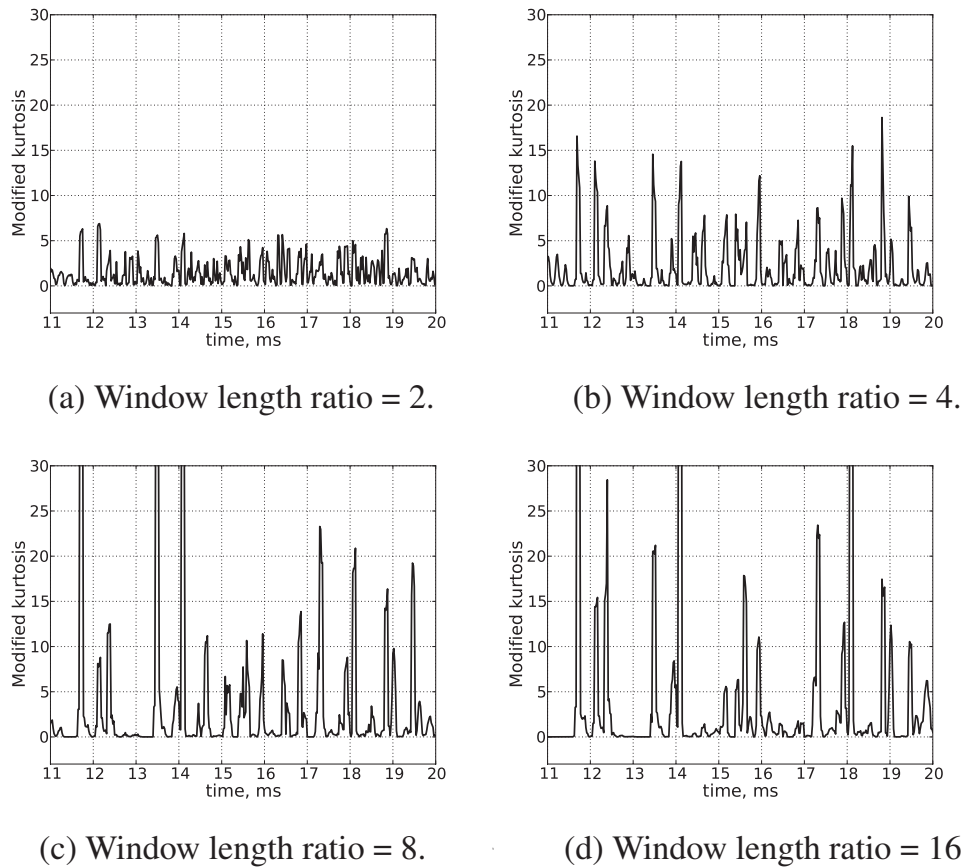


Fig. 3. Modified running kurtosis analysis of the RIR in Fig. 1. The window length ratio corresponds to m/l in the modified kurtosis equation [Eq. (4)], where the short window length l was always equal to four samples. Note that the false peak identified with the cross symbol in Fig. 1 was not detected.

dows to identify regions in the RIR where the distribution is non-normal. RIRs determined from both real and virtual rooms were used to validate the new method and to find optimum analysis window sizes, where a first window size of 0.8 ms and a second of 12.8 ms were found to accurately identify reflection onsets for the early part of the RIR. The effectiveness of identifying reflection onsets was found to be related to the relative length of the smaller window and the reflection density: The new method was reliable for densities less than 1-per-ms.

Acknowledgments

This work was partly supported by grants from iMP, Call No. FP7-ICT-2007-3 and Objective ICT-2007.4.4. The author is grateful for the comments from three anonymous reviewers.

References and links

- ¹J.-P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in music signals," *IEEE Trans. Speech Audio Process.* **13**, 1035–1047 (2005).
- ²G. Defrance, L. Daudet, and J.-D. Polack, "Finding the onset of a room impulse response: Straightforward?," *J. Acoust. Soc. Am.* **124**, EL248–EL254 (2008).
- ³M. R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J. Audio Eng. Soc.* **35**, 299–305 (1987).
- ⁴J. S. Abel and P. Huang, "A simple, robust measure of reverberation echo density," in *Proceedings of the 121st AES Convention*, preprint 6985 (2006), pp. 1–10.

⁵A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. (McGraw-Hill, New York, 2002).

⁶R. B. Darlington, "Is kurtosis really peakedness?," *Am. Stat.* **24**, 19–22 (1970).

⁷J. Usher and J. Benesty, "Enhancement of spatial sound quality: A new reverberation-extraction audio upmixer," *IEEE Trans. Audio, Speech, and Lang. Process.* **15**, 2141–2150 (2007).

⁸H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Science, Essex, UK, 1991).