CHARACTERIZING SOUND SOURCES FOR ROOM-ACOUSTICAL MEASUREMENTS

G. Defrance IJLRDA, LAM Team, UPMC Univ. Paris 06, CNRS UMR 7190,
L.Daudet Ministère de la Culture et de la Communication,
J-D Polack 11 rue de Lourmel, 75015, Paris, FRANCE
defrance@lam.jussieu.fr

1 INTRODUCTION

In room acoustics, room impulse responses (RIR) are measured to characterize quantitatively and qualitatively the acoustics of a hall. A convenient manner to measure RIRs is to use impulsive sound sources, such as: balloon bursts, gun shots, and crackers. However, unlike sweep or white noise, this kind of sources is known to be hardly reproducible. That is, the spectral and temporal components may differ from a measure to another. According to the nature of the source, sound source impulses do not have the same duration[1,2,3]. Moreover, for some experimental RIRs carried out with such sources, it becomes difficult to identify clearly these boundaries when the sound source is not recorded in the near field, or in a damped chamber. The identification of these boundaries may vary from an expert to another, since the background noise may often disturb the readability of the signal, as seen in Figure 1.

This study investigates two methods for estimating the impulse duration of some impulsive sound sources (cited above). The knowledge of the direct sound provides useful information on the sound source itself, and allows to whiten the RIR. This study aims at going one step further in the characterization of some frequently used sound sources in room acoustics measurements.

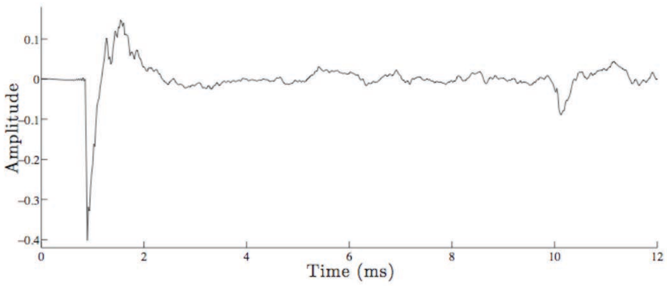

Figure 1: Experimental RIR for which determining the direct sound is not obvious.

Section 2 proposes to visualize temporal and spectral differences of such sources using the smoothed Wigner-Ville distribution. Section 3 presents two methods for estimating the temporal boundaries of the direct sound of experimental RIRs. The first method (Section 3.1) consists in low-pass filtering the local spectral energy of the signal, using the Short Time Fourier Transform (*STFT*), with high precision in time. Since this method shows well-known limitations, a second method, Matching Pursuit (MP) (Section 3.2), focuses on another approach, based on the inter-correlation between the RIR and the direct sound itself[4]. The estimation of the direct sound duration is achieved by learning the dictionary of atoms needed by MP. Section 5 presents experimental results derived from a hundred of measurements of RIRs in Salle Pleyel carried out with balloon bursts on the one hand, and with gun shots and crackers on the other hand[5].

Finally, the last section (Section 6) compares the results to impulses duration measured in a damped chamber, as well as to the relevance of each method, and discusses guidelines for choosing the best sound source.

2 VISUALIZING THE DIRECT SOUND USING THE WIGNER-VILLE DISTRIBUTION

This section aims at providing an original visualization and detection of impulses in noise in the time and frequency domains. A time-frequency energy distribution which is particularly indicated for this purpose is the Wigner-Ville distribution (*WVD*), which is defined in[6] as:

$$W_x(t,\upsilon) = \int_{-\infty}^{+\infty} x(t+\tau/2) \cdot x^*(x-\tau/2) \cdot e^{-2j\pi\upsilon\tau} d\tau \qquad (1)$$

This distribution satisfies a large number of desirable mathematical properties. In particular, the *WVD* is always real-valued, on the one hand, and it preserves time and frequency shifts, on the other hand. Unlike spectrogram, the *WVD* has non-zero interference terms regardless of the time-frequency distance between two signal terms. These interference terms are troublesome since they may overlap with auto-terms (signal terms) and thus make it difficult to visually interpret the *WVD* image. This drawback is solved by using a smoothing function, as presented in[7].

Figure 2 presents three examples of the direct sound of experimental RIRs measured with a balloon burst, a gun shot and a cracker impulse, respectively. As a first observation, this visualization highlights the longest duration of the balloon burst compared to the gun shot and to the cracker. Moreover, the time signal of the balloon burst presents large and long oscillations, making difficult to identify the temporal boundaries of the impulse precisely. Frequency components of the balloon burst are evolving with time, which is somehow related to the tearing of the balloon during its burst. The cracker impulse is short and relatively well localized in the frequency domain. However, its spectral content is impoverished compared to the gun shot. Indeed, even if this latest is not as brief as the cracker impulse, it covers a larger frequency range. The three sources studied here are all mainly low frequency.

3 METHODS FOR ESTIMATING THE DIRECT SOUND DURATION

3.1 Estimating the direct sound duration with STFT

When observing the temporal structure of a RIR, it is noticeable that the occurrence of the direct sound is accompanied by a sudden increase of amplitude, followed by a decay. The method proposed here aims at analyzing mean variations of spectra from the beginning (the onset) to the end (the offset) of the direct sound, windowing the signal. Temporal boundaries are estimated using an envelope follower, which can easily be constructed by low-pass filtering the local energy[8]. The onset time is estimated as follows. The maximum of $\hat{E}(n)$ is detected, and the signal analyzed from its beginning to its maximum (t_{max}). The ratio of two successive windows is calculated; the index (T_{On}) is found when the ratio is maximum (Eq. 2). While the offset is estimated as follows. The signal is analyzed from its maximum up to 5ms from the onset time, assuming a duration inferior or equal to 5ms. The ratio of two successive windows is calculated; the index (T_{Off}) is found when the ratio is minimum (Eq. 2). This can be written as:

$$\tilde{E}(n) = \sum_k |X(n.h,k)| \tag{2}$$

$$T_{On} = h \times \arg\max_n \left[\tilde{E}(n+1) / \tilde{E}(n) \right] \tag{3}$$

$$T_{Off} = h \times \arg\min_n \left[\tilde{E}(n+1) / \tilde{E}(n) \right], n > t_{\max} \tag{4}$$

where $X(n,k)$ is the Short Time Fourier Transform (*STFT*) of the signal $x(n)$, and h the time step between two windows.
In practice, we choose a window analysis width (that is the temporal resolution) of h=0.1ms.

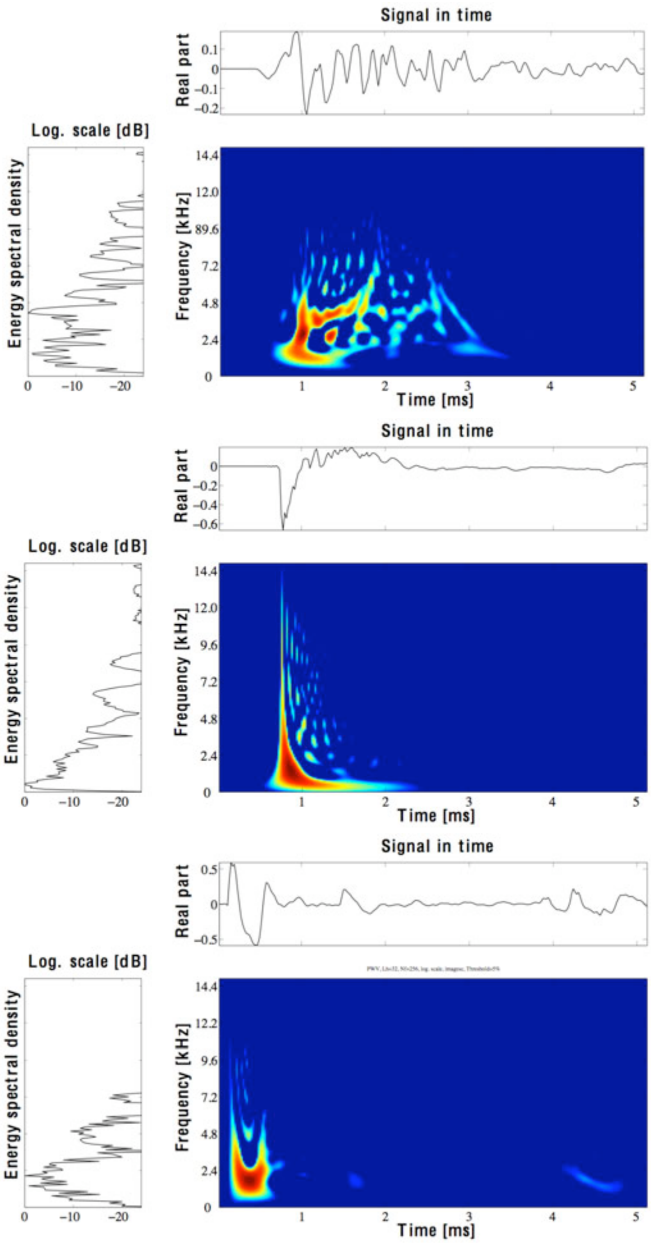

Figure 2: Smoothed Wigner-Ville distributions of the direct sound of three experimental RIRs measured with: -a balloon bursts (top); -a gun shot (middle); - a cracker (bottom) (at 48kHz).

3.2 Matching Pursuit applied to RIR

Since details and issues of using MP on RIRs has been developed in[9], this section briefly presents theoretical and practical aspects of MP, and focuses on the method for estimating the direct sound. A RIR can be seen as a linear set of occurrences of the direct sound translated in time, filtered by reflections on the surfaces of the hall. For this latter reason, it is believed that MP can help for understanding more deeply the architecture of a RIR, since this algorithm, introduced by[4], provides information, which can be seen as maxima of correlation (Eq. 5) between two signals[10]: the RIR and a dictionary composed of the direct sound (the mother atom) delayed in time. The temporal boundaries of the direct sound are estimated by defining precisely the mother atom, that is, by learning the dictionary (Section 3.2.2).

3.2.1 Theoretical reviews

In theory, any signal x can be perfectly decomposed in a linear set of atoms for an infinity of iterations.

Algorithm of Matching Pursuit

$$Input : x \in \mathbb{R}^N, D = \left\{\phi_\gamma, \gamma\right\}$$

$$Output : \gamma_{opt}^k, \alpha^k, k = 1...(n-1)$$

$$n = 0$$

$$R^0 x \leftarrow x$$

Repeat until the stopping criterium is reached:

$$\gamma_{opt}^{(n)} \leftarrow \arg\max_{\gamma\in\Gamma} \left|\left\langle R^n x \cdot \phi_\gamma \right\rangle\right| \tag{5}$$

$$\alpha^{(n)} \leftarrow \left|\left\langle R^n x \cdot \phi_{\gamma_{opt}^{(n)}} \right\rangle\right| \tag{6}$$

$$R^{n+1}x \leftarrow R^n x - \alpha^{(n)}\phi_{\gamma_{opt}^{(n)}} \tag{7}$$

$$n \leftarrow n + 1 \tag{8}$$

$$x^{(n)} = \sum_{k=0}^{n-1} \alpha^{(k)}\phi_{\gamma_{opt}^{(n)}} \tag{9}$$

with x being the original RIR, R the residual, $x^{(n)}$ the synthesized signal, < · > the scalar product, and ϕ_γ the dictionary of atoms γ.

In practice, this number of iterations must be finite and a stopping criterium has to be set. The authors propose to use the signal/noise ratio (*SRR*) in dB of x over the residual (R).

The quality of the decomposition of x in atoms depends on the value of *SRR*. On the one hand, for a *SRR* too low, the residual has a too high energy level and the rebuilt signal $x^{(n)}$ is an impoverished approximation of x. On the other hand, a *SRR* too high leads to a high number of iterations, that becomes useless over a certain value, since $x^{(n)}$ is almost identical to x. Comparing the acoustical indices, used in Room Acoustics[11], calculated on to those calculated on $x^{(n)}$ for different values of a *SRR* allows to set the stopping criterium on a rational ground. This is achieved by running MP on a RIR for which the determination of the direct sound is obvious. The acoustical indices used here are the reverberation times at 20dB (RT_{20}), 30dB (RT_{30}), the Early Decay Time at 10dB (EDT_{10}), and the Central Time (T_C). According to[9], a convenient *SRR* would be 20dB, since variations of acoustical indices lie under 5%. Figure 3 shows a RIR, its direct sound, the linear set of coefficients and the residual.

3.2.2 Detection of the direct sound: learning the dictionary of atoms ϕ_γ

It is extremely important to know the exact temporal boundaries of the direct sound, since the number of iterations of MP is the lower for a more precise atom definition, and bigger for a totally approximative atom. A fine estimation is possible using a method of grid search, as follows.

As in Section 3.1, we assume the impulse duration to be inferior to 5ms, defining the mother atom γ. The time index (t_0) of the maximum of the mother atom is detected. A dictionary of atoms ϕ_γ is constituted of atoms with temporal boundaries that are varying (with a step of 0.1ms) from 0ms to t_0 and from t_0 to 5.00ms for the first and last indices respectively. For each couple of [first index : last index], MP is ran. The temporal boundaries that are thought to be the best correspond to the lowest number of iterations. Impulses durations are estimated running MP onto experimental RIRs, with a stopping criteria of *SRR* = 20dB to reach.

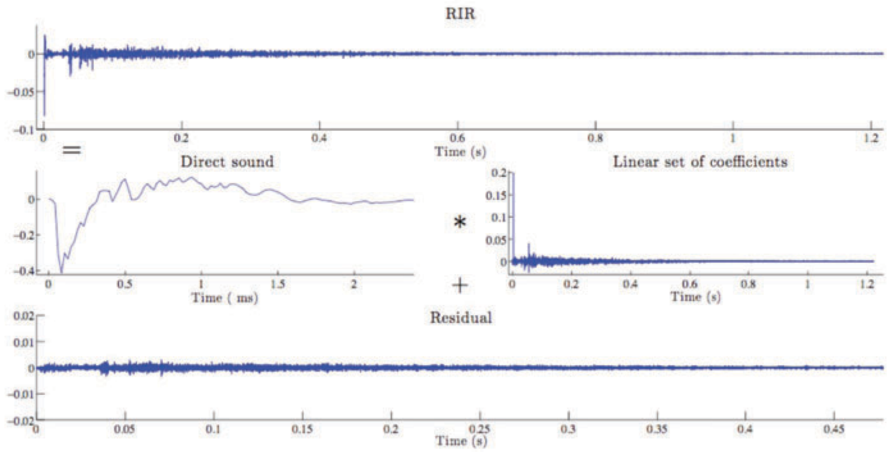

Figure 3: Matching Pursuit ran on an experimental RIR (SRR=20dB) (experimental RIR, direct sound, linear set of coefficients ($\gamma^{(n)}_{opt}$) and residual).

4 REFERENCE IMPULSES

For the three types of sound sources, 20 measurements of their impulse is carried out in a damped room with a 1/2" omnidirectionnal microphone B&K (type 4191), a amplifier B&K Nexus 2692C (calibrated at 316\mu V/Pa) and a numerical recorder NAGRA V. Recordings were made at a rate of 48kHz and for 24bits. The receiver is placed normally to the sound source, that is: pointing towards the piercing point of the balloon, perpendicular to the barrel of the gun, and facing the cracker. Temporal boundaries are estimated at hand for each impulse. Figure 4 shows the different durations measured. Variations of results are a consequence of the poor reproducibility of these sources.

5 RESULTS

The two detection methods are tested over 100 audio wavfiles. These RIRs have been measured in Salle Pleyel in Paris[5], according to ISO3382[11], with balloon bursts, gun shots and cracker impulses. The sound sources were of the same type than those used in the damped room. For each RIR, the onset and offset times are estimated by each method (figures 5-6), and compared to the references (figure 4). Statistics of the results are presented as interquantiles. Boxes have lines at the lower quartile, median, and upper quartile values. Whiskers show the extent of the rest of the data, while outliers are data with values beyond the ends of the whiskers.

On the one hand, the two methods return results that are in agreement with the reference durations: balloon bursts last longer than gun shots and cracker impulses. Durations of balloon bursts estimated present larger variations than for the two other kinds of sources. On the other hand, time durations estimated by MP are closer (in terms of average and spreading) to the references than those estimated using the *STFT*. This point is further discussed in Section 6. The good agreement between results from MP and references is a hint for choosing this correlation-based method to estimate the direct sound duration.

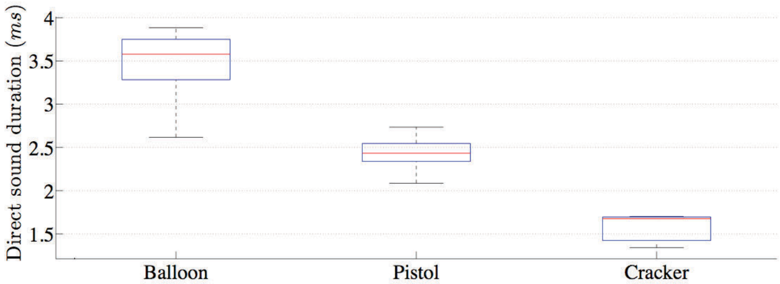

Figure 4: Impulse duration of balloon bursts, gun shots and crackers measured in a damped room.

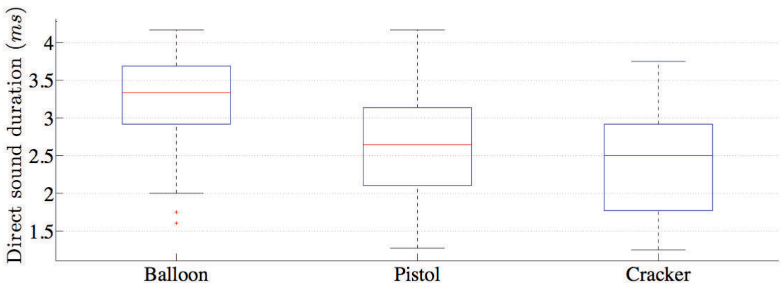

Figure 5: Impulse duration calculated on RIRs using the STFT method.

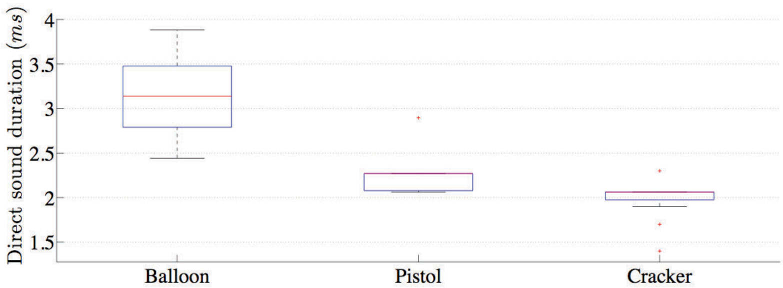

Figure 6: Impulse duration calculated on RIRs using Matching Pursuit (SRR=20dB).

6 DISCUSSION

Large variations of results are not only related to the robustness of the presented estimators, but also to the lack of reproducibility of the considered sound sources. According to results, the shorter and more reproducible source is the cracker (since variations are the smallest), followed by the gun shot and then by the balloon burst.

Large spreading of results returned by the *STFT* method are a consequence of the drawback of this method: that is, its resolution. Indeed, the time resolution is proportional to the effective duration of the analysis windows. Similarly, the effective frequency resolution is proportional to the effective band-width of the analysis window. Consequently, for the *STFT* a *trade-off* between time and frequency resolution has to be made: on the one hand, a good time resolution requires a short window analysis; on the other hand, a good frequency resolution requires a long window. This limitation leads to choose another method.
It is believed that the estimation of the impulse duration using MP provides more accurate and realistic results than the *STFT* method. First, results are closer to the references, and second, MP does not undergo any frequential or temporal constraint. However, MP seems to underevaluate impulse durations, except for the crackers. This can be explained by the fact that reference impulses duration have been determined by hand, and the determination of the offset mixed into the background noise leads to variations, as mentioned in Introduction. Furthermore, differences between experimental conditions of measurements in the damped room and in Salle Pleyel also explain these variations of estimation.

The three kind of sources studied in this article are certainly not recommended for accurate RIRs measurements, because of the lack of reproducibility in the time and frequency domains, yet they are often used by practitioners. Temporal and frequential properties of balloon bursts depend on the volume of the balloon, on the homogeneity of the latex, and overall on the location of the piercing point relatively to the microphone. Guns shots are also poorly reproducible for two main reasons. First, during the explosion of the bullet, the sound wave not only escapes the barrel, but also the cylinder, since it is in fact a revolver. Second, the bullets used with gun are not exactly filled with the same amount of powder, according to manufacturers. Among the three sources studied, cracker impulse is the shortest, and the more reproducible. However, the drawback of this source is its impoverished spectra, mainly a low frequency spectrum. As for the gun, variations of durations of the cracker impulse can be explained by its manufactory issues.

All these parameters have a strong influence on the directivity, on the impulse duration and on frequency components of the sources.

7 CONCLUSIONS

Three different types of frequently used sound sources are studied: balloon bursts, gun shots and cracker impulses. Their relative duration are estimated using a method based on the *STFT* and Matching Pursuit. Results are compared to some reference impulses measured in a damped room. Matching Pursuit returns the best results. This method is believed to be more accurate to determine the temporal boundaries of the direct sound of an experimental RIR than a method based on *STFT*. High disparities in results, specially for the balloon bursts and the guns shots, underlies their poor reproducibility in time. Cracker impulses are the sharpest, followed by gun shots and then by balloon bursts. Results confirm that these sound sources must be used with knowledge of their physical limitations.

Finally, further studies should investigate on the one hand, the effect of window analysis width used in the two presented methods, and on the other hand, the directionality of such sound sources.

8 ACKNOWLEDGMENTS

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

9 REFERENCES

1. D. Griesinger, Beyond MLS- Occupied hall measurement with fft techniques, URL http://world.std.com/~griengr/sweep.pdf, 2004.
2. A. Nash, On the acoustical characteristics of a balloon, in Proc. Int. Symp. Room Acoustics, Sevilla, Sept. 2007.
3. P. Fausti and A. Farina, Acoustic measurements in opera houses: Comparison between different techniques and equipment, Journal of Sound and Vibration, vol. 232, no. 1, pp.213-229, June 2000.
4. S. Mallat and Z. Zhang, Matching Pursuit with time-frequency dictionaries, IEEE Trans. Signal Process., vol. 40, no. 12, pp.3397-3415, December 1993.
5. G. Defrance, J-D. Polack and B-FG. Katz, Measurements in the new Salle Pleyel, in Proc. Int. Symp. Room Acsoutics, Sevilla, September 2007.
6. P. Flandrin, A time-frequency formulation of optimum detection, in IEEE Trans. Acoust. Speech Sig. Proc., vol. 36, pp.1377-1384, 1988.
7. D.J. Hermes, Vowel-onset detection, J. Acoust. Soc. Am., vol. 87, no. 2, pp. 866-873, 1990.
8. G. Defrance, L. Daudet and J-D Polack, Finding the onset of a room impulse response: straightforward?, J. Acoust. Soc. Am. (Express Letter), vol. 124, no. 4, October 2008 (to be published).
9. G. Defrance, L. Daudet and J-D Polack, Detecting arrivals within Room Impulse Response using Matching Pursuit, in Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland, September 1-4, 2008.
10. S. Krstulovic and R. Gribonval, Mptk: Matching Pursuit made tractable, ICASSP'06, Toulouse, France, May 2006.
11. ISO3382, Acoustic-measurements of the reverberation time with reference to other acoustical parameters, 1997.