How precise should we reconstruct a room impulse response? Physical and perceptual points of view

Guillaume Defrance, Jean-Dominique Polack

10ème **C**ongrès **F**rançais d'**A**coustique

Lyon, 12-16 Avril 2010

How precise should we reconstruct a room impulse response? Physical and perceptual points of view.

Guillaume Defrance[1], Jean-Dominique Polack[2]

[1] Department of Architecture, University of Sheffield, Conduit Road, S10 2TN, Sheffield, United Kingdom, g.defrance@sheffield.ac.uk

[2] UPMC Univ. Paris 06, IJLRDA LAM, CNRS UMR 7190, 11 rue de Lourmel, 75015 Paris, France, polack@lam.jussieu.fr

In a recent publication, we show that the algorithm of Matching Pusuit is applicable to estimating arrivals within room impulse responses. The study of the time distribution of arrivals allows one to define a time beyond which arrivals are statistically distributed. We call that time the cross-over time. However, these estimates are extremely sensitive to the stopping criterion of Matching Pursuit - which is closely linked to the precision of the approximation. When the precision is too low, we arrive at an unrealistic number of arrivals. On the other hand, when the precision is too high, the model of room impulse responses gives poor estimates. The best stopping criterion can be determined based on a comparison of room acoustical indices or using listening tests. In this paper, we use both approaches in order to estimate arrivals and the cross-over time of experimental room impulse responses of a concert hall from their Matching Pursuit decomposition. We show how perceptually determining the stopping criterion leads to a perceptual estimation of the cross-over time. We are thus able to make comparisons with the heuristic formula of the mixing time proposed by Polack, which is related to the volume of the hall.

1 Introduction

One measures room impulse responses (RIRs) in order to study and document the acoustics of a room. In practice, a *source*, e.g., spark gun, emits a wideband signal into the room, and *receivers*, e.g., microphones, measure local changes in pressure at specific locations. We can model a RIR, in the limit of high frequencies by modeling sound as rays that leave a source at the speed of sound c, and undergo reflections at boundaries before arriving at each receiver. An *arrival* is a sound ray emitted by the source that has undergone at least one reflection during its journey to the position of the receiver. What we wish to do is accurately detect arrivals in an RIR in order gain knowledge of the room, e.g., its volume. Such a relationship is embodied in the expected number arrivals received at t seconds after excitation [1]

$$\mu_A(t) = \frac{4\pi c^3}{3V} t^3 \tag{1}$$

where V is the volume of the room in cubic meters.

Given a measured RIR, we wish to accurately find the times and amplitudes of the arrivals. This problem has been addressed by a few methods within the discipline of room acoustics. One approach uses an adaptive thresholding technique [2], but this requires empirically testing a range of variables to make the detection algorithm give reasonable results. Furthermore, this approach essentially detects local peaks in the RIR, and then equates those to arrivals. A different approach uses greedy sparse approximation to first decompose the RIR as a linear combination of the measured direct sound, and then to detect arrivals in a domain more sparse than the original RIR [3, 4]. However, this approach is very sensitive to the parameters of the decomposition algorithm (Matching Pursuit [5]), especially to the stopping criterion of the decomposition. In a previous work, we have proposed to determine this criterion by studying variations of usual room acoustical indices [3, 6]. In this paper, we investigate another approach based on listening tests.

In the following, we briefly present Matching Pursuit (MP) and recall some results that we have obtained in previous work. Estimation of arrivals permits one to detect the cross-over time, which is the transition time between early reflections and late reverberation [4, 7]. We also recall how the cross-over time (also called the mixing time in the room acoustics literature) is related to the stopping criterion and how it can be measured from the set of arrivals estimated by MP. We summerize our previous approach of the determination of the stopping criterion of MP. We then use listening tests to set a perceptual meaningful stopping criterion of MP, i.e., we look for estimating the degree of approximation at which there are not any perceptual differences between high and low orders of decomposition. We finally discuss our results obtained with both approaches and compare the models of RIRs.

2 Matching Pursuit applied to RIRs

In this Section, we present the alogrithm of MP and show how the stopping criterion of MP can be set based on variations of room acoustical indices. We also briefly recall how the cross-over time can be estimated from the linear set of estimated arrivals and how it is dependent

on the stopping criterion of the pursuit.

2.1 Algorithm of MP

In a previous work [3, 4], we show that it is possible to estimate arrivals within RIRs, based on the assumption that a RIR can be approximated as a linear combination of the direct sound delayed in time and filtered by the boundaries of the room. Hence, a technique based on finding the times at which the source impulse (i.e., the direct sound) is highly correlated to the signal of the RIR is well indicated for estimating arrivals (times of occurrence and amplitudes).

Matching Pursuit works as follows:

1. Initialization: $m = 0$, $x_m = x_0 = x$
2. Computation the correlations between the signal x_m and every atom γ of a dictionary ϕ, using inner products:
$$\forall \gamma \in \phi \quad : \quad CORR(x_m, \gamma) = |\langle x_m, \gamma \rangle| \qquad (2)$$
The dictionary ϕ is a set of atoms γ, of the same length than x, constituted by the direct sound and translated in time, by step of one sample.
3. Search the most correlated atom, by searching for the maximum inner product:
$$\tilde{\gamma}_m \quad = \quad argmax(CORR(x_m, \gamma))_{\gamma \in \phi} \qquad (3)$$
4. Substracting the corresponding weighted atom $\alpha_m \tilde{\gamma}_m$ from the signal x_m:
$$x_{m+1} \quad = \quad x_m - \alpha_m \tilde{\gamma}_m \qquad (4)$$
$$x_R^{(m)} \quad = \quad \sum_{k \leq m} \alpha_k . \tilde{\gamma}_k \qquad (5)$$
where $\alpha_m = \langle x_m, \tilde{\gamma}_m \rangle$;
5. - stops if the desired level of accuracy is reached: $R = x_{m+1}$.
   - otherwise, re-iterate the pursuit to step 2: $m \leftarrow m + 1$.

where x is the RIR, R the residual, γ the atom (here, the direct sound), ϕ the dictionary of atoms γ, and $x_R^{(m)}$ the reconstructed signal.

2.2 Previous results

In theory, any signal x can be perfectly decomposed in a set of atoms for an infinity of iterations. In practice, this number must be finite and a stopping criterion has to be set. In [3, 4], the authors propose to use the signal/residual ratio (SRR) in dB of the norm L2 of x over the norm L2 of the residual (R) defined as:

$$SRR \quad = \quad 20 log_{10} \left(\frac{||x||_2}{||R||_2} \right). \qquad (6)$$

We also show that for estimating arrivals at any time, that is, with the same probability, it is necessary to compensate for the energy decay of the signal applying an inverse exponential based on the reverberation time and the mean absorption of the room [4]. In the following, this compensation is applied to all RIRs that are decomposed by MP.

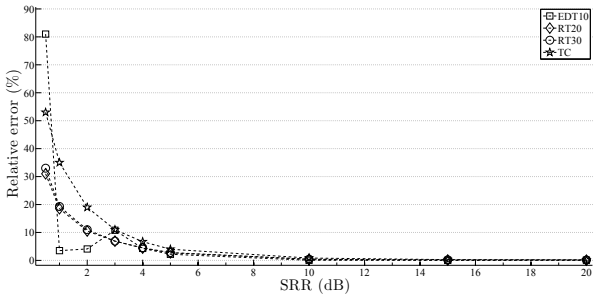

Figure 1: Variations in % of some room acoustical indices (EDT_{10}, RT_{20}, RT_{30}, T_C) versus the SRR in dB (with compensation of the energy decay).

2.2.1 Determining the stopping criterion

The number of estimated arrivals depends on the value of the SRR (6). Indeed, if the SRR is low, the number of estimated arrivals is low. On the other hand, to a large SRR corresponds a large number of arrivals. The problem to address is thus to determine the best value of SRR to use.

Defrance *et al.* [4] investigate an approach based on variations of room acoustical indices. For different values of the SRR, they calculate signed variations of room acoustical indices between reconstructed RIRs and original ones. According to [8], acceptable variations of acoustical indices are 5% and below. This previous study indicates that, when the energy decay of the RIR is compensated, the SRR should be equal or greater than 5 dB.

2.2.2 Estimating the cross-over time

What we call the cross-over time is better known as the mixing time in the literature of room acoustics. In this domain, this particular time defines the transition time between early reflections and late reverberation. However, mixing is by definition a property of some dynamical systems that are ergodic [9]. In [7], we investigate experimental estimation of the mixing time. We show that the mixing character of large halls is not proved yet. Therefore, we propose the term of *cross-over time*, in reference to the cross-over frequency proposed by Schroeder [10].

In [11], Polack proposes a heuristic formulation of the cross-over time, based on perception. It is reached when 10 arrivals occur in a window of 24 ms [1]. According to Polack, a possible formulation of the cross-over time could be:

$$\Delta\mu_A(t) \quad = \quad \frac{4\pi c^3 \Delta t}{3V} t^2 \qquad (7)$$
$$t \quad = \quad \sqrt{\frac{3V}{4\pi c^3} \frac{\Delta\mu_A}{\Delta t}}, \qquad (8)$$

where $\Delta\mu_A = 10$ arrivals and $\Delta t = 24$ ms. We can thus approximate (8) by:

$$T \approx \sqrt{V}, \qquad (9)$$

where T is the cross-over time expressed in ms.

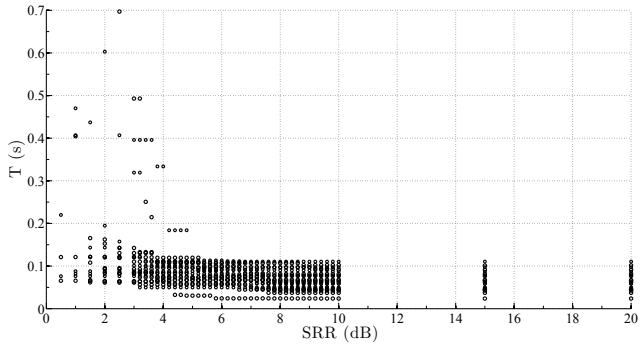

Figure 2: Cross-over times estimated on 21 experimental RIRs using different values of SRR (dB) (with compensation of energy decay). RIRs are measured in Salle Pleyel, using an omnidirectional microphone [14]. Note that at low SRR, some cross-over times are not always found.

Using MP, the estimation of the cross-over time, which is counted from the time of emission of the impulse by the source, can be achieved by looking for [4]:

$$t = argmin(t_{i+1} - t_i \leq \tilde{d}_{source}), \qquad (10)$$

where t_i is the time of occurrence of the ith estimated arrival, and $\tilde{d}_{source}$ is the equivalent duration of the source impulse [12, 13].

Obviously, when SRR is large, the number of arrivals is large too, and thus, the cross-over time is statistically low. On the other hand, if the SRR is low, the number of estimated arrivals is low, and thus, the cross-over time is either large, or even does not exist (Fig. 2). See Ref. [4] for more details.

3 A novel approach for estimating the SRR

In the following, we investigate a different approach for determining the SRR. We use listening tests in order to set the order of decomposition of MP such that there are no perceptual differences between a reference RIR and the approximated version returned by MP.

3.1 Experimental set up

In Salle Pleyel, we have measured 21 RIRs using a B-format microphone [15] and spark guns, in the same experimental configurations as in [14].

In the following, we only use three different RIRs, at thee different receiver positions:

- The microphone is placed in front of the stage, on the floor. The source is seen from the receiver position ($d_{SR} = 8$ m)[1].
- The microphone is placed on the first balcony in front of the stage. The source is seen from the receiver position ($d_{SR} = 30$ m).
- The microphone is placed on the second balcony (right side of the stage). The source is not seen from the receiver position ($d_{SR} = 20$ m).

Each of the four channels of the B-format recording is decomposed using MP for several values of SRR [0.5:20] dB. Each approximated channel is then convolved by three different pieces of music [16] recorded in an anechoic chamber:

1. W.A. Mozart, The marriage of Figaro (first 80 s);
2. Johann and Joseph Strauss, Pizzicata Polka, (first 80 s);
3. J. Brahms, first mouvement, Symphony # 4, bars 386-407 (first 35 s).

Therefore, nine pieces of music are played and sent on an Ambisonics sound system in a damped room, that comprises 12 loudspeakers and one subwoofer [17].

3.2 Listening test

Subjects that run the listening test have to answer one question from a user interface developped under MatLab (Fig. 3). The task is the following: the subject faces three buttons, each one corresponding to a sound. One of these buttons is called "Reference"; the reference sound uses a RIR decomposed with $SRR = 20$ dB, that is, a high order of approximation since variations of room acoustical indices are below 0.1% (Fig. 1). The two other buttons are called "Sound A" and "Sound B". If "Sound A" (or alternatively "Sound B") is the reference sound, then "Sound B" uses the same RIR, but approximated with a different SRR (SRR= $[0.5 : 20]$ dB). Further, the reference sound is always either "Sound A" or "Sound B", but never both. The subject is asked to identify the reference sound between "Sound A" and "B". Each time the subject is able to make the difference between the two sounds, he/she is asked to confirm his/her answer (Fig. 4). When the answer is the right one, the test uses a higher SRR (e.g., SRR goes from 0.5 to 1.0 dB). As long as the subject gives the good answer, the test becomes more difficult.When the subject is not able to make the difference between the two sounds, he/she is asked to answer the question using an inferior SRR (e.g., SRR goes from 9 to 8 dB). The test ends when a total of three wrong answers has been given. In practice, the test starts using a SRR equal to 0.5 dB.

The goal of this procedure is to estimate the value of SRR for which subjects are not able to make the difference between two orders of decomposition. Therefore, final answers of the subject vary around a limit value of the SRR (SRR_{lim}). For this particular SRR, one can assume that the two sounds are perceptively identical. This test is derived from the *Adaptive staircase technique* proposed by Levitt [18] and studied by Kollmeier *et al.* [19].

3.3 Results

Twenty six persons have run the test. Table 1 presents the results (averaged SRR_{lim} and the standard deviation) as functions of the type of music played. We first

[1] d_{SR} stands for the source/receiver distance.

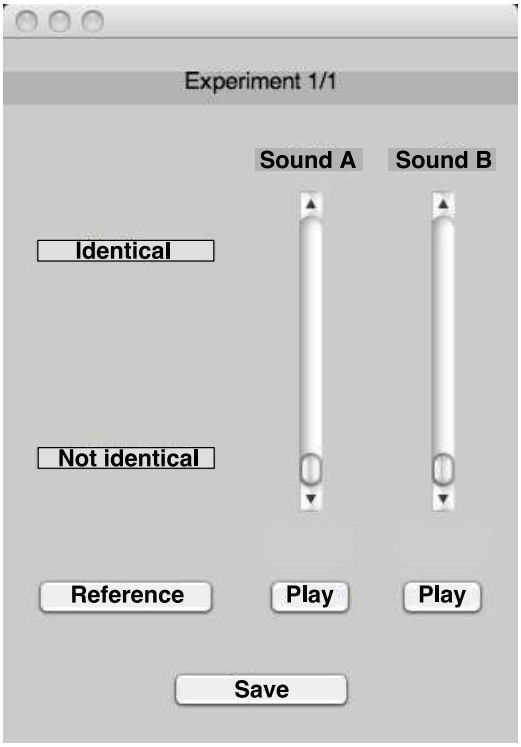

Figure 3: User interface used for listening tests. Subjects are asked to find the Reference sound within "Sound A" and "Sound B".

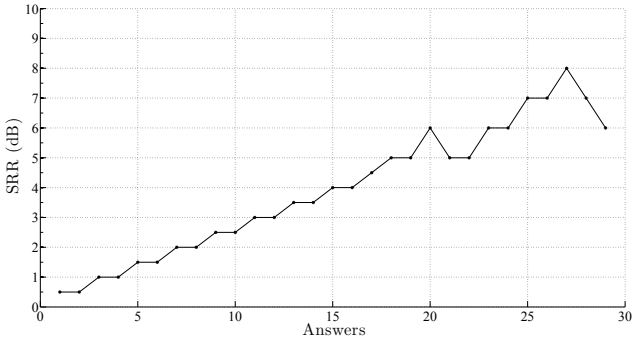

Figure 4: Example of the results of the listening test.

notice that the $\overline{SRR_{lim}}$ depends on the type of piece of music that is played. For instance, the $\overline{SRR_{lim}}$ is greater when percussive instruments are played than when it is only an ensemble of violins. Therefore, one can assume that, from a perceptual viewpoint, the "best" stopping criterion is the greatest ($\overline{SRR_{lim}} = 8$ dB). Note also the large standard deviations when pieces of music are slightly percussive (pieces 1 and 2, Table 1). The averaged $\overline{SRR_{lim}}$ of this test is approximately 6 dB, that is, approximately the estimated SRR found using room acoustical indices (Section 2.2.1).

4 Discussion

In [4] the cross-over time is found to be a function of the source/receiver distance. Using values of SRR found with the listening test (SRR = [5,6,8] dB), we can compute relationships between cross-over times and the source/receiver distances (Table 2). For SRR=8dB, the system is almot immediatly diffuse, that is, the transition to late reverberation occurs just after the direct sound has reached the receiver positions. The relationship obtained with SRR=6 dB is close to the one obtained in our previous work [4] (Fig. 5 and Table 2). It is difficult to conclude to the best SRR, since this study only concerns one concert hall and a few participants. However, results are quite consistent together if one considers the mean SRR estimated with listening tests and the one found in Section 2.2.1.

SRR (dB)	Cross-over time and d_{SR}
5	$Tm = 0.0026 \times d_{SR} + 0.026$ ($r = 0.82$)
6	$Tm = 0.0027 \times d_{SR} + 0.017$ ($r = 0.90$)
8	$Tm = 0.0030 \times d_{SR}$ ($r = 0.99$)

Table 2: Relationships between cross-over time and source/receiver distance (d_{SR}) for different SRRs in Salle Pleyel. Note that r is the correlation coefficient of the linear relationship.

SRR (dB)	RIR–1	RIR–2	RIR–3
5	**9**	38	32
6	**11**	32	25
7	17	57	61
8	25	74	75

Table 3: Mean numbers of estimated arrivals within a window of 24 ms at three given locations in Salle Pleyel. Note that the number of arrivals is counted for times earlier than the cross-over time. In bold font are the numbers of arrivals that correspond to the mean number predicted by Cremer and that agrees with Polack's proposition of the cross-over time [1, 11].

As said above, the order of decomposition of MP specifies the number of arrivals that one estimates within one given RIR. Table 2 shows that the greater the SRR, the larger the number of estimated arrivals and the earlier the cross-over time, thus the more diffuse the concert hall. We clearly see here the relationship that exists between the perceptual estimations of the SRR and the cross-over time.

One original aspect of using MP for estimating arrivals is that we are now able to compare our estimation of cross-over time to the formulation proposed by Polack (9). Table 3 presents the relationship between SRR and the mean number of estimated arrivals[2] in a windows of 24 ms. The number of estimated arrivals is not in total agreement with the one predicted by Cremer [1]. One obvious reason for that is that measurements are always carried out in particular configurations (balconies, ceiling, seats, etc.). However, one may notice that in the near field, the number of estimated arrivals agrees with Polack's proposition (5≤ SRR<8 dB (RIR–1), Table 3).

[2]Note that the number of arrivals is counted for times earlier than the cross-over times.

Piece of music #	Specifications	$\overline{SRR_{lim}}$ (dB)	σ (%)
1	Grand orchestra / mainly violins / not percussive	5.0	34
2	Harp / percussive	8.0	17
3	Small orchestra / mainly violins / quite percussive	7.0	22

Table 1: Specifications of each piece of music used in the test. To each piece of music is associated the averaged limit value of the SRR ($\overline{SRR_{lim}}$) and the standard deviation on the 26 subjects.

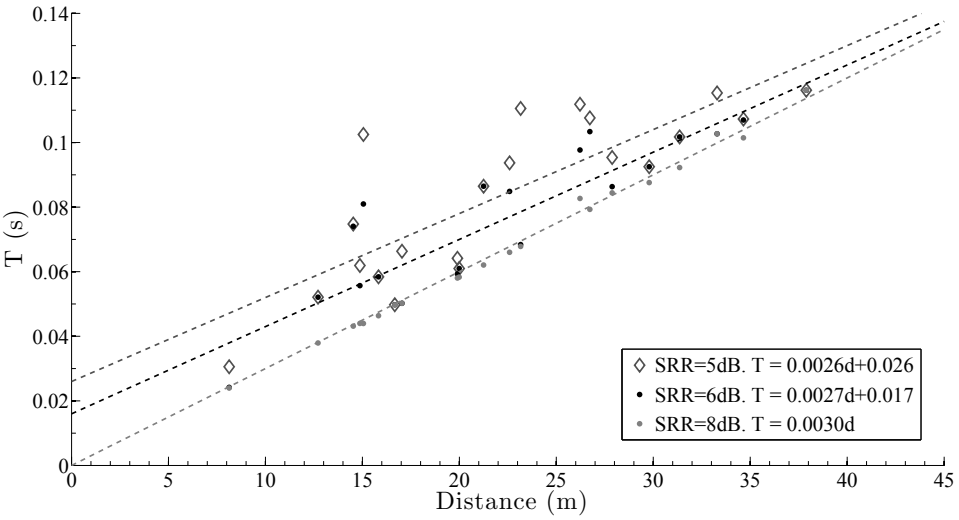

Figure 5: Cross-over times (T) estimated on 21 experimental RIRs using SRR [5,6,8]dB (with compensation of energy decay). The abscissa is the source/receiver distance. Note that as SRR increases, cross-over times occur just after the direct sound has reached the receiver position.

5 Conclusion

In this study, we use an algorithm based on maxima of correlation for estimating arrivals within RIRs. We show that the number of estimated arrivals, as well as the cross-over time are some functions of one important parameter of this algorithm. We investigate the use of listening tests to determining the best perceptual value of this criterion. We show that the mean value estimated is close to the one proposed in our previous work. Finally, we compare the number of arrivals estimated using this approach to the heuristic formulation of the cross-over time proposed by one of the authors and show that our results are in agreement in the near field. In a future work, we wish to carry out the same listening tests with a larger number of subjects and RIRs from several other concert halls.

References

[1] L. Cremer, H.A. Muller, and T.J. Schultz, *Principles and Applications of Room Acoustics*, vol. 1, Applied Science Publishers Ltd, London and New York, 1982.

[2] M. Kuster, "Reliability of estimating the room volume from a single room impulse response," *J. Acoust. Soc. of Am.*, vol. 124, no. 2, pp. 982–993, Aug. 2008.

[3] 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.

[4] G. Defrance, L. Daudet, and J-D. Polack, "Using Matching Pursuit for estimating mixing time within Room Impulse Responses," *Acta Acustica united with Acustica*, vol. 95, no. 6, pp. 1082–1092, December 2009.

[5] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionnaries," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 3397 3415, december 1993.

[6] Leo Beranek, *Concert and Opera Halls How They Sound.*, Acoustical Society of America, 1996.

[7] G. Defrance, *Characterization of mixing within room impulse responses. Application to the experimental estimation of the mixing time.*, Ph.D. thesis, University Pierre et Marie Curie, November 2009.

[8] X. Meynial, Polack. J-D, and G. Dodd, "Comparison between full-scale and 1:50 scale model measurements in Théâtre Municipal, Le Mans," *Acta Acustica*, vol. 1, pp. 199–212, 1993.

[9] W. B. Joyce, "Sabine's reverberation time and ergodic auditoriums," *J. Acoust. Soc. Am.*, vol. 58, no. 3, pp. 643–655, 1975.

[10] M.R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J. Acoust. Eng. Soc.*, vol. 35, no. 5, pp. 307–316, 1987.

[11] J-D. Polack, *La transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Thèse de doctorat d'Etat, Université du Maine, Le Mans, France, 1988.

[12] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurements Procedures*, New York: Wiley, 1971.

[13] W. D. Stanley and Peterson S. J., "Equivalent Statistical Bandwidths of Conventional Low-Pass Filters," *IEEE Transactions on Communications*, vol. 27, no. 10, pp. 1633–1634, October 1979.

[14] G. Defrance, J-D. Polack, and B-FG. Katz, "Measurements in the new Salle Pleyel," in *Proc. Int. Symp. Room Ac.*, Sevilla, September 2007.

[15] M. Gerzon, "Recording Concert Hall Acoustics for Prosperity," *J. Acoust. Eng. Soc.*, vol. 23, no. 7, pp. 569, 1975.

[16] Enkoji Masahiko (conductor), "Anechoic orchestral music recording," Compact Disc, Pure Gold Collection, DENON 1987.

[17] C. Gustavino, *Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain*, Ph.D. thesis, Université Pierre et Marie Curie (Paris VI), 2003.

[18] H. Levitt, "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.*, vol. 49, no. 2, pp. 467–477, 1971.

[19] B. Kollmeier, R. H. Gilkey, and U. K. Sieben, "Adaptive staircase tecnhiques in psychoacoustics: A comparison of human data and a mathematical model," *J. Acoust. Soc. Am.*, vol. 83, no. 5, pp. 1852–1862, 1988.