REVERBERATION-TIME PREDICTION METHOD FOR ROOM IMPULSE RESPONSES SIMULATED WITH THE IMAGE-SOURCE MODEL

Eric A. Lehmann, Anders M. Johansson and Sven Nordholm

Western Australian Telecommunications Research Institute, Perth, Australia
{Eric.Lehmann,ajh,sven}@watri.org.au

ABSTRACT

The image-source method has become a ubiquitous tool in many fields of acoustics and signal processing. A technique was recently proposed to predict the energy decay (energy–time curve) in room impulse responses simulated using the image-source model. The present paper demonstrates how this technique can be efficiently used to determine the enclosure's absorption coefficients in order to achieve a desired reverberation level, even with a non-uniform distribution of the sound absorption in the room. As shown in this work, classical expressions for the prediction of an enclosure's reverberation time, such as Sabine and Eyring's formulae, do not provide accurate results when used in conjunction with the image method. The proposed approach hence ensures that the image-source model effectively generates impulse responses with a proper reverberation time, which is of particular importance, for instance, for the purpose of assessing the performance of audio signal processing algorithms operating in reverberant conditions.

1. INTRODUCTION

Together with other modeling techniques of room acoustics such as ray and beam tracing, the image-source model (ISM) [2] represents a principle of considerable importance for the engineering and acoustics research community. An important practical application of the image-source concept is related to the performance assessment of various signal processing algorithms operating in reverberant environments. To name but a few examples, the ISM approach has been used in order to validate algorithms for blind source separation [3], channel identification and equalization [4], acoustic source localization and tracking [5], speech enhancement [6], as well as various other array signal processing techniques [7, 8]. In such a context, the ISM is typically used to test the considered algorithm in order to determine its robustness against increasing levels of environmental reverberation. For the sake of a fair and consistent comparison, it is hence important to ensure that all algorithms are assessed using the same measure of reverberation across all simulations. However, and although not usually addressed in the literature, a significant issue during this assessment process is related to predicting the reverberation time (RT) in the room impulse responses (RIRs) generated with the ISM.

In the recent literature [5, 6], a common approach to this assessment process is to use well-established RT formulae, such as Sabine or Eyring's equations, to determine the enclosure's reflection coefficients from a desired reverberation level. The RIRs are then generated with the ISM technique, and the algorithm's performance results are finally displayed against the *desired* RT value. As shown in the present paper however, this approach is subject to large inaccuracies. Consequently, a significant risk exists that the performance results are ultimately presented for a reverberation level that does not correspond to the actual testing conditions.

An alternative approach chosen by several authors is to present the performance results versus the room's reflection coefficient itself [3, 4]. However, contrary to more intuitive parameters such as the reverberation time T_{60} for instance, this way of presenting the results does not provide much insight into the practical reverberation characteristics of the considered environment.

The present paper details how the energy-decay approximation method recently proposed in [1] can be implemented to provide an effective solution to the above problem. Since this approach is based on a direct estimation of the energy decay curve (EDC) resulting from the ISM computations, it provides an unequivocal correspondence between the enclosure's reflection coefficients and the resulting reverberation time, without requiring time-consuming ISM computations. The paper first presents a brief review of the ISM principle and the EDC approximation method of [1]. Section 3 then describes the proposed RT prediction technique based on this EDC approximation, as well as several classical RT formulae for comparison purposes. The proposed approach is then validated with experimental simulations in Section 4.

2. REVIEW OF BASIC CONCEPTS

2.1. Image-Source Model

1) Original Approach. Allen and Berkley's implementation of the image-source method [2] is a well-established algorithm for generating simulated RIRs in a given room. Assume that a Cartesian coordinate system (x, y, z) is defined in the considered enclosure, and let $\boldsymbol{p}_s = [x_s, y_s, z_s]^{\mathrm{T}}$ and $\boldsymbol{p}_r = [x_r, y_r, z_r]^{\mathrm{T}}$ denote the positions of a source and a receiver, respectively. Similarly, let $\boldsymbol{r} = [L_x, L_y, L_z]^{\mathrm{T}}$ represent the vector of room dimensions, with length L_x, width L_y and height L_z. The acoustical property of each surface in the enclosure can be characterized by means of a sound reflection coefficient β, related to the absorption coefficient α according to $\alpha = 1 - \beta^2$. The reflection coefficients for each surface are denoted $\beta_{x,i}$, $\beta_{y,i}$ and $\beta_{z,i}$, with $i \in \{1, 2\}$.

The RIR from the source to the receiver is determined by considering image sources on an infinite grid of mirrored rooms expanding in all three dimensions. With the triplets $\boldsymbol{u} = (u_x, u_y, u_z)$ and $\boldsymbol{v} = (v_x, v_y, v_z)$ used as parameters controlling the image-source indexing, the RIR $h(\cdot)$ follows as [2]

$$h(t) = \sum\nolimits_{\boldsymbol{u}=0}^{1} \sum\nolimits_{\boldsymbol{v}=-\infty}^{\infty} A(\boldsymbol{u}, \boldsymbol{v}) \cdot \delta\big(t - \tau(\boldsymbol{u}, \boldsymbol{v})\big), \quad (1)$$

This work was supported by National ICT Australia (NICTA). NICTA is funded through the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council.

$$\hat{h}_E(t) = \frac{\pi}{2c\,t\,\bar{r}} \cdot \begin{cases} \Big(\mathrm{Ei}\big(\log\big(\frac{B_z}{B_x}\big)\big) + \log\big(\log\big(\frac{B_z}{B_x}\big)\big) - \mathrm{Ei}\big(\log\big(\frac{B_z}{B_y}\big)\big) - \log\big(\log\big(\frac{B_z}{B_y}\big)\big)\Big) \cdot B_z \Big/ \log\big(\frac{B_y}{B_x}\big) & \text{if } B_x \neq B_y \neq B_z\,, \\ \Big(\mathrm{Ei}\big(\log\big(\frac{B_z}{B}\big)\big) + \log\big(\log\big(\frac{B_z}{B}\big)\big) + \gamma\Big) \cdot B_z \Big/ \log\big(\frac{B_z}{B}\big) & \text{if } B_z = B_y \neq B_x \triangleq B \text{ or } B_z = B_x \neq B_y \triangleq B\,, \\ (B - B_z) \big/ \log\big(\frac{B}{B_z}\big) & \text{if } B_z \neq B_x = B_y \triangleq B\,, \\ B & \text{if } B_x = B_y = B_z \triangleq B\,. \end{cases} \tag{5}$$

where $A(\boldsymbol{u},\boldsymbol{v})$ is the amplitude factor and $\tau(\boldsymbol{u},\boldsymbol{v})$ the time delay associated with the path to the receiver from the image source with indices $(\boldsymbol{u},\boldsymbol{v})$, and $\delta(\cdot)$ denotes the Dirac impulse function. For conciseness, the sum over $\boldsymbol{u}$ (respectively $\boldsymbol{v}$) in (1) is used to represent a triple sum over each of the triplet's internal indices.

2) Frequency-Domain Computations. When dealing with discrete-time signals, subsample time delays can be achieved by carrying out the ISM computations in the frequency domain. The RIR then results as the inverse Fourier transform of the frequency data:

$$h(t) = \mathcal{F}^{-1}\Big\{ \sum\nolimits_{\boldsymbol{u}=0}^{1} \sum\nolimits_{\boldsymbol{v}=-\infty}^{\infty} A(\boldsymbol{u},\boldsymbol{v}) \cdot \mathrm{e}^{-j\omega\,\tau(\boldsymbol{u},\boldsymbol{v})} \Big\}, \tag{2}$$

with $j = \sqrt{-1}$ and the frequency variable ω. For band-limited and time-sampled signals, this approach essentially represents the frequency-domain equivalent to Peterson's method [9].

3) Negative Reflection Coefficients. Given a specific absorption coefficient α, the corresponding reflection coefficient follows as $\beta = \pm\sqrt{1-\alpha}$. The original ISM implementation uses the positive β definition. It was however shown in [1] that the frequency-domain approach used in conjunction with positive β coefficients typically leads to RIRs presenting an anomalous tail decay. On the other hand, using the negative β definition achieves better practical results, with the RIRs looking more like practically-measured transfer functions recorded in a real acoustic environment. The EDC approximation method proposed in [1] is therefore based on frequency-domain ISM computations using the negative definition of the β coefficients, which effectively implements a phase inversion upon every sound reflection. This can be seen as a special case of an angle-dependent definition of the reflection coefficient, which can become negative for a certain range of incidence angles.

4) Energy–Time Curve. Given a RIR $h(t)$ computed for a specific environment according to (2), the energy decay envelope $E(t)$ can be computed using a normalized version of the Schroeder integration method:

$$E(t) = 10 \cdot \log_{10}\Bigg(\frac{\int_t^\infty h^2(\xi)\,\mathrm{d}\xi}{\int_0^\infty h^2(\xi)\,\mathrm{d}\xi}\Bigg). \tag{3}$$

This expression can then be used as a basis for deriving an estimate of the reverberation time, such as T_{20} or T_{60} for instance.

2.2. Energy Decay Approximation

The EDC approximation method proposed in [1] is based on the following approach. Similarly to a ray-tracing model, each image source in the ISM technique can be seen as releasing a single sound "particle" (impulse) travelling in the direction of the receiver at the speed of sound. Each particle carries a unit amount of acoustic energy, which decreases upon each reflection on a boundary surface and as a function of the distance travelled to the receiver. Upon reaching the receiver, these sound impulses are then added at the corresponding time lags to create the RIR. As a result, the value of the energy impulse response[1] $h_E(t)$ corresponds to the addition of the energy contributions $a_i(\cdot)$ from all the image sources located on, or very close to a sphere of radius $R = c\cdot t$ around the receiver, with c denoting the sound velocity:

$$h_E(t) = \sum\nolimits_{i\in\mathcal{I}_s} a_i(R, \boldsymbol{r}, \beta_{x,1}, \ldots, \beta_{z,2})\,, \tag{4}$$

where $\mathcal{I}_s$ represents the index set of the sources located on the considered sphere. The energy contribution $a_i(\cdot)$ of the i-th image source can be easily approximated on the basis of its known position on the sphere. The basis of the proposed EDC approximation is then to consider (4) as a Riemann sum that can be represented as the integral of a continuous function $a(\cdot)$ over the entire sphere. The solution to this integral then leads to a closed-form estimate $\hat{h}_E(\cdot) \approx h_E(\cdot)$ of the energy transfer function, reproduced from [1] for convenience in (5) at the top of this page, with $\gamma = 0.5772157\ldots$ the Euler–Mascheroni constant, $\mathrm{Ei}(\cdot)$ denoting the exponential integral, the average room dimension $\bar{r} = (L_x + L_y + L_z)/3$, and $B_\xi = (\beta_{\xi,1}\beta_{\xi,2})^{c\cdot t/L_\xi}$, for $\xi = $ 'x', 'y', 'z'. From (3), the EDC approximation then follows as

$$\widehat{E}(t) \approx 10 \cdot \log_{10}\Bigg(\frac{\sum_{i=0}^{\infty} \hat{h}_E(t+iT)}{\sum_{i=0}^{\infty} \hat{h}_E(t_0+iT)}\Bigg), \tag{6}$$

where t_0 represents the time lag of the first value in the approximation curve. An example of numerical result is shown in Figure 1, which displays a typical EDC obtained from ISM computations together with the curve obtained using (6). If necessary, readers are referred to [1] for more details on this approximation method.

3. REVERBERATION-TIME PREDICTION

This section demonstrates how the proposed EDC approximation method can be used to determine the values of reflection coefficients achieving a desired reverberation time (RT) in a given enclosure. For comparison purposes, this work also considers the results obtained with some classical RT prediction formulae.

3.1. Preliminaries

Consider an enclosure where each boundary surface is assigned an absorption coefficient as follows: $\alpha_{x,1} \triangleq \alpha_1$, $\alpha_{x,2} \triangleq \alpha_2$, $\alpha_{y,1} \triangleq \alpha_3$, and so forth. For the sake of clarity, it is assumed that the coefficients are identical for each pair of opposing walls, i.e., $\alpha_1 = \alpha_2 \triangleq \bar{\alpha}\cdot w_x$, $\alpha_3 = \alpha_4 \triangleq \bar{\alpha}\cdot w_y$, and $\alpha_5 = \alpha_6 \triangleq \bar{\alpha}\cdot w_z$, where w_x, w_y and w_z are absorption weighting factors for the walls in the x, y and z dimension, respectively. This representation allows for clearer derivations related to non-uniform absorption coefficients, which can be simply characterized in terms of the single parameter $\bar{\alpha}$ used in conjunction with the weights vector $\boldsymbol{w} = [w_x, w_y, w_z]$, or alternatively, the weighting ratios $(w_x : w_y : w_z) \equiv (\alpha_1 : \alpha_3 : \alpha_5)$. It must be stressed however that this restriction is without loss of generality as the derivations presented here as well as in [1] are also valid for the case where all coefficients have different values.

In the following, the reverberation time will be characterized using the T_{20} parameter, defined in this work as the time required

[1] The subscript "E" emphasizes the fact that the current developments are based on acoustic energy rather than power.

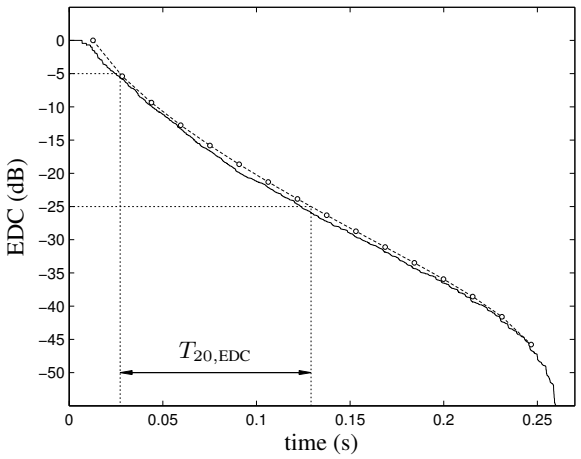

Figure 1: Measurement of T_{20} from the proposed EDC approximation curve (dashed line), simulated for non-uniform reflection coefficients, $\boldsymbol{r} = [3.2\text{m}, 4\text{m}, 2.7\text{m}]^{\text{T}}$, $\boldsymbol{p}_s = [1.1\text{m}, 1\text{m}, 1.2\text{m}]^{\text{T}}$, and $\boldsymbol{p}_r = [2\text{m}, 3\text{m}, 2\text{m}]^{\text{T}}$. The solid line is the EDC obtained from ISM computations for the same setup.

by the RIR energy $E(\cdot)$, defined by (3), to decay from -5dB to -25dB:

$$T_{20} = E^{-1}(-25) - E^{-1}(-5)\,, \tag{7}$$

where $E^{-1}(\xi)$ corresponds to the time lag t_ξ for which $E(t_\xi) = \xi$. The reason for using T_{20} instead of the more common T_{60} parameter is solely in order to reduce the computational load when measuring the reverberation time in a series of simulated RIRs in Section 4. By using the parameter T_{20} instead of T_{60}, it is only necessary to compute the EDC down to -25dB (or slightly below), which involves significantly less image sources during the RIR computations. It is however emphasized that the present developments are valid for any quantity of interest defined on the basis of the EDC, such as T_{60}, T_{30}, early decay time, etc.

3.2. Classical RT Formulae

Previous literature works have made extensive use of well-known RT formulae to solve the problem under consideration. Many RT expressions can be found in the acoustics literature [10], and the present work investigates some of the most commonly used definitions, namely Sabine, Eyring, Millington and Fitzroy's formulae:

$$T_{60,\text{Sab}}(\bar{\alpha}, \boldsymbol{w}) = \frac{0.161 \cdot V}{\sum_{i=1}^{6} S_i\,\alpha_i}\,, \tag{8}$$

$$T_{60,\text{Eyr}}(\bar{\alpha}, \boldsymbol{w}) = \frac{0.161 \cdot V}{-S \cdot \log\!\big(1 - \sum_{i=1}^{6} S_i\,\alpha_i / S\big)}\,, \tag{9}$$

$$T_{60,\text{Mil}}(\bar{\alpha}, \boldsymbol{w}) = \frac{0.161 \cdot V}{-\sum_{i=1}^{6} S_i \cdot \log(1-\alpha_i)}\,, \tag{10}$$

$$T_{60,\text{Fit}}(\bar{\alpha}, \boldsymbol{w}) = \frac{0.161 \cdot V}{S^2} \cdot \bigg(\frac{-2L_yL_z}{\log\big(1-(\alpha_1+\alpha_2)/2\big)} - \frac{2L_xL_z}{\log\big(1-(\alpha_3+\alpha_4)/2\big)} - \frac{2L_xL_y}{\log\big(1-(\alpha_5+\alpha_6)/2\big)} \bigg), \tag{11}$$

where V represents the room volume, S is the total surface area of the enclosure, and S_i, $i \in \{1, \ldots, 6\}$, are the surface areas of each individual wall. Because the expressions in (8)–(11) were derived on the basis of the average sound absorption within the room, or assuming a homogenous spatial distribution of the sound energy (diffuse field), they implicitly assume a linear energy decay in the resulting EDC. It follows that for each of these cases, the T_{20} value is simply defined as $T_{20,(\cdot)}(\bar{\alpha}, \boldsymbol{w}) = T_{60,(\cdot)}(\bar{\alpha}, \boldsymbol{w})/3$.

Given a specific weighting vector $\boldsymbol{w}$, the problem of determining the value of $\bar{\alpha}$ that achieves a desired RT value, denoted here as $T_{20,\text{des}}$, can be seen as a nonlinear optimization problem:

$$\bar{\alpha}_{(\cdot),\text{des}} = \arg\min_{\bar{\alpha}\in[0,1]} \big| T_{20,\text{des}} - T_{20,(\cdot)}(\bar{\alpha}, \boldsymbol{w}) \big|\,. \tag{12}$$

This minimization problem is here solved numerically using a golden section search algorithm with parabolic interpolation.

3.3. RT Prediction using the EDC Approximation

The EDC approximation method described in Section 2.2 can be used in a straightforward manner for the purpose of RT prediction. As depicted in Figure 1, the numerical value of $T_{20,\text{EDC}}(\bar{\alpha}, \boldsymbol{w})$ for a given absorption parameter $\bar{\alpha}$ and weighting vector $\boldsymbol{w}$ is obtained by simply computing the EDC approximation curve according to (6), and determining the corresponding RT value directly from it. A first-order interpolation between points is also used in this process to refine the estimate. Using this approach, the value of $\bar{\alpha}$ yielding the desired reverberation time $T_{20,\text{des}}$ is finally obtained as the solution to the same optimization problem as given by (12).

4. EXPERIMENTAL SIMULATIONS

1) Numerical Results. The RT prediction accuracy of the classical formulae in (8)–(11) as well as the method proposed in Section 3.3 is assessed as follows. Given a specific target reverberation time $T_{20,\text{des}}$ and weight vector $\boldsymbol{w}$, the absorption parameter $\bar{\alpha}$ is determined for each method via (12). The image-source model is then used with the resulting $\bar{\alpha}$ value to generate a number of RIRs in the considered environment, and the "true" reverberation-time value $T_{20,\text{meas}}$ is measured directly from each RIR using the definition in (7). For the proposed EDC approximation method, the frequency-domain ISM algorithm with negative reflection coefficients is used to simulate the RIRs, as described in Section 2.1. Peterson's implementation [9] of the ISM algorithm with positive reflection coefficients is used with the classical RT prediction methods, which represents the approach that is currently widely used in the literature. The resulting error ε is then simply defined, for each RIR, as $\varepsilon = |T_{20,\text{des}} - T_{20,\text{meas}}|$. For a given $T_{20,\text{des}}$, this process is repeated for a total of 30 randomly selected source and receiver positions, in each of eight different rooms, with volumes ranging from 20.25m^3 to 202.5m^3. Figure 2 presents the root-mean-square error (RMSE) $\bar{\varepsilon}$ for each method, averaged over the $K = 240$ resulting error values:

$$\bar{\varepsilon} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} \varepsilon_k^2}\,, \tag{13}$$

where ε_k represents the error value for the k-th simulation. Each plot in Figure 2 presents the results obtained with different ratios $(w_x : w_y : w_z)$ of absorption coefficients, representing various levels of non-uniform sound absorption. Note that the omitted part of the curves corresponds to RT values that are not "physically" achievable given the considered room $\boldsymbol{r}$ and weighting vector $\boldsymbol{w}$.

2) Discussion. Figure 2 shows that the proposed EDC approximation method is able to maintain a low level of error for the considered test scenarios, and provides the best results among the five considered T_{20} prediction methods. None of the classical

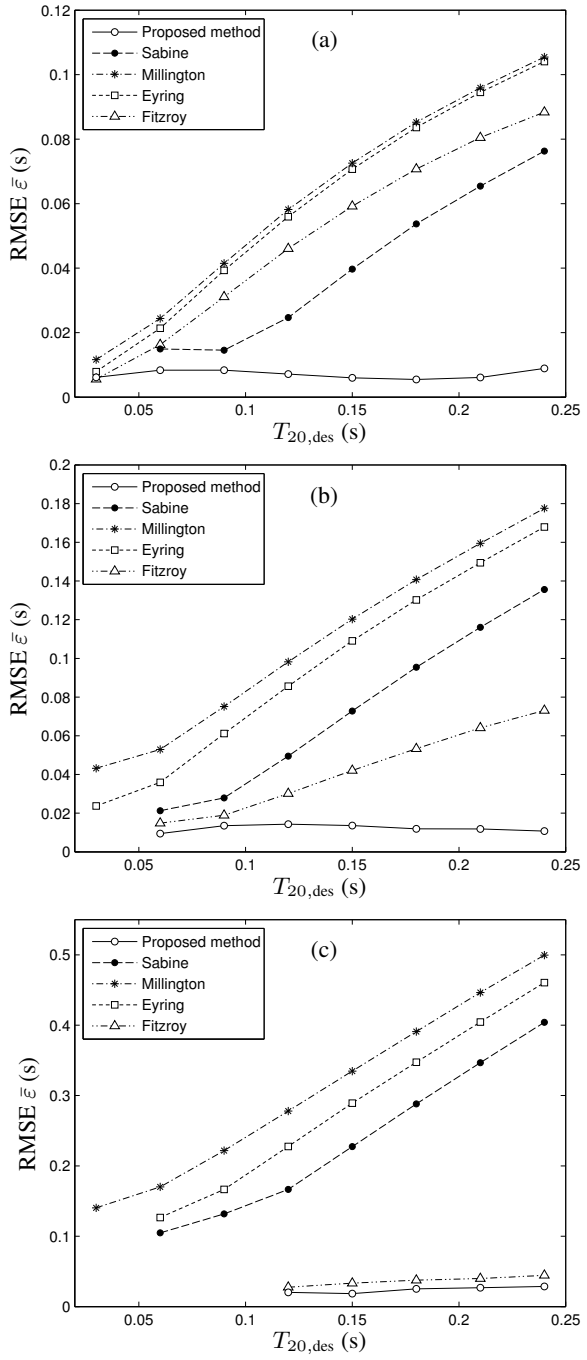

Figure 2: T_{20} prediction error for each considered method. The absorption coefficient ratios $(w_x : w_y : w_z)$ are defined as (a) $(1.0 : 0.8 : 0.6)$, (b) $(1.0 : 0.6 : 0.3)$, and (c) $(1.0 : 0.5 : 0.1)$. Note the different scale of the RMSE axis in each plot.

formulae is able to provide a consistently low level of estimation error, and for most of them, the error becomes larger as the desired RT value increases. This consequently raises some doubts regarding their usefulness in predicting the reverberation level in simulated RIRs, which can be regarded as being of some concern since many publications have been published which provide some sort of performance results based on these formulae (see, e.g., [6–8]).

5. CONCLUSION

This paper demonstrates that the recently proposed EDC approximation method can be used as an efficient tool in the process of assessing the performance of acoustic signal processing algorithms, evaluated on the basis of image-source simulations. This approach establishes a direct relationship between environmental factors, such as the walls' absorption coefficients, and the level of reverberation resulting in the considered enclosure; as shown in this work, this relation is not currently well modeled by classical reverberation-time formulae. Experimental results show that the accuracy of the proposed approximation technique allows to generate a reverberation time in the simulated impulse responses within a small percentage of the targeted value. In order to test audio processing algorithms, the proposed method hence provides researchers with a way of generating realistic-looking impulse responses whose main parameter of interest, namely the reverberation level, can be accurately tuned. This consequently ensures that the algorithms are tested, and performance results are presented, versus a reverberation level that corresponds more or less exactly to what was used during the algorithm simulations. The uniformity of the performance results can thus be guaranteed when comparing the different approaches to a specific audio processing problem presented by various researchers in different publications.

6. REFERENCES

[1] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," submitted to *J. Acoust. Soc. Am.*

[2] J. Allen and D. Berkeley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[3] M. Ikram and D. Morgan, "A multiresolution approach to blind separation of speech signals in a reverberant environment," in *Proc. IEEE ICASSP*, vol. 5, Salt Lake City, UT, USA, May 2001, pp. 2757–2760.

[4] B. Radlović, R. Williamson, and R. Kennedy, "Equalization in an acoustic reverberant environment: robustness results," *IEEE Trans. Speech Audio Proc.*, vol. 8, no. 3, pp. 311–319, May 2000.

[5] E. Lehmann and A. Johansson, "Particle filter with integrated voice activity detection for acoustic source tracking," *EURASIP J. Adv. Sig. Proc.*, vol. 2007, Article ID 50870.

[6] P. Aarabi and G. Shi, "Phase-based dual-microphone robust speech enhancement," *IEEE Trans. Systems, Man, Cybernetics—Part B*, vol. 34, no. 4, pp. 1763–1773, Aug. 2004.

[7] M. Brandstein and H. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. IEEE ICASSP*, vol. 1, Munich, Germany, Apr. 1997, pp. 375–378.

[8] S. Doclo and M. Moonen, "Robust adaptive time delay estimation for speaker localization in noisy and reverberant acoustic environments," *EURASIP J. Appl. Sig. Proc.*, vol. 2003, no. 11, pp. 1110–1124, 2003.

[9] P. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Am.*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.

[10] R. Neubauer and B. Kostek, "Prediction of the reverberation time in rectangular rooms with non-uniformly distributed sound absorption," *Archives of Acoustics*, vol. 26, no. 3, pp. 183–202, 2001.