A SPARSITY MEASURE FOR ECHO DENSITY GROWTH IN GENERAL ENVIRONMENTS

Helena Peić Tukuljac[1*], *Ville Pulkki*[2†], *Hannes Gamper*[3], *Keith Godin*[3], *Ivan J. Tashev*[3], *Nikunj Raghuvanshi*[3]

[1]École polytechnique fédérale de Lausanne (EPFL), Switzerland
[2]Aalto University, Espoo, Finland
[3]Microsoft Research, WA, USA

ABSTRACT

We study the detailed temporal evolution of echo density in impulse responses for applications in acoustic analysis and rendering on general environments. For this purpose, we propose a smooth *sorted density* measure that yields an intuitive trend of echo density growth with time. This is fitted with a general power-law model motivated from theoretical considerations. We validate the framework against theory on simple room geometries and present experiments on measured and numerically simulated impulse responses in complex scenes. Our results show that the growth power of echo density is a promising statistical parameter that shows noticeable, consistent differences between indoor and outdoor responses, meriting further study.

***Index Terms*—** impulse response, echo density, mixing time, outdoor acoustics, parametric models, statistical signal processing

1. INTRODUCTION

Statistical parameters that characterize impulse responses in enclosures, such as the reverberation time, have been extensively studied in room acoustics [1], along with fairly standard estimation algorithms [2]. These parametric models provide insight into impulse responses and enable efficient, natural sounding artificial reverberation [3, 4] and efficient acoustical encoding [5] for interactive auralization. However, parameters characterizing enclosures are insufficient for convincing spatial audio rendering in augmented and virtual reality applications which increasingly feature a rich variety of spaces that are partially or fully outdoors, such as courtyards, forests, and urban street canyons.

We investigate how acoustic responses in such spaces might differ from enclosures, whether obtained through measurement or simulation [6, 7, 8]. In particular, motivated by the common observation that outdoor scenes are sparsely reflecting [7], we study the temporal growth of echo density in the impulse response. Our goal is to characterize how this growth might differ - if at all - between indoor and outdoor acoustic impulse responses, using a parametric power-law model. To our knowledge, such an investigation has not be done before. Prior techniques, compared in [9], study echo density primarily for classifying the first moment when the impulse response is sufficiently diffuse, called the mixing time [10]. This is in contrast to our goal, which is to quantify and analyze the detailed echo density evolution *before* the mixing time.

Our main contribution is a *sorted density* (SD) measure of echo density that enables such an investigation. We show SD to be theoretically meaningful while being robust to complex 3D scenes. In contrast to simple scenes such as a cuboid (shoebox), echo density in complex scenes cannot be defined as number density of non-zero values in the impulse response. Firstly, surface details and irregularities cause wave scattering so that strong reflections do not appear as exact copies of the source pulse in the impulse response, but rather contain substantial linear distortion. Secondly, the distorted strong arrivals are intermixed with numerous weak arrivals from diffuse scattering caused by geometric clutter. This makes it challenging to define and separate out "salient" peaks to measure their temporal density, such as in [11] to estimate mixing time, as compared in [9]. Our sorted density function (illustrated in Figures 1 and 2) is an aggregate measure that avoids peak separation or detection, obviating such difficulties.

We validate our SD measure against theoretical notion of echo density on simple enclosures and observe good agreement. We then apply our technique to measured and simulated impulse responses on complex scenes and observe that echo density growth with time can be modeled well as t^n, where the growth power, $n \approx 2$ indoors and $n \approx 1$ outdoors, with intermediate values in mixed cases. Based on these results, we observe that the growth power of echo density during early reflections is a promising new statistical parameter that discriminates indoor and outdoor acoustics.

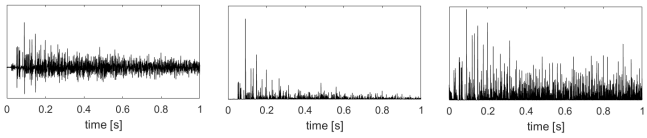

Fig. 1. Input impulse response (left) is converted to an echogram (middle). Local energy normalization factors out the energy decay envelope (right).

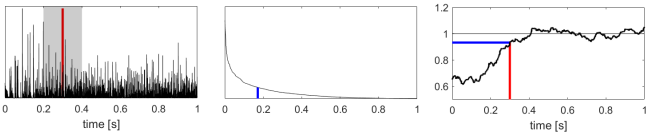

Fig. 2. Normalized echogram is analyzed (left) with a rectangular sliding window (shaded) centered at each sample (red line). Sorted density is computed, as a fraction of window width (middle, blue line). Processing for each sample and normalizing with expected value for Gaussian noise yields echo density (right).

2. PROPOSED ECHO DENSITY MEASURE

Given an input band-limited impulse response $h_i(t)$ we find the first-arrival delay of the direct sound, τ_0. This can be estimated by manual

*Work done as research intern at Microsoft Research, Redmond
†Work done as consulting researcher at Microsoft Research, Redmond

inspection to locate the signal onset, or using a detection algorithm [5]. Direct sound is removed by setting: $h_i(t) = 0, t < \tau_0 + 10$ ms. This yields the input response to the echo density estimation, $h(t) \equiv h_i(t + \tau_0), t \geq 0$. Echo density is then computed using a two-pass procedure, illustrated in Figures 1 and 2 respectively.

2.1. Local energy normalization

The input response is converted to an echogram, $e(t) \equiv h^2(t)$. The first pass performs local energy normalization which factors out the energy decay in the response thus ensuring that number density estimates are not biased by the overall energy envelope of the response, making the measure fairly insensitive to the reverberation time. We normalize each sample value with the local mean of surrounding samples weighted with a Tukey window w,

$$\tilde{e}(t) = \frac{e(t)}{\int e(t+\tau)\, w(\tau)\, d\tau}\,. \tag{1}$$

We have used continuous time notation for brevity, the integrals are to be understood as discrete summation. The width of the window defines the temporal locality for normalization. A half-width of $T_n = 10$ ms corresponds to the interval of perceptual echo fusion [12] and was found to work well in practice. The Tukey window is normalized so that $\int w(\tau)\, d\tau = 1$. The symmetric cosine tapering segments have width of 5 ms each with a 10 ms long constant segment in the middle. As the example in Figure 1 shows, the resulting signal is much more amenable for sparsity analysis, emphasizing peaks without explicit detection.

2.2. Sorted Density (SD)

We employ a simple measure of sparsity in a discrete positive signal $s[i]$. Our main idea is to sort the signal to yield a monotonically decreasing signal $\hat{s}[i]$. The sparser s is, the faster $\hat{s}$ will fall off as a function of number of samples. Any smooth measure of the width of $\hat{s}$ normalized with number of samples should then yield a notion of fractional energy density in the signal. An example is shown in Figure 2.

A natural way to compute width is via first-moment of sample index i with $\hat{s}$ serving as weight. This is the sorted density functional,

$$\mathcal{D}[s] \equiv \frac{1}{m} \frac{\sum_{i=1}^{m} i\, \hat{s}[i]}{\sum_{i=1}^{m} \hat{s}[i]} \tag{2}$$

where m is the number of samples in the observed window. The sorted density is a unitless measure with values ranging between 0 and 0.5 corresponding respectively to minimal echo density when s contains a single non-zero sample, to maximum when all values are non-zero and equal. Gaussian noise has an intermediate (expected) value of $\mathcal{D}[g] = 0.18$.

We then estimate the echo density function for the input response, $h(t)$, by employing a sliding rectangular time window on the normalized echogram, $e_n(t)$ and computing the sorted density in each window:

$$N'_{sd}(t) = \frac{\mathcal{D}[\tilde{e}\,(t \in (t - T_l, t + T_l)]}{\mathcal{D}[g]}, \tag{3}$$

where any samples $e_n(t)$ for $t < 0$ are discarded from the analysis. Note the normalization with $\mathcal{D}[g]$, so that an echo density of $N'_{sd} = 1$ indicates Gaussian noise. T_l is the half-width of the rectangular window and we empirically found $T_l = 100$ ms to work well. As shown in Figure 2 this yields an intuitive trend of echo density that initially increases and then settles near some maximum value (close to 1 indoors) as the response transitions to late reverberation.

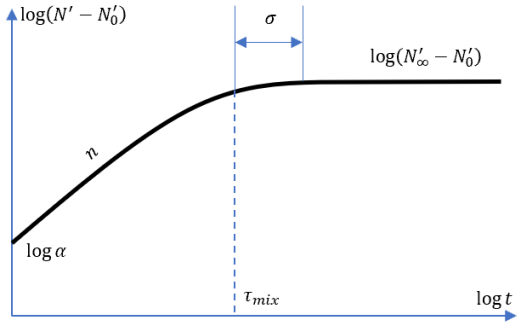

Fig. 3. Log-domain parametric model that is fitted to extracted echo density trend.

3. STATISTICAL MODEL

We describe our general model for echo density growth, analytical motivation and fitting procedure.

3.1. Analytical motivation

For simple geometries such as a shoebox (rectangular) room where geometric acoustics is accurate the echo density may be defined rigorously by counting the number $N(t)$ of geometric paths that arrive at the listener within time t after the source emits an impulse. For any source location, the corresponding image sources form a periodic, discrete sampling of 3D space. Observing that the maximum propagation path length until time t is ct where c is the speed of sound, we have: $N(t) \propto (ct)^3$ by counting all image sources in the spherical ball with radius ct. Taking the time derivative to convert echo count to echo density, the full expression is [1, p. 110],

$$N'(t)_{indoor} = \alpha t^2, \tag{4}$$

where α is a geometry-dependent parameter, given by $\alpha = 4\pi c^3/V$ for room volume V. This result also holds true under theoretically ideal diffuse field conditions. Note that this model describes the behaviour only up to the mixing time τ_{mix} where the impulse response approaches noise so that $N'(t)$ approaches a constant.

Removing the roof of the shoebox yields a courtyard-like geometry with 4 surrounding walls and a ground. This represents a reverberant outdoor scene where most reflectors surround the source and listener horizontally. Ignoring edge diffraction from the top wall edges, the image sources occupy a periodic sampling of 2D (rather than 3D) space, so that number of echoes $N(t) \propto (ct)^2$ and the echo density, $N'(t)_{outdoor} \propto t$. Based on these observations, we hypothesize the general model for any acoustical environment:

$$N'(t; N'_0, \alpha, n, \tau_{\text{mix}}) = \begin{cases} N'_0 + \alpha t^n, t < \tau_{\text{mix}} \\ N'_\infty, t \geq \tau_{\text{mix}} \end{cases}, \tag{5}$$

where $N'_\infty \equiv N'_0 + \alpha \tau_{\text{mix}}^n$ to ensure continuity, and $\{N'_0, \alpha, n, \tau_{\text{mix}}\}$ are the model parameters. The analytical results above do not apply near $t = 0$ or $t = \tau_{\text{mix}}$. Near $t = 0$ one must have some non-zero echo density, N'_0, due to first reflections, followed by power-law growth that remains continuous and then stabilizes near some maximum value, N'_∞ at the mixing time, τ_{mix}. The continuous parameter n is the focus of our experiments, with the hypothesis that it should be ~ 1 outdoors and ~ 2 indoors based on analytical considerations above. Some geometric information about scene size is also contained in α, although its interpretation has a dependence on n, whose study we leave for future work.

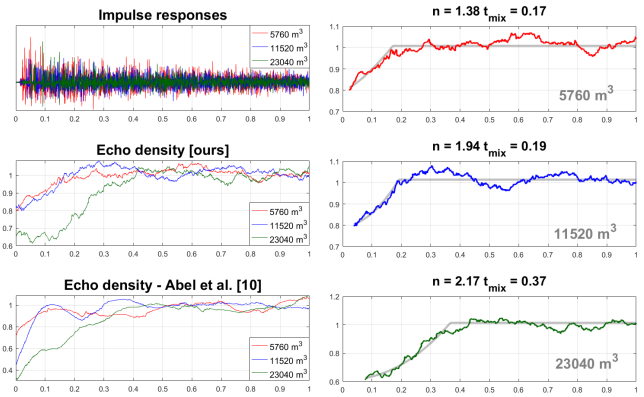

Fig. 4. Validation of method on shoebox scenes. Impulse responses are on left top. Three rooms are tested with volumes increasing by factor of two. Fitted models are plotted in grey color. Our echo density measure shows a growth power $n > 1$ as expected for indoors (right column).

3.2. Model fitting

To robustly estimate the growth power, n, we first separately estimate N'_0. We then perform fitting on $\log(N' - N'_0)$. As illustrated in Figure 3, this simplifies the model in Eq. 5 to two linear segments respectively that meet at $t = \tau_{\text{mix}}$: $\log(\alpha) + n\log(t)$ and $\log(N'_\infty - N'_0)$. To reduce sensitivity in fitting due to non-smooth model at $t = \tau_{\text{mix}}$, we cross-fade between the two linear segments via a sigmoid window

$$W(t; \tau_{\text{mix}}, \sigma) = \frac{1}{2}\Big(1 - \tanh\Big(\frac{t - \tau_{\text{mix}}}{\sigma}\Big)\Big). \tag{6}$$

The parameter σ controls width of the cross-fade, which we set to $\sigma = 20$ ms. The resulting smoothed parametric model is

$$\begin{aligned}\log(N'(t; \alpha, n, \tau_{\text{mix}}) - N'_0) = W \cdot (\log\alpha + n\log t) + \\ (1 - W) \cdot \log(N'_\infty - N'_0).\end{aligned} \tag{7}$$

Given the observed echo density profile N'_{sd} from Eq. 3, we estimate N'_0 as the minimum value of the echo density, $\min\{N'_{sd}(t)\}$ and then fit the above model to $\log(N'_{sd} - N'_0)$ using non-linear least squares. We constrain the search space to accelerate convergence. The search for α is unbounded, but for n is bounded by $[0, 5]$ and for $(N'_\infty - N'_0)$ is bounded by $[0, 2]$. With this choice of bounds we have avoided manual tuning in the fitting procedure, since the observations have implied that the sufficient upper bounds would be 2.5 and 0.5, respectively.

4. RESULTS

Our experiments have two goals. First, we compare against theory on enclosures to validate our technique. Second, we compare the echo density growth power, n, between indoor and outdoor cases.

4.1. Experimental data

Experiments are performed on impulse responses acquired from both measurements and 3D wave simulations. Simulations allow tests with tightly controlled 3D geometry, but are necessarily band-limited due to computational cost restrictions. We use the time-domain spectral wave solver [13]. All simulations are band-limited to 1kHz with sampling frequency of 6kHz with the source and microphone placed close to the center of the room, but off the axes of symmetry and more than 1m apart. With these constraints, the results were not found to be sensitive to exact placement. Surface absorptivity was set to 0.05 for all frequencies in all simulations. While measured responses necessarily contain more noise, we have noticed that the higher sample rate improves the reliability of our technique, presumably because there is a larger number of samples within each analysis window for statistical estimation.

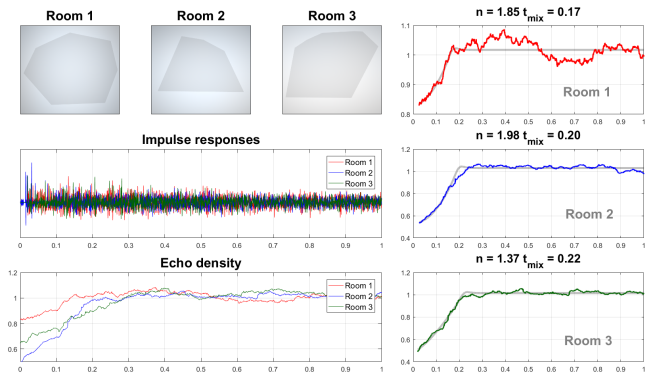

Fig. 5. Echo density on simulated convex polyhedral rooms with flat ground and ceiling.

4.2. Validation on simple enclosures

If our sorted density measure (Eq. 3) is a valid generalization of the theoretical notion of echo density (Eq. 4), we expect $n \approx 2$ on simple enclosures where geometric acoustics underlying Eq. 4 is reasonably accurate. We test this hypothesis with simulations on two types of such geometries: shoebox and convex polyhedron.

Figure 4 shows experiments on three shoebox rooms with volume increasing by factor of two. Input responses are on left top. Compare our echo density measure (left middle) to [14] (left bottom), with the latter using the same window half-width T_l as our method. Both techniques are normalized so a value of 1 indicates late reverberation. Both techniques show an increasing trend, reaching around 1 at similar mixing times, τ_{mix}. However, our measure is designed to also model echo density growth before τ_{mix}, as shown on the right. All cases show a growth power $n > 1$ as expected for indoors, with the two larger rooms agreeing well with theory with $n \approx 2$. For the smallest room however, n is smaller. We observe this systematic bias for smaller spaces with our technique. Echo density buildup is quick in small rooms, leaving a short span for model fitting. Our sorted density analysis window is also quite wide with $T_l = 100$ ms which is a contributing factor, but we found this width necessary to build reliable statistics.

Figure 5 tests three general convex polyhedral room geometries with large flat reflectors. The polyboxes were randomly generated such that their volume is within $[10000, 20000]\text{m}^3$. The echo density shows a close to quadratic growth in the first two cases with more irregular geometry, agreeing well with theory. In some cases, like "Room 3," we observe a decrease in n, perhaps because of flutter echoes between the two large near-parallel faces. Such periodicity in the response also motivated avoidance of symmetry axes in the shoebox tests.

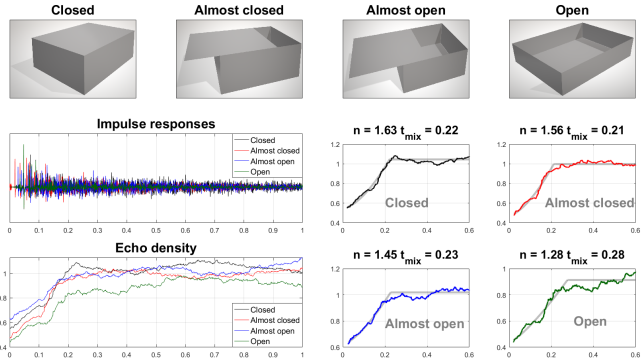

Fig. 6. A simulated shoebox room that gradually transforms to a courtyard. Echo density growth power, n, decreases smoothly as the scene progresses towards outdoors.

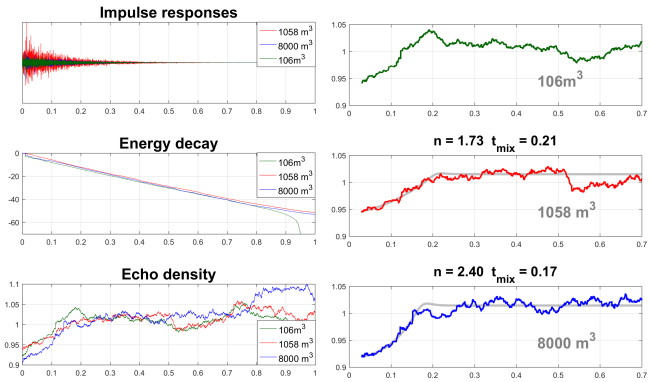

Fig. 7. Comparison on measured responses in three rooms with different volumes, but same reverberation time. For the two larger rooms, n is around 2, agreeing well with expectations.

4.3. Indoor to outdoor scene modification

As discussed in Section 3.1 if we remove the roof of a shoebox to turn it into a "courtyard", we theoretically expect $n = 1$, with some deviations caused by edge diffraction. We performed simulations in a shoebox room with a ceiling that gradually opens, as shown in Figure 6. As the roof is removed, the value of n smoothly decreases from near 2 towards 1, with intermediate values in the middle. This fits with theoretical expectations on the closed and open extremes, and also illustrates that the technique is resilient to mixed cases somewhere between indoors and outdoors.

4.4. Varying volume with fixed reverberation time

Figure 7 compares measured impulse responses on three enclosures with large variation in scene volume but differing absorptivity so that the reverberation times are similar. The three measurements were taken from the Reverb Challenge corpus ("Room 2", 106 m^3) [15], and from the Open AIR database ("Dixon Studio, York University Theatre", 1058 m^3, "Central Hall, York University", 8000 m^3) [16]. The energy decay curves are nearly identical (left column, middle). All of the measurements have a sampling frequency of 16kHz. In all cases the echo density trend is plausible, increasing and settling near 1. For the two larger rooms, we observe values of $n \approx 1.7$ and 2.4, corresponding well to indoors, with the smaller of the two rooms producing smaller value, a bias we noted earlier. Regression fails on the smallest room with volume similar to a small office ($\approx 100\text{m}^3$) indicating that our regression could be improved to handle small rooms better.

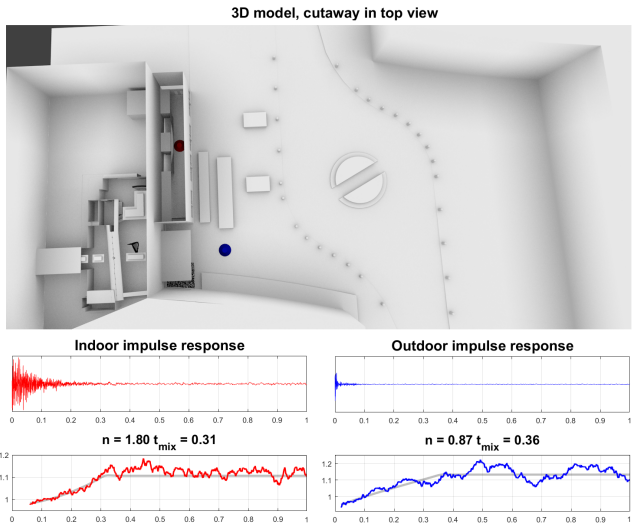

Fig. 8. Measurements were performed in the two locations shown in the 3D cutaway top view, indoors (red) and outdoors (blue). The two locations are clearly differentiated by $n = 1.8$ and 0.87 respectively.

4.5. Indoor versus outdoor location in urban area

We measured impulse responses in urban office building at two locations inside and outside, shown on a 3D cutaway top view in Figure 8. Sampling frequency was 48kHz. We find values of n in good agreement with expectations, 1.80 indoors and 0.87 outdoors, showing a clear difference between indoor and outdoor acoustics in a highly complex scene.

5. CONCLUSION AND FUTURE WORK

We described a novel sorted density method for estimating echo density growth in acoustic impulse responses which is fitted with a simple power-law model for general scenes. The method is found to agree well with theory. Experiments on measured and simulated responses show that the growth power of echo density, n, shows promise as a salient statistical parameter differentiating indoor and outdoor acoustics. We wish to improve the robustness of the method in the future, especially for small rooms. The size parameter, α, and mixing time, τ_{mix}, contain geometric information about the scene. But in outdoor cases ($n \approx 1$) they no longer admit interpretation in terms of "room" volume. A study on the geometric interpretation of these parameters in general scenes could prove to be a fruitful future direction.

6. ACKNOWLEDGEMENTS

Thanks to Jon Hanzelka and Pedro Urbina Escos from the Microsoft Synthetics team for the LIDAR scan and 3D model of the building used in our "urban area" tests.

7. REFERENCES

[1] Heinrich Kuttruff, *Room Acoustics*, Taylor & Francis, 4 edition, 2000.

[2] Anders Gade, "Acoustics in Halls for Speech and Music," in *Springer Handbook of Acoustics*, Thomas Rossing, Ed., chapter 9. Springer, 2007 edition, May 2007.

[3] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, July 2012.

[4] S. J. Schlecht and E. A. P. Habets, "Feedback delay networks: Echo density and mixing time," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 374–383, Feb 2017.

[5] Nikunj Raghuvanshi and John Snyder, "Parametric directional coding for precomputed sound propagation," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 108:1–108:14, July 2018.

[6] K Spratt and J.S. Abel, "A digital reverberator modeled after the scattering of acoustic waves by trees in a forest," vol. 2, pp. 1284–1293, 01 2008.

[7] F. Stevens, D. T. Murphy, L. Savioja, and V. Välimäki, "Modeling sparsely reflecting outdoor acoustic scenes using the waveguide web," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1566–1578, Aug 2017.

[8] Ravish Mehra, Nikunj Raghuvanshi, Anish Chandak, Donald G. Albert, D. Keith Wilson, and Dinesh Manocha, "Acoustic pulse propagation in an urban environment using a three-dimensional numerical simulation," *The Journal of the Acoustical Society of America*, vol. 135, no. 6, pp. 3231–3242, 2014.

[9] Alexander Lindau, Linda Kosanke, and Stefan Weinzierl, "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," in *Audio Engineering Society Convention 128*, May 2010.

[10] Jean-Dominique Polack, "Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics," *Applied Acoustics*, vol. 38, no. 2, pp. 235 – 244, 1993.

[11] 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. 1071–1081, 2009.

[12] Ruth Y. Litovsky, Steven H. Colburn, William A. Yost, and Sandra J. Guzman, "The precedence effect," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.

[13] Nikunj Raghuvanshi, Rahul Narain, and Ming C. Lin, "Efficient and accurate sound propagation using adaptive rectangular decomposition," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 5, pp. 789–801, Sept. 2009.

[14] Jonathan S. Abel and Patty Huang, "A Simple, Robust Measure of Reverberation Echo Density," in *Audio Engineering Society Convention 121*, 2006.

[15] Keisuke Kinoshita, Marc Delcroix, Takuya Yoshioka, Tomohiro Nakatani, Armin Sehr, Walter Kellermann, and Roland Maas, "The reverb challenge: A common evaluation framework for dereverberation and recognition of reverberant speech," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013 IEEE Workshop on*. IEEE, 2013, pp. 1–4.

[16] Damian T Murphy and Simon Shelley, "Openair: An interactive auralization web resource and database," in *Audio Engineering Society Convention 129*. Audio Engineering Society, 2010.