Analysis of multi-exponential and anisotropic sound energy decay

Citation for published version:
Götz, G, Hold, C, McKenzie, T, Schlecht, S & Pulkki, V 2022, Analysis of multi-exponential and anisotropic sound energy decay. in *Jahrestagung für Akustik, Stuttgart, Germany, 21/03/2022.* German Acoustical Society (DEGA).

Document Version:
Peer reviewed version

Published In:
Jahrestagung für Akustik, Stuttgart, Germany, 21/03/2022

Analysis of multi-exponential and anisotropic sound energy decay

Georg Götz
Aalto University

Christoph Hold
Aalto University

Thomas McKenzie
Aalto University

Sebastian Schlecht
Aalto University

Analysis of multi-exponential and anisotropic sound energy decay

Georg Götz[1*], Christoph Hold[1], Thomas McKenzie[1], Sebastian J. Schlecht[1,2], Ville Pulkki[1]
[1] *Aalto Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland*
[2] *Media Lab, Department of Art and Media, Aalto University, Espoo, Finland*
[*] *Email: georg.gotz@aalto.fi*

Introduction

Reverberation is an important cue for determining the distance and location of sounds, and can be used to infer the size of a space. Although simple models of reverberation assume a single exponentially decaying late reverberation tail that is independent of direction, in practice, many rooms feature multiple exponential decays with directional anisotropy [1, 2, 3].

In this paper, a framework for the directional analysis of spatial room impulse response reverberation decays is presented. The framework uses a recent neural network approach for multi-exponential decay analysis in conjunction with a spherical filterbank analysis. Additionally, the paper introduces the common-slope model of directional reverberation, in which directional or spatial decay variations are described in terms of exponential amplitudes, while fixing the corresponding decay times.

Sound energy decay analysis

The sound energy decay in rooms can be described with energy decay functions (EDFs), which can be calculated from room impulse responses (RIRs) via the backwards integration procedure proposed by Schroeder [4]. For an RIR $h(t)$, the EDF is given by [4]

$$d(t) = \frac{1}{E} \sum_{\tau=t}^{L} h^2(\tau)\,, \quad \text{with} \quad E = \sum_{\tau=1}^{L} h^2(\tau)\,, \tag{1}$$

where t is the sample index and L is called the upper limit of integration (ULI), corresponding to the number of samples in the EDF.

The sound field in coupled rooms or rooms with non-uniform absorptive material distributions can decay with more than one decay rate [1, 2, 3], thus making it practical to use a decay model that consists of multiple exponential decays and a noise term. In this paper, we model the sound energy decay with the model $d_K(t)$ given by [5]

$$d_K(t) = N_0(L-t) + \sum_{i=1}^{K} A_i \left[e^{\frac{\ln(10^{-6})\cdot t}{f_s T_i}} - e^{\frac{\ln(10^{-6})\cdot L}{f_s T_i}} \right]\,, \tag{2}$$

where K is the model order, T_i and A_i are the decay rate and the amplitude of the i^{th} exponential decay, respectively, N_0 is the amplitude of the noise term, $\ln(\cdot)$ denotes the natural logarithm, and f_s is the sampling frequency of the RIR. The second exponential term in the square brackets is a constant that accounts for the finite ULI [5].

In this paper, we use a recently proposed neural-network-based approach [6] for estimating the model parameters T_i, A_i, and N_0.

Directional Analysis

The spatial room impulse response (SRIR) input signal $\boldsymbol{h}_{nm}^{\text{in}}$ is typically given in the spherical harmonics (SH) domain. Directional sound field analysis usually requires dividing the sound field into multiple directionally-constrained regions. For this purpose, we use a set of spatial filters (or "beamformers"), applied in the SH domain. We previously proposed a spherical filter bank (SFB) that allows obtaining such directionally constrained analysis signals as a set of beamformer outputs [7]. These beamformers are designed such that they maintain the full spherical coverage, while preserving the input amplitude or energy. Additionally, the utilized spherical filter bank structure allows modifying the directional analysis patterns.

Previous studies have used the plane-wave decomposition (PWD) for analyzing the sound energy decay in different directions [8, 9]. Directional analysis with the PWD is equivalent to using a higher-order hypercardioid or maximum directivity pattern. Despite its widespread usage in various applications, the side- and backlobe suppression of the maximum directivity pattern is only moderate. However, backlobe suppression is especially critical in room acoustic analysis scenarios, in which large differences along one axis need to be resolved. Therefore, our directional decay analysis utilizes the described directional analysis [7] with a spatial Butterworth filter, which exhibits a greater front-to-back-separation as detailed in [10]. We extract the output signals $\boldsymbol{s}_\xi$ of beamformer $\xi \in [1 \dots J]$ as

$$\boldsymbol{s}_\xi = \boldsymbol{A}\boldsymbol{h}_{nm}^{\text{in}} \quad . \tag{3}$$

The analysis matrix $\boldsymbol{A} \in \mathbb{R}^{J\times(N+1)^2}$ is comprised of stacked beamformers $\boldsymbol{w}_{\xi,nm}$ and is formulated directly with stacked steering vectors $\boldsymbol{y}_n^m(\Omega_\xi)$ to a matrix $\boldsymbol{Y}$ as $\boldsymbol{A} = \boldsymbol{Y}\,\text{diag}_N(\boldsymbol{c}_n^{\text{an}})$, where $\boldsymbol{c}_n^{\text{an}}$ describes the axisymmetric analysis pattern as modal weighting coefficients. For the spatial Butterworth filter, the modal weighting coefficients are given by $c_n^{\text{Butterworth}} = \frac{1}{\sqrt{1+(n/n_c)^{2k}}}$, where we set the cuton SH order $n_c = 3$ and filter order $k = 5$ for the analysis of this paper.

Utilizing the spherical filterbank analysis allows decomposing the sound energy decay in various directions. A directional sound decay analysis was previously proposed by Berzborn and Vorländer [8]. In the remainder of this paper, we will use the term directional EDF (DEDF) for EDFs calculated from the beamformer signals $\boldsymbol{s}_\xi$.

Common-slope modeling of directional reverberation

Sound energy decay analysis is traditionally carried out in a limited number of frequency bands. However, generally, the sound field contains a large number of sinusoidal room modes with various decay rates [11]. While the mode frequencies and decay rates depend only on the room geometry and wall properties, the mode amplitudes depend on the source-receiver configuration and excitation signal [11]. The relationship between decaying room modes and EDFs has been detailed by Kuttruff [11, pp. 82-88].

In the following, we introduce the common-slope modeling of directional reverberation, which is inspired by the common-acoustical-pole and residue (CAPR) model by Haneda et al. [12]. While the CAPR model assumes constant mode frequencies and decay rates for different source-receiver configurations, we assume that the decay rates T_i of Eq. (2) are independent of source and receiver position and orientation. In other words, the decay rates T_i only depend on the room geometry and wall properties. This assumption allows estimating a set of reference T_i values that are kept constant while analyzing multiple EDFs of the same room. Therefore, the common-slope analysis can be applied to a set of directional EDFs or to EDFs of different source-receiver configurations. By keeping the T_i values fixed, energy variations of the late reverberation become more evident because they can be represented as changes in the A_i values.

The reference T_i values can be estimated as usual with currently available decay analysis approaches like the DecayFitNet [6] or Bayesian decay analysis [13]. After fixing the T_i values, the decay analysis for all other EDFs simplifies to a linear least-squares problem, in which only the amplitudes A_i and noise values N_0 need to be estimated.

Proposed analysis framework

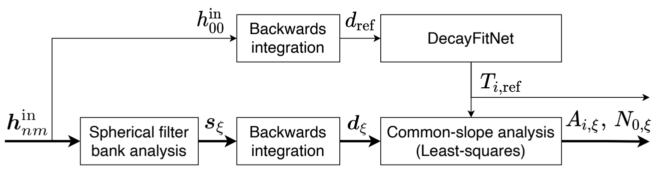
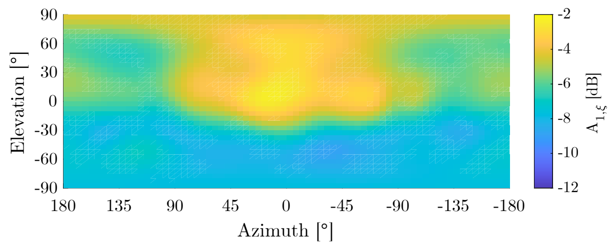


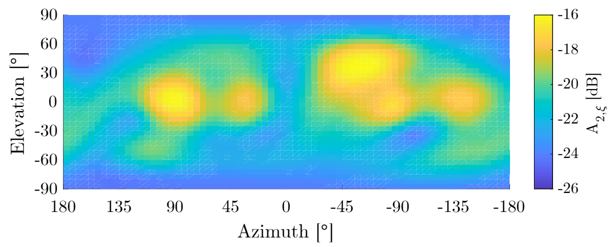
Figure 1: Block diagram of the proposed framework for multi-exponential and anisotropic sound energy decay analysis with common slopes. Bold arrows indicate multiple signals or values are passed for Ambisonic channels or beamformers.

Figure 1 summarizes our proposed analysis framework, which combines the above concepts of spherical filterbanks, multi-exponential sound decay analysis, and common-slope modeling. An SRIR $\boldsymbol{h}_{nm}^{\text{in}}$ in Ambisonics format is analyzed with the spherical filterbank to produce the beamformer output signals $\boldsymbol{s}_\xi$. The resulting signals $\boldsymbol{s}_\xi$ are then backwards-integrated to yield the DEDFs $\boldsymbol{d}_\xi$. Similarly, an EDF is calculated from the omnidirectional SRIR channel h_{00}^{in}. The omnidirectional EDF is chosen as the reference EDF d_{ref}, which is subsequently analyzed with the DecayFitNet to obtain a set of reference decay rates $T_{i,\text{ref}}$. Finally, a common-slope decay model with the decay rates $T_{i,\text{ref}}$ is fit to all DEDFs to yield the corresponding $A_{i,\xi}$ and $N_{0,\xi}$ values for all directions.

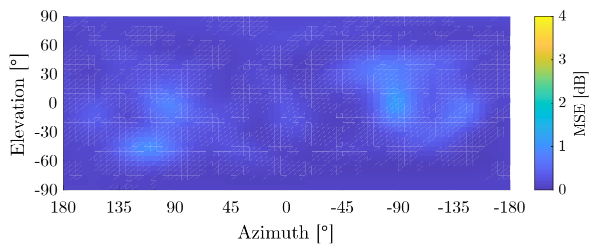
Results: concert hall



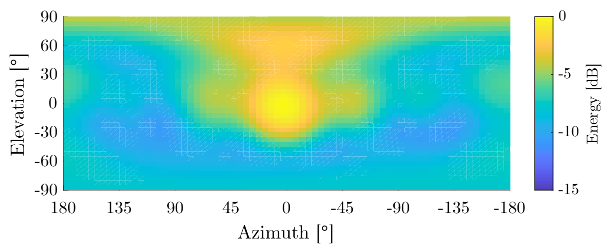
(a) $A_{1,\xi}$ values ($T_{1,\text{ref}} = 1.81\,\text{s}$)



(b) $A_{2,\xi}$ values ($T_{2,\text{ref}} = 2.55\,\text{s}$)



(c) Mean squared error (MSE) [c.f. Eq. (4)] between common-slope model fits and measured DEDFs (average MSE: 0.21 dB)



(d) Total energy

Figure 2: Common-slope amplitude and total energy variation across different directions for a concert hall measurement. The analysis is based on the DEDFs in the 1 kHz octave band and uses a maximum spherical harmonics order $N_{\text{sph}} = 4$ and decay model order $K = 2$.

First, we demonstrate our analysis framework on a concert hall SRIR measurement. The analyzed SRIR is part of the *SRIR compilation captured at the WDR Broadcast Studios* [14]. It was measured with a VariSphear spherical scanning microphone array (110 microphone positions on a Lebedev grid, including rigid sphere extension) and an omnidirectional Sonic Ball sound source at the large broadcast studio (Klaus-von-Bismarck-Saal; volume: $6461.56\,\text{m}^3$).

Using our proposed analysis framework, we analyzed the concert hall SRIR with maximum spherical harmonics order $N_{\mathrm{sph}} = 4$ and decay model order $K = 2$. The analysis is based on the DEDFs in the 1 kHz octave band. Figure 2a and 2b show respectively the estimated $A_{1,\xi}$ and $A_{2,\xi}$ values for various beamformer directions. The reference decay rates were estimated on the omnidirectional EDF and amount to $T_{1,\mathrm{ref}} = 1.81\,\mathrm{s}$ and $T_{2,\mathrm{ref}} = 2.55\,\mathrm{s}$, respectively. Both decay amplitudes vary approximately 6 dB across all directions. A distinct maximum can be observed for the $A_{1,\xi}$ values at frontal directions. Furthermore, the $A_{1,\xi}$ values are generally higher in the upper than in the lower hemisphere. This effect could be caused by increased absorption due to seating. The $A_{2,\xi}$ values are considerably lower than the $A_{1,\xi}$ values and also show multiple clear peaks. While the $A_{2,\xi}$ values seem to be highest at lateral and rear directions, the peaks do not coincide exactly with the side and back walls.

Figure 2c shows the mean squared error (MSE) between the measured DEDFs $d_{\mathrm{dB}}(t)$ and common-slope model fits $d_{K,\mathrm{dB}}(t)$

$$\mathrm{MSE} = \frac{1}{L} \sum_{t=0}^{L-1} \left[d_{\mathrm{dB}}(t) - d_{K,\mathrm{dB}}(t)\right]^2 , \qquad (4)$$

where both EDFs are converted to a logarithmic scale (in dB) prior to calculating the MSE.

The fitting error is below 1.2 dB across all directions and has an average value of 0.21 dB, demonstrating that the common-slope model is suitable for describing the directional energy decay in the analyzed concert hall.

Figure 2d depicts the total beamformer signal energy across directions. The plot features a strong peak at the sound source direction due to the direct sound. Slightly increased energy can also be observed for reflections from the right side and rear walls, ceiling, and floor. However, in contrast to Figures 2a, the directionality of the late reverberation does not become evident.

Results: coupled room

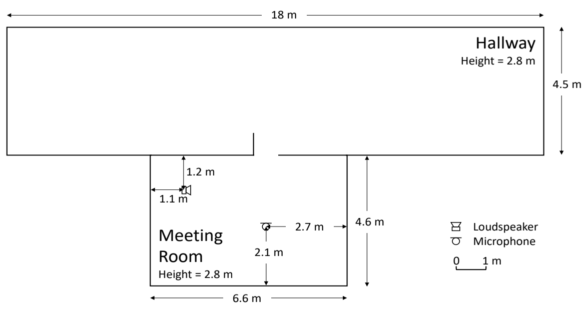
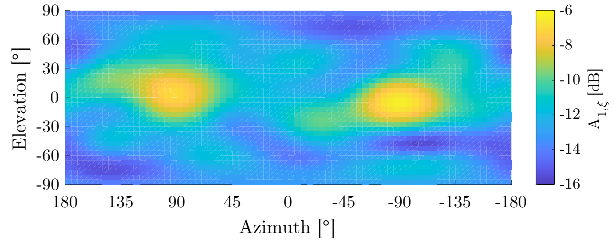
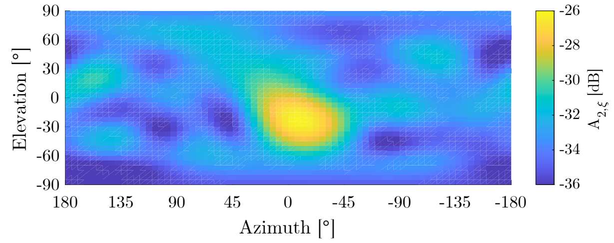


Figure 3: Coupled room measurement setup [9].

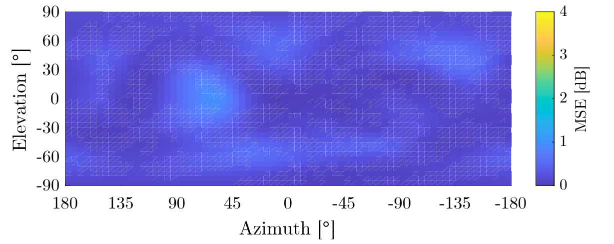
Next, we demonstrate our analysis framework on a coupled room SRIR measurement from the dataset in [9]. The analyzed measurement is taken at the 0 cm position of the transition *"meeting room to hallway, source in the meeting room, no continuous line of sight"*. It was obtained using a Genelec 8331A coaxial loudspeaker which



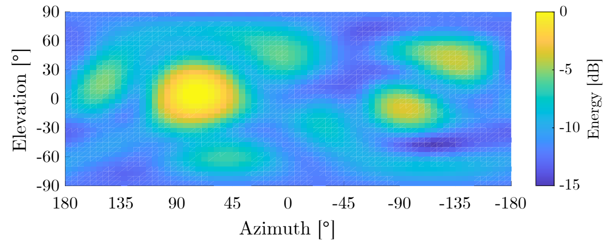
(a) $A_{1,\xi}$ values ($T_{1,\mathrm{ref}} = 0.47\,\mathrm{s}$)



(b) $A_{2,\xi}$ values ($T_{2,\mathrm{ref}} = 1.47\,\mathrm{s}$)



(c) Mean squared error (MSE) [c.f. Eq. (4)] between common-slope model fits and measured DEDFs (average MSE: 0.23 dB)



(d) Total energy

Figure 4: Common-slope amplitude and total energy variation across different directions for a coupled room measurement. The analysis is based on the DEDFs in the 1 kHz octave band and uses a maximum spherical harmonics order $N_{\mathrm{sph}} = 4$ and decay model order $K = 2$.

is a directional source. The room geometry and measurement setup of the utilized SRIR is presented in Figure 3.

Figure 4 summarizes the common-slope analysis results for the coupled room measurement. The analysis is based on the 1 kHz octave band, maximum spherical harmonics order $N_{\mathrm{sph}} = 4$, and decay model order $K = 2$. The directionality of the $A_{1,\xi}$ and $A_{2,\xi}$ values is depicted in Figure 4a and 4b, respectively. The reference decay rates of the omnidirectional EDF amount to $T_{1,\mathrm{ref}} = 0.47\,\mathrm{s}$ and $T_{2,\mathrm{ref}} = 1.47\,\mathrm{s}$, respectively. It is straightforward to associate those decay rates with the involved rooms ($T_{1,\mathrm{ref}}$: meeting room, $T_{2,\mathrm{ref}}$: hallway).

The $A_{1,\xi}$ values have a distinct maximum at the source direction and the opposite wall. This directionality can be attributed to the loudspeaker directivity and its effect on the reverberation. In contrast, the $A_{2,\xi}$ values exhibit a peak at the aperture direction, which highlights the energy contribution from the coupled room. Interestingly, the peak is slightly shifted toward the lower hemisphere. This phenomenon could be caused by the aperture size and position. In this measurement, the aperture is a door, the height of which is smaller than the wall height. While there is a direct line of sight from the receiver to the floor of the coupled room, its ceiling is slightly obstructed by the wall portion above the door.

Figure 4c shows the MSE [c.f. Eq. (4)] between the measured DEDFs and common-slope model fits. The fitting error is below 1 dB across all directions, resulting in an average MSE of 0.23 dB. The low fitting error demonstrates that the common-slope model is also suitable for describing the sound decay in the coupled room scenario.

Figure 4d depicts the total beamformer signal energy across different directions. While the plot exhibits a distinct peak at the source direction and the opposite wall, the other smaller peaks cannot be attributed to specific room features. In contrast to Figure 4a and 4b, the coupled room characteristic and the directionality of the late reverberation are not evident from this figure.

Applications

The previous results show that the common-slope model of directional late reverberation is a valuable tool for analyzing the sound energy decay in rooms with complex geometries. Apart from improving our understanding of room acoustics, common slopes can also be used as a model for room acoustic synthesis tasks.

For example, we recently proposed a SRIR denoising approach, which combines a spherical filterbank and multi-exponential decay analysis to replace the late reverberation of a noisy SRIR with a resynthesized noise-free late reverberation tail [10]. With classical decay analysis methods, slightly different decay rates would be determined for the various analysis directions. The different decay rates would overfit the measured data and require different reverberators, thus increasing the computational complexity significantly. By applying the common-slope model, the resynthesis could be based on a more realistic late reverberation model that avoids overfitting and reduces the number of reverberators and computational cost. The framework could also be applied to an extended reality (XR) audio engine, in which the late reverberation is synthesized for various receiver orientations and positions.

Conclusion

This paper proposed a framework for multi-exponential and anisotropic sound energy decay analysis. A spherical filterbank allows analyzing the energy decay in various directions. The energy decays of all analyzed directions are modeled with one set of common decay rates. By using the common-slope model, the directionality of late reverberation becomes evident. The proposed model has been successfully applied to spatial room impulse response measurements of a concert hall and a coupled room scenario. Apart from room acoustic analysis, the presented framework can also be applied during the synthesis of late reverberation tails.

Acknowledgments

The project has received funding from from the Academy of Finland, project no. 317341, and the Human Optimised XR (HumOR) project.

References

[1] C. F. Eyring, "Reverberation time measurements in coupled rooms," *J. Acoust. Soc. Am.*, vol. 3, no. 2, pp. 181–206, 1931.

[2] F. V. Hunt, L. L. Beranek, and D. Y. Maa, "Analysis of Sound Decay in Rectangular Rooms," *J. Acoust. Soc. Am.*, vol. 11, no. 1, pp. 80–94, 1939.

[3] N. Xiang, J. Escolano, J. M. Navarro, and Y. Jing, "Investigation on the effect of aperture sizes and receiver positions in coupled rooms," *J. Acoust. Soc. Am.*, vol. 133, no. 6, pp. 3975–3985, 2013.

[4] M. R. Schroeder, "New Method of Measuring Reverberation Time," *J. Acoust. Soc. Am.*, vol. 37, no. 3, pp. 409–412, 1965.

[5] N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayeisan decay time estimation," *J. Acoust. Soc. Am.*, vol. 117, no. 6, pp. 3707–3715, 2005.

[6] G. Götz, R. Falcón Pérez, S. J. Schlecht, and V. Pulkki, "Neural network for multi-exponential sound energy decay analysis," *submitted to J. Acoust. Soc. Am.*, 2021.

[7] C. Hold, A. Politis, L. McCormack, and V. Pulkki, "Spatial Filter Bank Design in the Spherical Harmonic Domain," in *29th Eur. Signal Process. Conf. (EUSIPCO)*, (Online conference), pp. 106–110, 2021.

[8] M. Berzborn and M. Vorländer, "Directional sound field decay analysis in performance spaces," *Building Acoustics*, pp. 1–15, 2021.

[9] T. McKenzie, S. J. Schlecht, and V. Pulkki, "Acoustic Analysis and Dataset of Transitions Between Coupled Rooms," in *Int. Conf. Acoust., Speech, Sig. Proc. (ICASSP)*, (Online conference), pp. 481–485, 2021.

[10] C. Hold, T. McKenzie, G. Götz, S. J. Schlecht, and V. Pulkki, "Resynthesis of Spatial Room Impulse Response Tails With Anisotropic Multi-Slope Decays," *submitted to J. Audio Eng. Soc.*, 2022.

[11] H. Kuttruff, *Room Acoustics*. London, UK: Spon Press, 4th ed., 2000.

[12] Y. Haneda, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 709–717, 1999.

[13] T. Jasa and N. Xiang, "Efficient estimation of decay parameters in acoustically coupled-spaces using slice sampling," *J. Acoust. Soc. Am.*, vol. 126, no. 3, pp. 1269–1279, 2009.

[14] P. Stade, B. Bernschütz, and M. Rühl, "A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios," in *27th Tonmeistertagung (TMT)*, (Cologne, Germany), pp. 1–17, 2012.