Deep Learning for Artificial Room Reverberation

A Research Report

By

Chad McKell

Department of Music
University of California San Diego

March 20, 2020

ROUGH DRAFT

1 Artificial Room Reverberation

1.1 Abstract Methods

Reverberation filters (Schroeder reverberators: allpass filters connected in series to set of parallel feedback comb filters)

1.2 Physical Modeling

1.2.1 Geometric Methods

Ray tracing

1.2.2 Wave-based Methods

Detailed 3D models of sound propagation in rooms can be found in various textbooks and research papers [1–4] . Given various assumptions about the sound environment, sound propagation in rooms can be modeled using the 3D wave equation given by [4]

$$\Psi_{tt} = c^2 \Delta \Psi + c\eta \Delta \Psi_t \tag{1}$$

where $\Psi = \Psi(\mathbf{r}, t)$ is the acoustic velocity field in units of m^2/s (or, equivalently, the acoustic pressure field in units of Pa), c is the speed of sound in m/s, and η is a constant that describes the viscous or thermal characteristics of the air.

Given a particular geometric model of a room and a set of boundary conditions for the walls of the room, one may discretize the 3D wave equation above using finite-difference time domain (FDTD) methods to obtain detailed simulations of the acoustic velocity field (or acoustic pressure field) over time. Such direct simulation schemes allow us to make accurate and precise predictions about various reverberation parameters. However, the computational cost of these schemes on even the most advanced modern hardware is often prohibitive, thus restricting the simulations to non-real time applications.

1.3 Neural Audio Synthesis

1.3.1 Prior Work

Using deep learning schemes, we aim to predict reverberation parameters with higher computational efficiency than physical modeling and with sufficient accuracy and precision. Prior studies have used deep learning algorithms to model reverberation in noisy audio for purposes of source localization [5], speech recognition [6, 7], and sound design [8]. While closely related, our goal is to use deep learning to add physically accurate room reverberation to a dry signal. That is, the resulting reverberation must accurately match the reverberation expected in a particular physical room. This deep learning-based technique of room reverberation simulation falls within the category of neural audio synthesis.

Researchers have used various deep learning schemes to model reverberation, including convolutional neural networks [9].

1.3.2 Training Data

The training data used for deep learning consists of physical features of the room and corresponding reverberation parameters. The reverberation parameters we aim to predict in testing scenarios include the direct sound travel time, early and late reflection travel times, and the reverberation time. Figure 1 illustrates the relationship between early and late reflections as well as the salient features of the room impulse response.

The reverberation time is often described as the time required for reflections of a direct sound to decay by 60 dB relative to the direct sound, in which case it is denoted as T_{60}. This value can be approximated mathematically using the Sabine equation shown below.

$$T_{60} = 0.161 \frac{V}{S\alpha + mV} \tag{2}$$

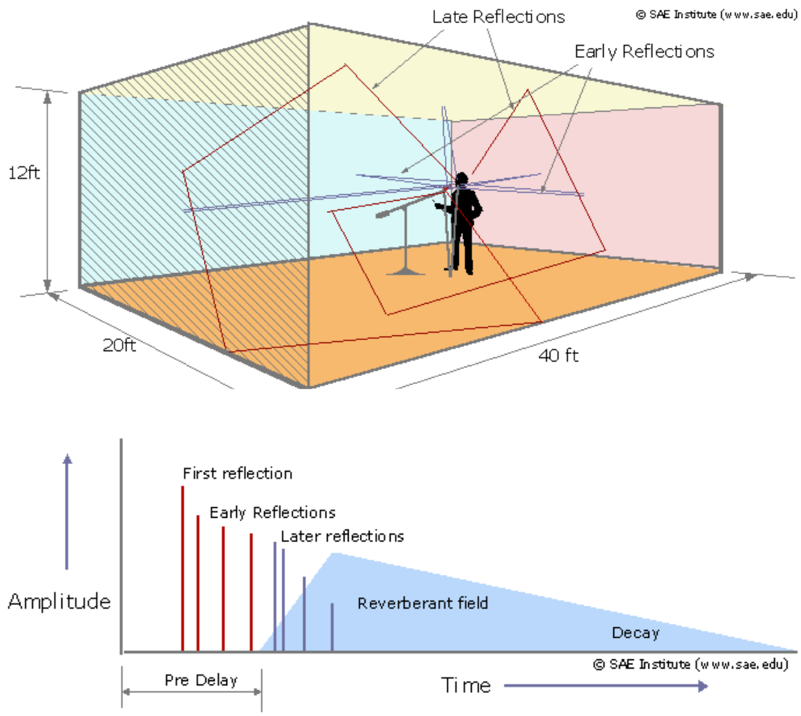

Figure 1: Top: Diagram of the early and late reflections in a room. Bottom: Illustration of the room impulse response. Note that the reverberation field is shown to have both an attack time and a decay time. Copyright SAE Institute (www.sae.edu).

Here, V is the total volume of the room, S is the total surface area of the room, $\alpha = \Sigma w_i \alpha_i$ is the composite absorption coefficient of the interior room surfaces, and m is a constant describing the absorption in air. Given the reverberation time T_{60} of the room, we can compute the corresponding impulse response of the room $h(t)$, then convolve $h(t)$ with a dry input signal to create a simulation of the room reverberation. To speed up the computation of the reverberation, we perform convolution as a multiplication in the frequency domain when the sample size of the input signal is sufficiently large[1]. This increase in speed is due to the fact that convolution in the time domain scales as $\mathcal{O}(n^3)$ while convolution as multiplication in the frequency domain scales as $\mathcal{O}(n \log n)$ [8].

[1]When the sample size of the input signal is small, it is preferable to perform convolution in the time domain because—unlike convolution in the frequency domain—it does not require that we compute the Fourier transform.

References

[1] P. M. Morse and K. U. Ingard, *Theoretical acoustics.* Princeton university press, 1986.

[2] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, "Fundamentals of acoustics, john wiley & sons," *Inc, New York*, 2000.

[3] S. Bilbao, B. Hamilton, J. Botts, and L. Savioja, "Finite volume time domain room acoustics simulation under general impedance boundary conditions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 1, pp. 161–173, 2015.

[4] B. Hamilton, "Finite difference and finite volume methods for wave-based modelling of room acoustics," 2016.

[5] S. Chakrabarty and E. A. Habets, "Broadband doa estimation using convolutional neural networks trained with noise signals," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 136–140, IEEE, 2017.

[6] K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, E. Habets, R. Haeb-Umbach, V. Leutnant, A. Sehr, W. Kellermann, R. Maas, *et al.*, "The reverb challenge: A common evaluation framework for dereverberation and recognition of reverberant speech," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–4, IEEE, 2013.

[7] R. Giri, M. L. Seltzer, J. Droppo, and D. Yu, "Improving speech recognition in reverberation using a room-aware deep neural network and multi-task learning," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5014–5018, IEEE, 2015.

[8] J. Engel, L. Hantrakul, C. Gu, and A. Roberts, "Ddsp: Differentiable digital signal processing," *arXiv preprint arXiv:2001.04643*, 2020.

[9] H. Gamper and I. J. Tashev, "Blind reverberation time estimation using a convolutional neural network," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 136–140, IEEE, 2018.