Loudness-Based Display and Analysis Applied to Artificial Reverberation

Patty Huang and Julius O. Smith III

Center for Computer Research in Music and Acoustics, Stanford University

pph@ccrma.stanford.edu, jos@ccrma.stanford.edu

Abstract

In this paper, we propose a psychoacoustically motivated method to analyze the quality of reverberation and other audio signals. A time-varying loudness model is used as a front end to produce a visual display of a signal which emphasizes its perceptual features. A time-frequency display of reverberation impulse responses based on specific loudness is shown to produce a psychoacoustically relevant visualization of response features. In addition, a metric based on instantaneous loudness is proposed as an objective measure of quality and texture of late reverberation.

1   Introduction

Reverberation has traditionally been described in terms of acoustical measures for concert halls and rooms. These include reverberation time in octave bands, clarity, initial decay time, interaural cross-correlation coefficient, and center time, among others (Beranek 1996; ISO 1997). However, these metrics, whose values are effective discriminators of hall preference and usage (e.g., chamber music, opera, lecture, etc.) do not directly illuminate the finer details of a reverberant impulse response, such as the modal density, rate of echo growth, or the texture of the late reverberation.

Especially in artificial reverberation applications, the finer structure of reverberation can become important not only for descriptive purposes, but as something which can be manipulated by users to produce a wider range of reverberant effects. The majority of digital reverberators try to produce "sufficiently" or "maximally" dense reflections (often resulting in a pleasingly generic reverberant tail) and then allow user control over equalization, early reflections (pattern, level, and timing), and the late reverberation (level, timing, and decay time over several frequency regions). Though the temporal and modal structure of the reverberation is often tied to the algorithm used and is not parameterized, some programs do offer control over diffusion or density, for example, with varying results.

We have two goals toward gaining further insight into the character of a reverberant response. The first is visual display of a reverberant impulse response which highlights (psychoacoustically) important features of the signal, and the second is to develop metrics for describing and quantifying the temporal quality of the reverberation. For the first task, we propose a time-frequency display based on specific loudness (Zwicker and Fastl 1999). For the second task, we tackle a smaller portion of the problem, and focus on developing an objective metric which helps to distinguish between different subjective late reverberation textures.

These methods are perceptually motivated in order to best match the way humans hear. Though other auditory models may give similar results, we use Glasberg and Moore's time-varying loudness model (Glasberg and Moore 2002) as the front end for analyzing a reverberant response.

2   Specific Loudness Time-Frequency Display of Reverberation

2.1   Displays for reverberant impulse responses

A classical way to visualize a reverberant response is to plot its *energy-time curve*, which is squared amplitude versus time, as shown in Fig. 1 (top). The logarithmic scaling in the energy-time curve (see Fig. 1, top) conforms better to human hearing than linear amplitude scaling, and emphasizes the general structure of the response, where typically a few early reflections develop into a late reflections/early reverberation region and then into the late reverberant field (usually seen as a nearly uniform linear decay on a log amplitude scale). The *Energy Decay Curve* (EDC), developed by Schroeder (1965), computes, for each time t, the integral with respect to time of all signal energy after time t (Fig. 1, bottom). It thus shows very clearly the overall decay rate versus time. The decay rate of the reverberation can be estimated at any time by fitting a straight line to the log-EDC about that time. One can also see if there is a two-stage decay and how smoothly (ideally exponentially) it decays.

Frequency information, however, is needed to give a more complete picture of the response characteristics. The classic spectrogram produces a time-frequency display, but the
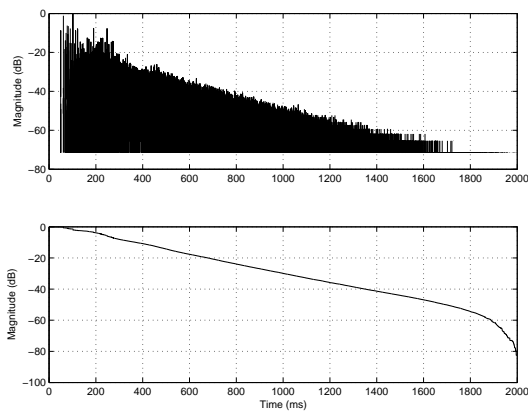
Figure 1: Energy-time curve (top) and energy decay curve (bottom) of an artificial reverberation impulse response, normalized to 0 dB.

choice of window size forces a tradeoff between capturing larger-scale features and getting a good time resolution. In addition, the noise-like nature of a reverberant signal results in neighboring bins which vary considerably in magnitude, visually complicating what may simply be heard by the ear as a uniform, smooth decay. The *Energy Decay Relief* (EDR), developed by Jot (1992), is the generalization of the EDC to multiple frequencies: $EDR(t_n, f_k) = \sum_{m=n}^{M} |H(m,k)|^2$, where $H(m,k)$ is the short-time Fourier transform value of the reverberant response $h(t)$ at time frame m and frequency bin k. The EDR provides an excellent basis for the measurement of reverberation time at each frequency. The integration process, however, smooths away much of the temporal information, so that these features are obscured.

Lokki and Karjalainen (2002) have proposed a visualization and analysis method for room responses based on auditory perception. After applying a level sensitivity filter, the response is fed into a gammatone filterbank to divide the signal into equivalent rectangular bandwidth (ERB) bands, compressed, and then filtered with an integrating window which simulates forward and backward temporal masking. The resulting time-frequency plot is made by decompressing the final result and plotting on a dB scale. The psychoacoustic frequency scale and the temporal smoothing result in a display which corresponds more closely to auditory perception, and whose values may be used in other applications such as analyzing the directional characteristics of sound fields in time and frequency.

2.2 Specific loudness spectrogram

A visualization method in which one can see what the ear hears would be a valuable tool for evaluating the quality and recognizing the salient features of reverberation. A psychoacoustic model can best represent this, and we apply the time-varying loudness model of Glasberg and Moore (2002) to pro-

duce a time-frequency plot of specific loudness, or a "loudness spectrogram". Their model was chosen for its applicability to a wide range of sounds, validation against multiple experimental datasets, nearly literal interpretation of loudness across time and frequency, known relationship to real-world units (e.g., SPL, sones, phons), and good time-domain resolution.

Calculating loudness The calculation of loudness is based on three recent papers (Glasberg and Moore 1990; Moore, Glasberg, and Baer 1997; Glasberg and Moore 2002) and implemented in Matlab. The specific loudness, or loudness per ERB, is a loudness measure in time-frequency. Briefly, the steps are the following:

- The signal is filtered, simulating transmission through the outer and middle ear.

- A multi-resolution short-time Fourier transform is computed, using longer windows for lower frequencies and shorter windows at higher frequencies to approximate the time-frequency resolution of the auditory system (Fig. 2). (Figs. 1 to 4 in this paper are derived from this same artificial reverberation sample.)
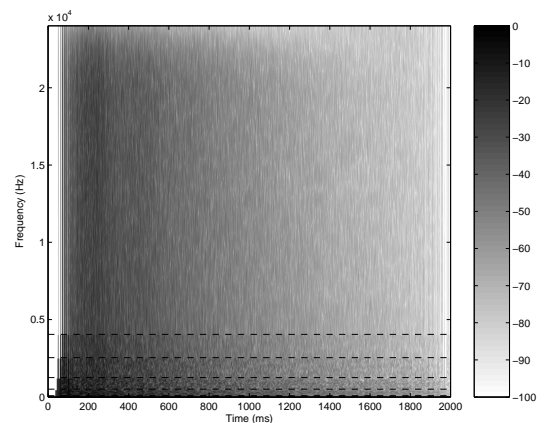

Figure 2: Multi-resolution short-time Fourier transform of an artificial reverberation impulse response, normalized to 0 dB max. The horizontal dashed lines mark the frequencies at which the STFT resolution changes. The gray-level scaling is in dB.

- The excitation pattern is calculated at 0.25 ERB intervals from the multi-resolution STFT, using rounded-exponential auditory filters whose shapes vary with center frequency and level. This results in frequency smoothing.

- A compressive nonlinearity applied to the excitation pattern, simulating compression that occurs in the cochlea, transforms it to specific loudness (Fig. 3).
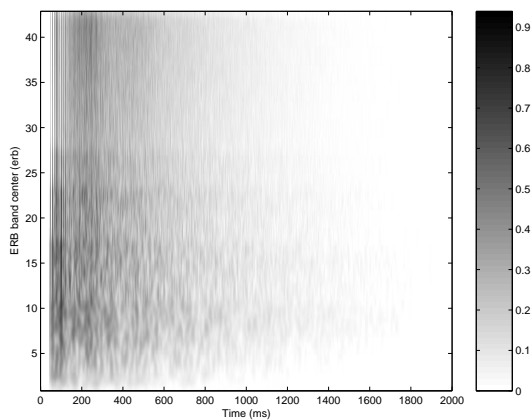
Figure 3: Specific loudness of an artificial reverberation impulse response. Note that the frequency axis is on an ERB scale. The gray-level scaling is in sones/erb.

- The overall instantaneous loudness in units of sones is computed by summing the specific loudness across all ERB bands and then normalizing according to the amount of ERB band overlap. Fig. 4 shows the instantaneous loudness in sones and in phons (Rossing, Moore, and Wheeler 2002).
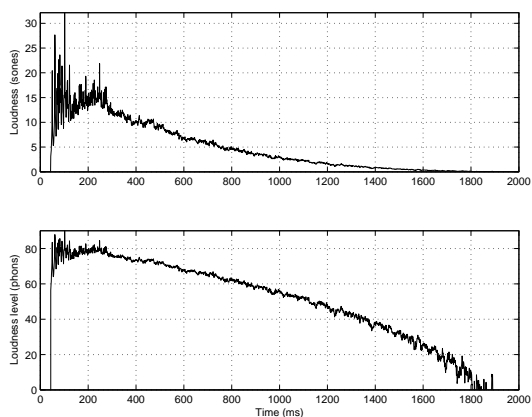

Figure 4: Instantaneous loudness of an artificial reverberation impulse response, in sones (top) and in phons (bottom).

- The single-valued short-term loudness and long-term loudness can be derived, if desired, by smoothing the instantaneous loudness over time, as detailed in (Glasberg and Moore 2002).

Display of time-varying sounds The artificial reverberation sample represented in the plots in this article (Figs. 1 to 4) is heard as having two closely-spaced "events" and then a uniform decay where all frequencies decay at nearly the same rate. The first "event" is a group of crackly bursts, while the second "event" is an accent at the start of the late reverb – a sort of splash that the late decay grows out of. There are two dark regions (indicating acoustic patterns), one centered around 100 ms and another broader region centered around 250 ms (Fig. 3). These are the two "events".

The early reflections are especially visible at high frequencies, since each distinct reflection contains much high-frequency content. Due to the higher time resolution, the pattern appears as alternating dark and light vertical stripes. In the late reflection/early reverberation region (approximately 150 ms to 300 ms), vertical striations are still visible but are not as sharply defined. In the late field, the striations disappear and remaining loudness pattern has no distinguishable features. The slightly darker horizontal bands which run across the specific loudness spectrogram are an artifact of using a multi-resolution STFT to calculate the short-term spectrum. This can be remedied by frequency warping the (filtered) signal to conform to an ERB scale and then applying a fixed-resolution STFT, instead of taking separate STFTs at different resolutions.

It should be noted that the specific loudness spectrogram would also be very useful for perceptual display of other types of sounds. For example, a loudness spectrogram of a speech sample would emphasize the formant structure and main harmonics at low frequencies, and time-domain pitch-pulses at high frequencies (indicating glottal closure instants). Also, the relative loudness of different time and frequency regions can be observed, as well as the audibility of the speech at different sound levels. Additionally, time-frequency loudness can function as the basis for a psychoacoustical error criterion comparing an original signal with its synthesized, modeled, or compressed counterpart.

3 Loudness-Derived Objective Metric for Late Reverberation Texture

Late reverberation is often described and simulated as filtered, exponentially decaying white Gaussian noise. The filtering and decay characteristics result from air propagation losses and absorption from room surfaces. The Gaussian amplitude distribution arises when the dense superposition of reflections approximates a random sum of plane waves from all directions (the diffuse field). In artificial reverberation algorithms, however, the late reverb sometimes does not reach a smooth, nondescript Gaussian noise-like quality due to an insufficient echo density or poor temporal distribution of echoes, resulting in a somewhat grainy textured decay.

Reflections which stand out from the decaying tail or regions which are temporally more sparse will affect the time-frequency distribution. The instantaneous loudness values can be considered to be a psychoacoustically based summary, collapsed in frequency but having fine resolution in the time domain. How much the instantaneous loudness values vary in time will give a perceptually weighted measure of how much the frequency content has changed.

3.1 Metric calculation

The metric proposed for describing and quantifying the late reverberation temporal quality is the standard deviation (in phons) of the instantaneous loudness decay during late reverb. To calculate this, instantaneous loudness values (in phons) are computed over a length of time during the late reverberation portion of the impulse response. The loudness decay rate (slope) is estimated using least squares, and this linear trend is then subtracted from the instantaneous loudness values. The standard deviation is calculated from these normalized loudness values (see Fig. 5, right column).
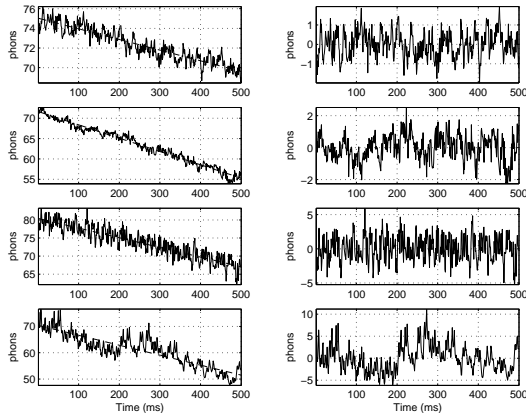

Figure 5: Instantaneous loudness of the four corresponding artificial reverberation samples as listed in Table 1. Left column: instantaneous loudness (solid line) and the estimated linear trend representing loudness decay (dotted line) during a 500 ms period of late reverb. Right column: instantaneous loudness values with the linear trend subtracted off.

3.2 Simulation results

The late reverberation of four different artificial reverb samples were analyzed. The first three samples are taken from commercial reverb plug-ins, while the fourth sample is an order 4 feedback delay network (Jot and Chaigne 1991) with the delay line lengths chosen to produce a somewhat "pathological" reverb tail. (It is not known by the authors what algorithms were used for the commercial plug-ins.) The second artificial reverberation sample is the one used for the other plots in this paper. The analysis results are shown in Table 1, and the standard deviation values correlate strongly with the degree of textural graininess.

4 Conclusions and Future Work

A time-frequency display of reverberation impulse responses based on specific loudness was shown to produce a psychoacoustically relevant visualization of response features. In

artificial reverb sample	qualitative description	standard deviation
1	smooth, transparent	0.6624
2	smooth, sandy	0.7965
3	bit crackly, rougher sandy	1.8527
4	crackly	2.8967

Table 1: Calculated standard deviation of instantaneous loudness decay (in phons) during late reverberation. The artificial reverberation samples are listed in order of increasingly grainy reverb tails.

addition, a metric based on instantaneous loudness has shown potential as an objective measure of temporal quality of late reverberation. Future work includes testing of the metric with a larger reverberation sample set, and conducting empirical listening experiments. The model will be developed further to help classify and to provide an accurate prediction of the perceived texture of late reverberation.

References

Beranek, L. (1996). *Concert and Opera Halls: How They Sound.* Woodbury, NY: Acoustical Society of America. Note that a 2003 edition is now available.

Glasberg, B. R. and B. C. J. Moore (1990). Derivation of auditory filter shapes from notched noise data. *Hearing Research 47*, 103–138.

Glasberg, B. R. and B. C. J. Moore (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society 50*(5), 331–342.

ISO (1997). ISO 3382. acoustics – measurement of reverberation time of rooms with reference to other acoustical parameters.

Jot, J.-M. (1992). An analysis/synthesis approach to real-time artificial reverberation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Volume 2, San Francisco, CA, pp. 221–224.

Jot, J.-M. and A. Chaigne (1991). Digital delay networks for designing artificial reverberators. In *The 90th Convention of the Audio Engineering Society*, Paris, France. preprint 3030.

Lokki, T. and M. Karjalainen (2002). Analysis of room responses, motivated by auditory perception. *Journal of New Music Research 31*(2), 163–169.

Moore, B. C. J., B. R. Glasberg, and T. Baer (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society 45*(4), 224–240.

Rossing, T. D., F. R. Moore, and P. A. Wheeler (2002). *The Science of Sound* (Third ed.). San Francisco: Addison Wesley.

Schroeder, M. R. (1965). New method of measuring reverberation time. *Journal of the Acoustical Society of America 37*, 409–412.

Zwicker, E. and H. Fastl (1999). *Psychoacoustics: Facts and Models* (Second ed.). Berlin: Springer Verlag.