Audio Engineering Society

Convention Paper

Presented at the 121st Convention
2006 October 5–8 San Francisco, CA, USA

A Simple, Robust Measure of Reverberation Echo Density

Jonathan S. Abel[1,2] and Patty Huang[2,3]

[1] *Universal Audio, Inc., Santa Cruz, CA, 95060, USA*

[2] *Center for Computer Research in Music and Acoustics (CCRMA), Stanford Univ., Stanford, CA, 94305, USA*

[3] *Laboratory of Acoustics and Audio Signal Processing, Helsinki Univ. of Tech., Espoo, FI-02015 TKK, Finland*

Correspondence should be addressed to Jonathan S. Abel (`abel@uaudio.com`)

ABSTRACT

A simple, robust method for measuring echo density from a reverberation impulse response is presented. Based on the property that a reverberant field takes on a Gaussian distribution once an acoustic space is fully mixed, the measure counts samples lying outside a standard deviation in a given impulse response window and normalizes by that expected for Gaussian noise. The measure is insensitive to equalization and reverberation time, and is seen to perform well on both artificial reverberation and measurements of room impulse responses. Listening tests indicate a correlation between echo density measured in this way and perceived temporal quality or texture of the reverberation.

1. INTRODUCTION

Sound radiated in a reverberant environment will interact with objects and surfaces in the environment to create reflections. These propagate and subsequently interact with additional objects and surfaces, creating even more reflections. Accordingly, an impulse response measured between a sound source and listener in a reflective environment will record an increasing arrival density over time. After a sufficient period of time, the echo density will be so great that the arriving echoes may be treated statistically, and the impulse response would arguably be indistinguishable from Gaussian noise with an evolving color and level [1].

The rate of echo density increase has to do with factors such as the size and shape of the space and whether it is empty or cluttered with reflecting objects. Although it has not been widely studied in the literature, echo density and its time evolution are commonly thought to influence the perceived time-domain timbre of a space, with a large, rapidly increasing echo density being desirable.

In [2], the variance over time of reverberation impulse response energy (which can be indicative of echo density) was seen to correlate with the subjective texture of late reverberation. In another recent work [3], temporal correlations in reverberation impulse response frequency bands were shown to relate to the "temporal color" of the reverberator. The idea was that noise-like segments of a reverberation impulse response would maintain small autocorrelation sidelobes compared to their peak values, whereas segments with undesirable periodicities would have relatively larger sidelobes.

In the context of developing an artificial reverberator based on pseudorandom sequences of echoes [4, 5], Rubak and Johansen cited literature and presented listening tests suggesting that reverberation impulse responses maintaining echo densities greater than about four thousand echoes per second may be treated statistically and sound similar to one another, independent of their actual echo densities.

Echo or reflection density has more commonly been expressed in terms of the number of reflections per second. However, this measure is not always meaningful because it is unclear how to best define a reflection, and sampling rate becomes a limiting factor in counting reflections. As an estimate of reflection density in a given synthetic reverberator impulse response window, Griesinger suggested counting echoes within 20 dB of the window maximum [6, pg. 192]. This method was used to measure echo density in [7], where an echo was defined as an impulse response peak—impulse response taps larger in absolute value than either of their neighbors. One difficulty with this approach is that any given reflection can produce a large number of impulse response peaks, depending on the details of its propagation as well as factors such as impulse response equalization [6, pg. 188]. In addition, the level of the loudest tap, and therefore the number of taps considered reflections, is sensitive to the phase of its associated reflection and other characteristics which are not likely psychoacoustically meaningful.

In this paper, the problem of measuring echo density as a function of time is considered. The approach taken follows from the fact that once the room or reverberator is sufficiently mixed, the impulse response taps take on a Gaussian distribution, irrespective of the actual reflection density. Over a sliding impulse response window, the echo density measure presented here simply counts the number of taps outside the standard deviation for the window, and normalizes by that expected for Gaussian noise.

Reflections arriving separated in time or in level skew the tap histogram away from its limiting Gaussian form, and the percentage of taps outside a standard deviation is a computationally simple indicator of this skew. In this way, the measure is normalized, producing values on a scale from zero to around one. The presence of a few prominent reflections results in a low echo density value, since they contribute to a larger standard deviation and consequently a smaller number of samples classified as outliers. By contrast, an extremely dense pattern of overlapping reflections approximating Gaussian noise will produce a value near one.

The proposed echo density measure is described in §2. The performance of the measure is examined in §3 by computing echo density profiles of room impulse response measurements and artificial reverberation generated by a feedback delay network reverberator with various amounts of state mixing. §4 concludes the paper with a discussion about properties of the echo density measure and how it may be useful in describing the temporal quality of reverberation.

2. ECHO DENSITY PROFILE

2.1. Echo Density Measure

Over a sliding reverberation impulse response window, the proposed echo density measure $\eta(t)$—called here the *echo density profile*—is the fraction of impulse response taps which lie outside the window standard deviation:

$$\eta(t) = \frac{1/\mathrm{erfc}(1/\sqrt{2})}{2\delta+1} \sum_{\tau=t-\delta}^{t+\delta} \mathbf{1}\{|h(\tau)| > \sigma\}, \tag{1}$$

where $h(t)$ is the reverberation impulse response (assumed to be zero mean), $2\delta+1$ is the window length in samples, σ is the window standard deviation,

$$\sigma = \left[\frac{1}{2\delta+1} \sum_{\tau=t-\delta}^{t+\delta} h^2(\tau)\right]^{\frac{1}{2}}, \tag{2}$$

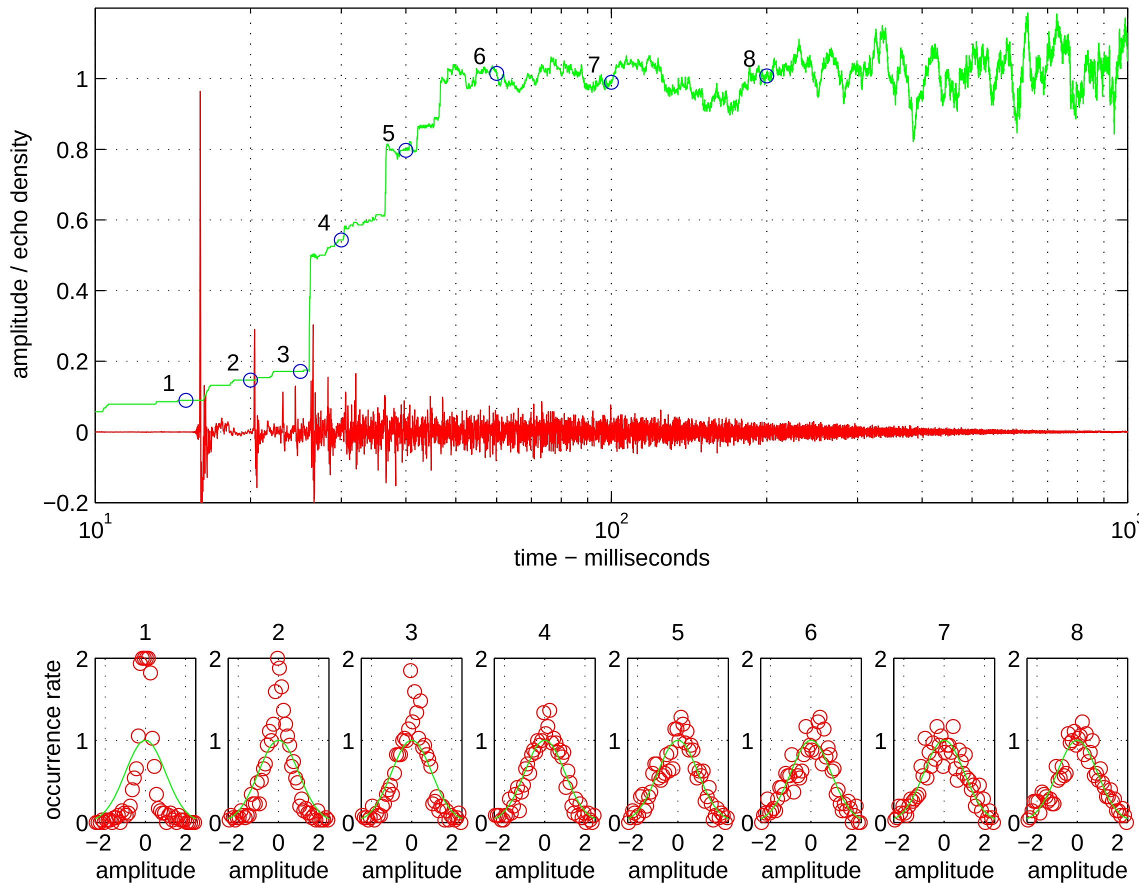

Fig. 1: Top: Measured room impulse response (red, near bottom, 44.1 kHz samples) and its echo density profile (green, near top). Note that the time axis is on a logarithmic scale. Bottom: Window histograms centered at times indicated by the blue circles in top plot. The Gaussian distribution is indicated by a green line. Normalized occurence rates greater than 2.0 are plotted at this value.

$\mathbf{1}\{\cdot\}$ is the indicator function, returning one when its argument is true and zero otherwise, and $\mathrm{erfc}(1/\sqrt{2}) \doteq 0.3173$ is the expected fraction of samples lying outside a standard deviation from the mean for a Gaussian distribution.

The echo density profile is more generally computed using a positive weighting function $w(t)$ so as to de-emphasize the impulse response taps at the sliding window edges:

$$\eta(t) = \frac{1}{\mathrm{erfc}(1/\sqrt{2})} \sum_{\tau=t-\delta}^{t+\delta} w(\tau)\mathbf{1}\{|h(\tau)| > \sigma\} \quad (3)$$

with

$$\sigma = \left[\sum_{\tau=t-\delta}^{t+\delta} w(\tau)h^2(\tau) \right]^{\frac{1}{2}} \quad (4)$$

and where $w(t)$ is normalized to have unit sum $\sum_\tau w(\tau) = 1$.

Consider the impulse response shown in Figure 1, measured in the lobby of the Knoll, a building at Stanford University. (This impulse response is also used in Figures 2–5.) The direct path is evident, as are a number of early reflections giving way to a noise-like late field response. Also shown is the echo density profile, $\eta(t)$, computed according to (1) using a 20 ms long sliding window, and tap histograms for several of the windows used in computing the echo density profile.

The echo density profile starts near zero and increases over time to around one, where it remains for the duration of the impulse response. This is typical of echo density profiles of measured impulse responses and high-quality synthetic reverberators. As we will see later, what sets one profile apart from another is the rate of increase and the time at which a value of one is first attained, interpreted here as the start of the late field.

The echo density profile is mirrored in the window histograms, which are calculated at 5, 10, 15, 20, 30, 50, 90, and 190 ms. For windows at the beginning of the impulse response, the histograms are peaked at small tap values compared to the Gaussian, indicating the concentration of the window energy in a relatively small number of taps. After the echo density profile achieves a value of one, the histograms take on a Gaussian form, consistent with the notion that the late field has started.

2.2. Window Length and Weighting

The behavior of the echo density profile $\eta(t)$ is affected by the choice of sliding window length and weighting function $w(t)$.

Figure 2 shows echo density profiles of a measured room impulse response calculated using various window lengths and a Hanning window weighting function. While the gross shape of the profiles are similar, there are some important differences.

A short window length naturally includes few impulse response taps and thus is expected to have a relatively high variance about its local mean. This is evident in the profiles of Figure 2, which exhibit a variance decreasing with increasing window length. In the late field, with the echo density profile varying about one, the profile variance will decrease roughly in proportion to window length.

While using a shorter window leads to a larger echo density profile variance, the profile produced will be more responsive to short-term echo density changes. However, the window should not be made so short that it does not contain any reflections. If so, the profile will contain jumps which are not meaningful. A window length of 20–30 ms works well, since it

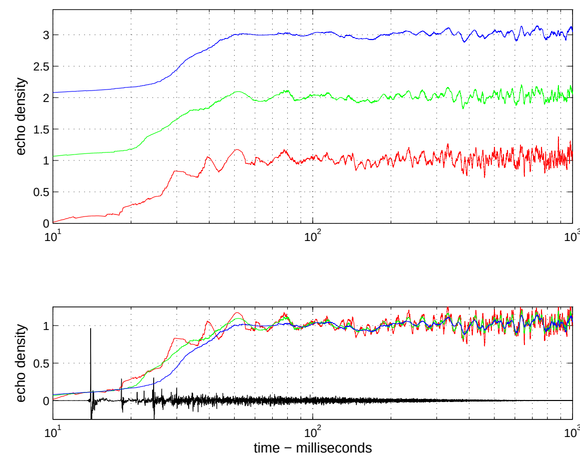

Fig. 2: Bottom: Echo density profiles of a measured room impulse response (black) using different window lengths—10 (red), 20 (green), and 40 (blue) ms. Top: Echo density profiles plotted in offset (from bottom to top, respectively).

is long enough to provide good statistics and contain at least a few reflections within each window, while being short enough to maintain sufficient time resolution for psychoacoustic purposes.

The length of the echo density profile window may even be varied over time in either a predetermined fashion or as a function of a previously computed echo density profile having a fixed window length. The idea would be to shorten the window at the beginning of the impulse response where the echo density is smaller and more quickly changing, and lengthen the window in the late-field where the echo density is relatively constant near one.

As seen in Figure 1, individual large echoes, such as those occurring at the impulse response onset, can cause abrupt changes in the echo density profile as they become included in the sliding window. In order to smooth these jumps, a weighting function emphasizing samples at the window center may be used.

In Figure 3, an echo density profile calculated using a Hanning window in (3) is plotted along with a profile calculated using a rectangular window, (1). The rectangular window profile is much more uneven and stepped than that using the Hanning weighting,

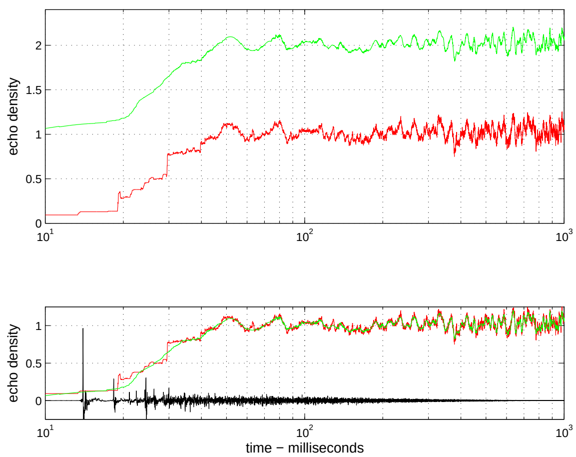

Fig. 3: Bottom: Echo density profiles of a measured room impulse response (black) using different weighting functions—rectangular window (red) and Hanning window (green). Top: Echo density profiles plotted in offset.

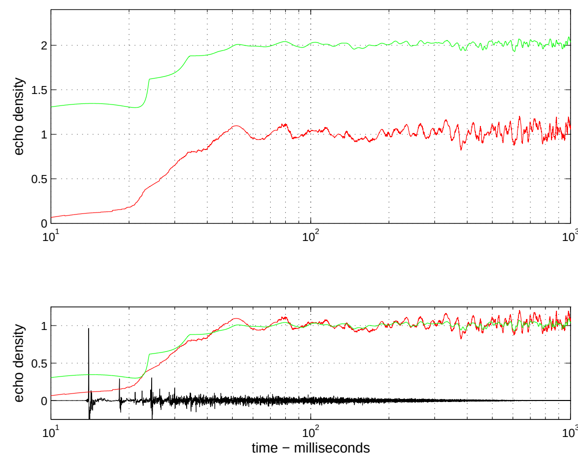

Fig. 4: Bottom: Echo density profiles of a measured room impulse response (black) using different statistical criteria—outlier-based (red) and kurtosis-based (green). Top: Echo density profiles plotted in offset.

which is seen to smoothly interpolate the "steps." Note that the smooth tapering of the Hanning window results in an effective window length of about half of the actual length. This is seen in Figure 3, where the Hanning-weighted profile used a window length of 20 ms and the rectangular-weighted profile used a window length of 10 ms.

2.3. Similar Measures

Echo density measures can also be formed based on other statistics indicating an impulse response segment's similarity to a Gaussian process. For instance, a measure related to the kurtosis [8], which describes the peakedness of a sample distribution, is a good example of such a statistic. A higher kurtosis value results when a relatively small number of taps contribute to the majority of the variance. This is what is expected during the early portion of a reverberation impulse response when the reflections are more sparse.

In the same sliding window, an alternative echo density measure $\eta_k(t)$ computes the fourth root of the fourth moment of the window taps, normalized by that expected for a Gaussian having the same variance:

$$\eta_k(t) = \frac{\sigma}{\left[\frac{1}{3}\sum_{\tau=t-\delta}^{t+\delta} w(\tau)h^4(\tau)\right]^{\frac{1}{4}}}, \tag{5}$$

where σ is defined as before, (4).

Figure 4 compares two echo density profiles using a 20 ms window with Hanning weighting, one defined by (3) and another related to the kurtosis. Though the kurtosis measure $\eta_k(t)$ has less variance in the late field (just as the median has larger variance than the mean), it does not perform as well when the echo density is small. Using the kurtosis measure, windows sliding over reflections often result in a false peak in the echo density profile. Furthermore, the measure based on the percentage of outliers seems to have a better dynamic range than that based on fourth moments.

Another measure involving the fourth moment is to form a mixture model for the window of impulse response taps, where each tap is either drawn from a Gaussian or is the sum of a sample from the Gaussian and a large-variance distribution. The idea is to estimate the probability that just the Gaussian is used. A measure very similar to $\eta_k(t)$ results.

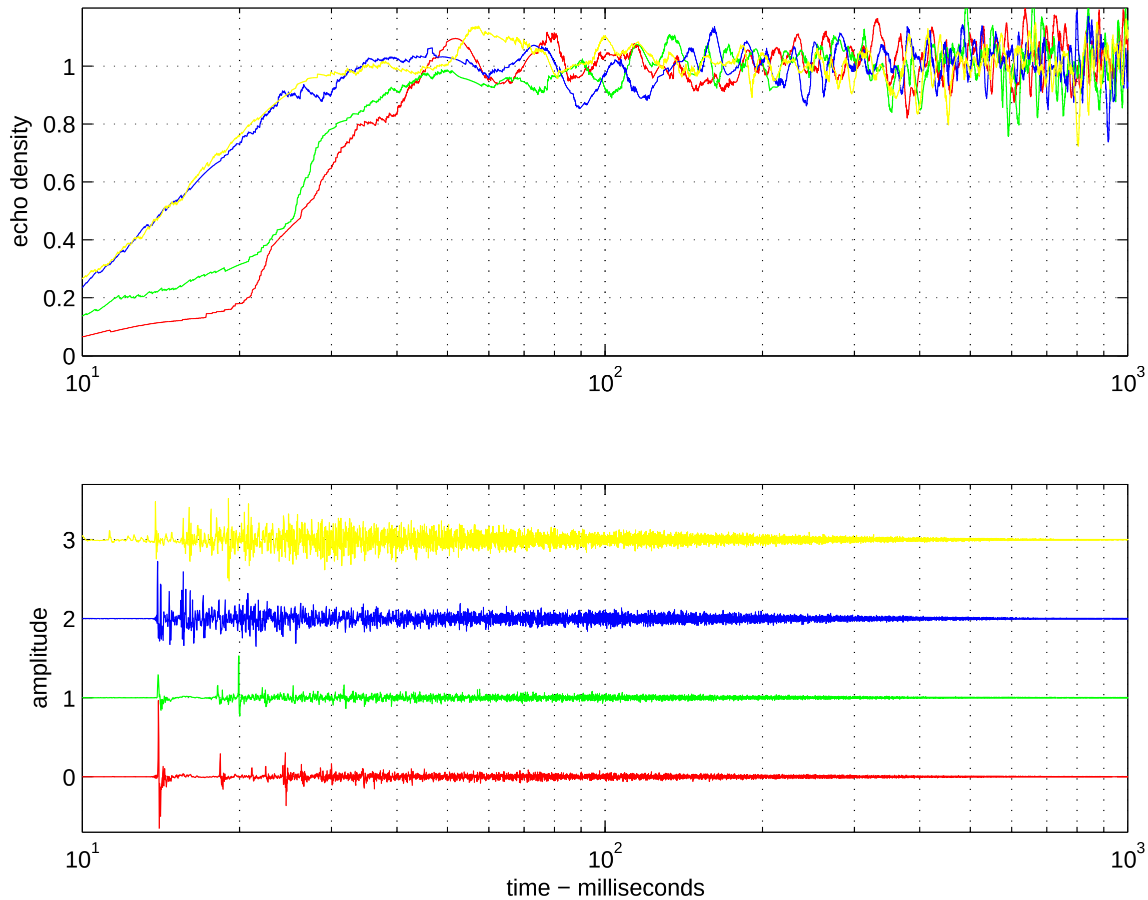

Fig. 5: Bottom: Measured room impulse responses of a lobby (red, green) and a hallway (blue, yellow) having different source and receiver placements, plotted respectively from bottom to top in offset. Top: Corresponding echo density profiles of the impulse responses below.

3. PERFORMANCE

The performance of the proposed echo density measure $\eta(t)$ was analyzed by calculating echo density profiles for a variety of different measured and artificial reverberation impulse responses. The following examines how well these profiles correlate to desired and/or expected behavior and how the echo density measure may be interpreted.

3.1. Measured Room Reverberation

A set of room impulse responses was studied. There were four impulse responses, two measured in each of two acoustical spaces—a lobby and a hallway. The locations and directional orientations of the speaker and microphone varied for all measurements. These measured impulse responses and their respective echo density profiles (window length 20 ms, Hanning weighting) are shown in Figure 5.

All four echo density profiles start from a value near zero and increase monotonically until leveling off around one, as is typical for a reverberation impulse response. Even though the impulse responses are visually very different, the profiles calculated for impulse responses measured in the same room closely track one another. This supports the property that the rate of increase of echo density in a particular room is only dependent on its volume and shape and thus would not be noticeably affected by any differences in source/microphone placement. Impulse responses measured in the same room should have similar echo density profiles and reach the late field at the same time.

The differences expected in the rate of echo density increase between the lobby and hallway impulse responses are also evident in their respective echo density profiles. The long and narrow hallway with several small alcoves supports many reflections, but its long length prevents the space from mixing efficiently. Its corresponding echo density profile has a higher initial echo density value but a relatively slower echo density increase. The spacious lobby takes more time to build up echoes due to the higher ceiling and longer distances between its closest walls, but soon the echo density increases considerably, as can be seen by the steeper echo density profile slope between 20 ms and 50 ms.

The number of sound source arrivals in an enclosed, empty space may be estimated by comparing the room volume to the volume swept out by sound radiating from the source in free-space [9, pp. 252–253]. In this way, echo density is seen to increase as time squared in three-dimensional rooms. A reflection rate on the order of 2000–4000 echoes/s produces a statistical late field [5, 10], leading to a 25–35 ms mixing time for the 150 cubic meter lobby. Indeed, the echo density profiles hit their late-field values of one about 30 ms after the direct path arrival.

3.2. Artificial Reverberation

Artificial reverberation impulse responses were generated using various diffusion settings, reverberation times (t_{60}), delay line lengths, and equalizations in order to test how the echo density measure performs on artificial reverberation impulse responses and how sensitive it is to different reverberation features. A feedback delay network (FDN) structure [11] was used to produce the impulse responses, as it allows parameters to be controlled independently. Low-order filters were designed to give the desired reverberation times and equalization contours. Diffusion was controlled via the feedback matrix, which was set to the identity matrix at zero diffusion and a Hadamard matrix at the maximum diffusion value of one.

Some of these impulse responses are plotted in Figures 6 and 8. For a given decay rate and equalization, the diffusion is varied from low (0.2) to high (1.0). Though the responses all begin the same way with a few sparse reflections, the effect of the diffusion control is immediately apparent. The response with the lowest diffusion setting (red) has sparse reflections visible throughout the majority of its response, while the response with the highest diffusion setting (yellow) quickly builds up echoes in the beginning and individual reflections are not visible past 100 ms. The responses with medium diffusion values (green and blue) have intermediate transitions regarding the buildup of echoes and the disappearance of large, individual reflections. The corresponding spectrograms in Figures 6 and 8 show that the energy and frequency-dependent time evolution within each set of four responses are the same—the only difference between these responses is the amount of diffusion, which affects the time-domain characteristics of the reverberation.

Figures 7 and 9 plot the echo density profiles corresponding to the impulse responses shown in Figures 6 and 8, respectively. A 26.7 ms long window was used and the weighting function was a rectangular window which tapers off at each side with half a Hanning window. (The Hanning halves together are the same length as the rectangular window component.) This window is roughly equivalent to a 20 ms long rectangular window, but produces a smoother echo density profile.

The echo density measure is able to discriminate well between different diffusion settings. A higher diffusion value sets the feedback matrix to mix more fully among the FDN's delay lines. More echoes are generated, and this leads to a quicker arrival at a Gaussian distribution in the impulse response. This can be seen clearly in Figures 7 and 9—the larger the diffusion setting value, the faster the corresponding echo density profiles reach full density.

Another important feature of the echo density measure is that it is robust with respect to a wide range of reverberation times and different equalizations. This can be seen by comparing the echo density profiles associated with the same diffusion setting in Figures 7 and 9. They are essentially the same, even though there is a significant difference in the timbre—one has a short decay time and a highpass

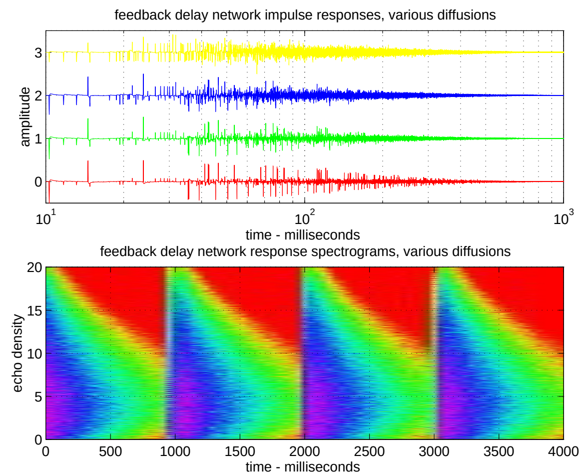

Fig. 6: Top: Feedback delay network impulse responses at different diffusion settings—0.2 (red), 0.4 (green), 0.6 (blue), and 1.0 (yellow)—with $t_{60} \approx 1.0$ s and roughly highpass equalization. Bottom: Corresponding spectrograms of impulse responses (from left to right). Color scaling runs from high (magenta) to low (red) dB values.

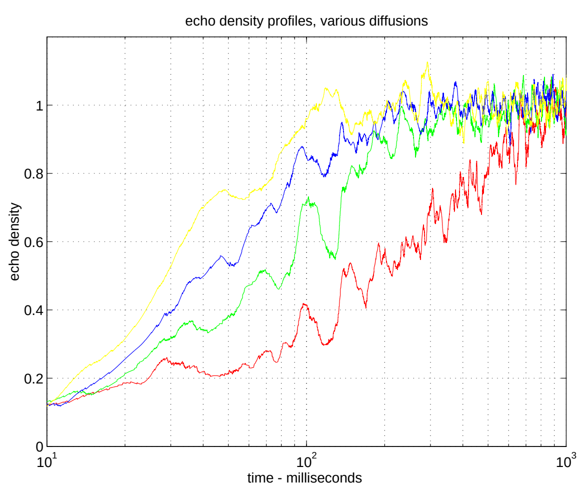

Fig. 7: Echo density profiles of feedback delay network impulse responses at different diffusion settings—0.2 (red), 0.4 (green), 0.6 (blue), and 1.0 (yellow)—with $t_{60} \approx 1.0$ s and roughly highpass equalization. These correspond to the impulse responses shown in Figure 6.

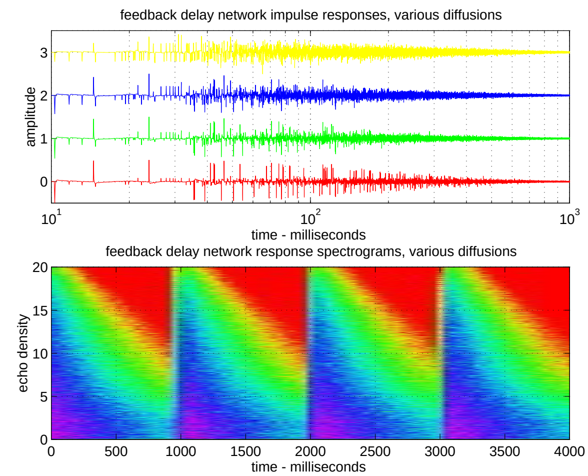

Fig. 8: Top: Feedback delay network impulse responses at different diffusion settings—0.2 (red), 0.4 (green), 0.6 (blue), and 1.0 (yellow)—with $t_{60} \approx 3.5$ s and lowpass equalization. Bottom: Corresponding spectrograms of impulse responses (from left to right). Color scaling runs from high (magenta) to low (red) dB values.

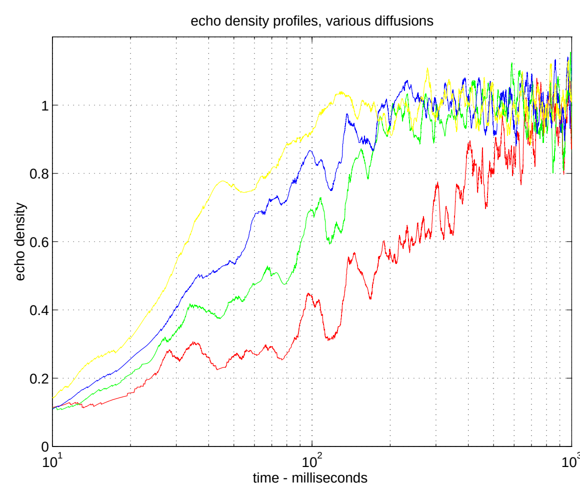

Fig. 9: Echo density profiles of feedback delay network impulse responses at different diffusion settings—0.2 (red), 0.4 (green), 0.6 (blue), and 1.0 (yellow)—with $t_{60} \approx 3.5$ s and lowpass equalization. These correspond to the impulse responses shown in Figure 8.

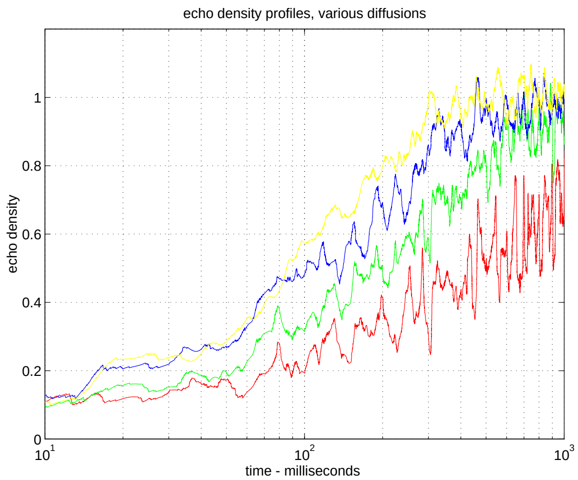

Fig. 10: Echo density profiles of feedback delay network impulse responses at different diffusion settings—0.2 (red), 0.4 (green), 0.6 (blue), and 1.0 (yellow)—with $t_{60} \approx 3.5$ s, lowpass equalization, and a combined delay length approximately twice as long as that of the FDN used to generate the impulse responses in Figure 8.

equalization; the other has a long decay time and a lowpass equalization that results in a dark-sounding reverb.

A last test was conducted to verify whether the rate of echo density increase changes in proportion to the changes in typical delay line length. Impulse responses were generated with parameter settings matching those shown in Figure 9, except that FDN delay line lengths roughly twice as long as the original were used. (Delay line lengths remained incommensurate in order to preserve the quality of the reverberator; they were not simply made twice as long as those the first FDN.)

It is expected that with delay lines approximately twice as long in the new FDN, the rate of echo density increase would be about twice as slow. Figures 9 and 10 illustrate this inversely proportional change in echo density increase. Comparing the echo density profiles associated with the same diffusion value, the time at which the late field is reached (when $\eta(t)$ first has a value of 1 or greater) in Figure 10 is approximately twice that in Figure 9.

4. DISCUSSION

4.1. Properties of the Echo Density Measure

The echo density measure presented here is normalized to take on values in the range from zero to about one, with zero indicating a sparse set of reflections and one indicating a Gaussian late field. It was seen to perform well with both measured and artificial reverberation, being sensitive to the level of diffusion while remaining insensitive to decay rate, equalization, level, and sampling rate.

The echo density profile is useful for examining how the echo density increases before the late field start and, in particular, for estimating when the late field begins. It is logical to define the start of the late field as the point in time when the echo density profile $\eta(t)$ first attains a value of one, as this is its expected value in the presence of Gaussian noise. However, to accomodate the natural variation due to the finite window length, we suggest defining the late-field start as the time at which the profile is first within σ_l of one, where σ_l is the standard deviation of the $\eta(t)$ in the late field.

4.2. Temporal Reverberation Quality

The echo density measure can be interpreted as a psychoacoustic measure which acts as an indicator of temporal quality and is a good alternative to counting echoes when determining quality in artificial reverberation. Since an impulse response is maximally diffuse when a Gaussian distribution is reached, it follows that the late reverberation portion of the response has begun, reflections are fully dense, and that the perceptual quality is high.

The temporal quality of the feedback delay network impulse responses is strongly correlated to the diffusion settings used to generate the responses. For example, the reverberation resulting from a diffusion setting of 0.2 can be described as "crackly" or "sputtery," while the reverberation resulting from the highest diffusion setting has more of a smooth, "sandy" texture. For many of the impulse responses, the transition over time in texture from "crackly" to "sandy" is audible, tracking the echo density profile on its climb to one.

Traditional acoustical parameters for reverberation [12] have not included measures related to reflection density or the description of temporal timbre, but the time-domain quality or texture of a reverberant signal can be as varied and audible as the (frequency-dependent) reverberation time. Since the echo density measure is able to discriminate so well between different diffusion settings and the resulting rate of echo density increase, it has much potential as a tool for evaluating the time-domain timbre of reverberation. Such an analysis would be helpful in the more precise design of artificial reverberators intended to emulate a particular space or impulse response.

5. REFERENCES

[1] Jot, J.-M., Cerveau, L., and Warusfel, O., "Analysis and synthesis of room reverberation based on a statistical time-frequency model," in *Proceedings of the 103rd AES Convention*, preprint 4629, New York, September 26–29, 1997.

[2] Huang, P. and Smith, J. O., "Loudness-based display and analysis applied to artificial reverberation," in *Proceedings of the 2004 International Computer Music Conference*, pp. 37–40, Miami, Florida, November 1–6, 2004.

[3] Rubak, P., "Evaluation of artificial reverberation decay quality," in *Proceedings of the 118th AES Convention*, preprint 6395, Barcelona, Spain, May 28–31, 2005.

[4] Rubak, P. and Johansen, L., "Artificial reverberation based on a pseudo-random impulse response," in *Proceedings of the 104th AES Convention*, preprint 4875, Amsterdam, Netherlands, May 16–19, 1998.

[5] Rubak, P. and Johansen, L., "Artificial reverberation based on a pseudo-random impulse response II," in *Proceedings of the 106th AES Convention*, preprint 4875, Munich, Germany, May 8–11, 1999.

[6] Griesinger, D., "Practical processors and programs for digital reverberation," in *Proceedings of the 7th International Conference of the Audio Engineering Society*, pp. 187–195, April, 1989.

[7] Väänänen, R., Välimäki, V., Huopaniemi, J., and Karjalainen, M., "Efficient and parametric reverberator for room acoustics modeling," in *Proceedings of the 1997 International Computer Music Conference*, pp. 200–203, Thessaloniki, Greece, September 25–30, 1997.

[8] Bickel, P. and Doksum, K., *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, 1977.

[9] Pierce, A., *Acoustics*, McGraw-Hill, 1981.

[10] Kuttruff H., *Room Acoustics*, Fourth edition, Spon Press, 2000.

[11] Jot, J. M. and Chaigne, A. "Digital delay networks for designing artificial reverberators,"in in *Proceedings of the 90th AES Convention*, preprint 3030, Paris, January 1991.

[12] ISO 3382, "Acoustics—measurement of reverberation time of rooms with reference to other acoustical parameters," 1997.