Audio Engineering Society

Convention Paper

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

Reverberation Echo Density Psychoacoustics

Patty Huang, Jonathan S. Abel, Hiroko Terasawa, and Jonathan Berger

[1] *Center for Computer Research in Music and Acoustics (CCRMA), Department of Music, Stanford University, Stanford, CA 94305, USA*

Correspondence should be addressed to Patty Huang (pph@ccrma.stanford.edu)

ABSTRACT

A series of psychoacoustic experiments were carried out to explore the relationship between an objective measure of reverberation echo density, called the normalized echo density (NED), and subjective perception of the time-domain texture of reverberation. In one experiment, 25 subjects evaluated the dissimilarity of signals having static echo densities. The reported dissimilarities matched absolute NED differences with an R2 of 93%. In a 19-subject experiment, reverberation impulse responses having evolving echo densities were used. With an R2 of 90%, the absolute log ratio of the late field onset times matched reported dissimilarities between impulse responses. In a third experiment, subjects consistently reported breakpoints in echo pattern character at NEDs of 0.3 and 0.7.

1. INTRODUCTION

Reverberation is a principal component of timbre, effecting the perceived clarity and loudness of a sound, as well as providing vital cues regarding inference of distance, direction and physical space. Perceptual studies of reverberation primarily focus on decay time [1], and the effect of reverberation on loudness, distance [2], and room size perception [3].

The contribution of reverberation to the perceived clarity of a sound, or more generally, a sound's perceived *texture*, is typically recognized in terms of cross-modal adjectival analogies to visual clarity, focus and blur. Composer Alexandre Cellier described a particular reverberant space as producing sound that was

> blurred and confused, somewhat resembling — in the domain of optics — the vision obtained by an opera-glass out of focus. [4]

More common, particularly in audio engineering and architectural acoustics contexts, is an association of the ratio of direct and reflected signal in terms of relative moisture, where the distorting effect of reverberation of an unreflected (*dry*) signal is described in terms of adding *wetness*.

Sabine's pioneering work establishing a perceptually based measure of acoustic absorption and the widely accepted index of acoustic clarity (C80) are the foundations of physical and perceptual correlates describing reverberation. A statistical description of the number of reflections per second of an acoustic signal is another useful correlate between perception and physics [5].

Recently there has been interest in the transition between early reflections and late reverberation in artificial reverberation and room impulse responses [6, 7, 8, 9].

The temporal evolution of echo density, and in particular its rate of increase, influence the perceived time-domain attribute of timbre, which we refer here to as a sound's *texture*[10].

In [6], a method was presented for measuring the echo density of reverberant impulse responses over time by examining the percentage of impulse response taps lying outside the local standard deviation. This measure, called the normalized echo density or NED, ranges from near zero, indicating few echoes, to around one, indicating a fully dense reverberation having Gaussian statistics. Two recent studies [8, 9] share the common theoretical direction with the proposed model of assessing the "Gaussianness" of the echo density used in this study [6], although there were differences in the analytical methods of these studies.

However, the effect of this aspect of reverberation on the perception of timbre and texture has yet to be explored in any quantifiable investigation. Although the statistical properties of early reflections and late reverberation are well studied for both artificial reverberation and room impulse responses, the relationship between such statistical descriptions and the relevant perceived quality of sound texture demands further research and experimentation.

Our earlier work [11] on the validity of describing noise-like sound texture using NED constitutes one such psychoacoustic study. In these informal listening tests, NED showed promise as a predictor of the perception of such noise textures. NED correlated well to the perception of smoothness and roughness

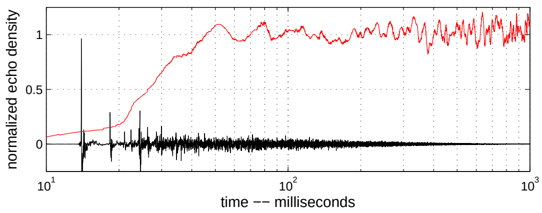

Fig. 1: Normalized echo density profile (red) of a measured room impulse response (black). Note that the time axis is on a logarithmic scale.

(or degree of granularity) in noiselike echo patterns, while being insensitive to the bandwidth (or, loosely speaking, the echo duration) of the noises. For example, noises with low NED (e.g., NED $= 0.2$) were perceived as "sputtery" regardless of bandwidth, and noises with high NED (e.g., NED $= 0.8$) were perceived as "smooth" regardless of bandwidth.

These results motivated us to conduct more formal psychoacoustic experiments to quantitatively investigate (1) the perception of static noise textures, and (2) dynamic mixing time perception, both modeled by NED.

In this paper we report the results of psychoacoustic experiments which show how objective measures of reverberation echo density relate to subjective perception.

2. NORMALIZED ECHO DENSITY

Over a sliding reverberation impulse response window, the *normalized echo density profile* $\eta(t)$ is the fraction of impulse response taps which lie outside the window standard deviation, normalized to that expected for Gaussian noise:

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

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

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

$\mathbf{1}\{\cdot\}$ is the indicator function, returning one when its argument is true and zero otherwise, and $\text{erfc}(1/\sqrt{2}) \doteq 0.3173$ is the expected fraction of samples lying outside a standard deviation from the mean for a Gaussian distribution [6].

The normalized echo density profile (NEDP) 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:

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

with

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

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

Figure 1 shows the normalized echo density profile of a measured room impulse response using a 20 ms Hanning window. NED values are near zero during the early reflection portion of the reverberation, indicating a low echo density. The NED value increases over time to a value near one, suggesting Gaussian-like statistics, where it remains for the duration of the impulse response. As described in [6], what sets one NED profile apart from another is the rate of increase and the time at which a value near one is first attained, indicating the start of the late field.

As developed in [11], the normalized echo density η can be related to the absolute echo density ρ, measured in echoes per second, by the following expression:

$$\eta = \frac{\delta\rho}{\delta\rho + 1}, \quad (5)$$

where δ is the echo duration in seconds, or alternatively the inverse echo bandwidth in 1/Hz .

3. STATIC ECHO DENSITY

In order to conduct a systematic analysis of echo density psychoacoustics, artificial echo patterns were synthesized for a variety of static echo densities and echo bandwidths.

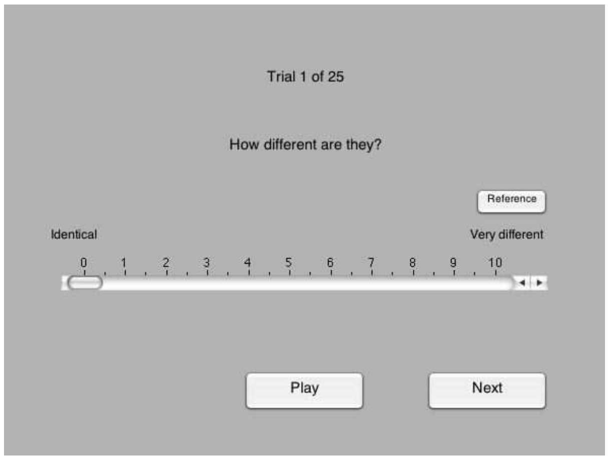

Fig. 2: Dissimilarity experiment graphical user interface.

A Poisson process was used to generate echo arrival times using absolute echo densities ranging from 10 echoes/s to 2.8e5 echoes/s. Echo amplitudes were drawn from Gaussian distributions with variance scaled by the echo density so that energy is roughly constant across echo patterns. Sinc interpolation was used to convert echo times and amplitudes into an echo pattern, and echoes having a range of different durations were generated by applying second-order Butterworth lowpass filters having bandwidths from 1.0 kHz to 10 kHz. [11]

Stimuli for the following experiments described in this section were selected from this large collection of synthesized echo patterns based on the combination of echo pattern bandwidth and echo density desired.

3.1. Static Echo Pattern Dissimilarity

In this first experiment, we investigated the relationship between NED and perception of echo patterns with static echo densities. Our primary interests are (1) whether the texture descriptions (NED, AED, and log AED) relate in a simple way to perceived texture dissimilarity, and (2) if those relationships are consistent across bandwidths, i.e., echo durations.

3.1.1. Method

Participants. Twenty-five normal-hearing participants, graduate students and faculty members from

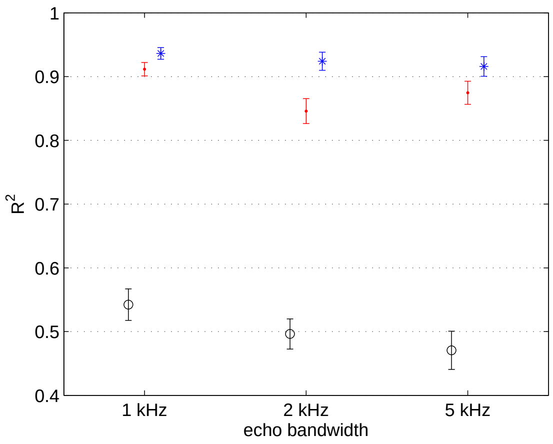

Fig. 3: Mean R^2 and 95% confidence intervals of linear regression on dissimilarity ratings of echo patterns having static echo densities, using AED (o), log AED (.), and NED (*) as the independent variable.

Center for Computer Research in Music and Acoustics at Stanford University volunteered for the experiment. All of them were experienced musicians and/or audio engineers with various degrees of training.

Stimuli. Three sets of echo patterns having five different static echo densities (NED $= 0.13, 0.24, 0.57, 0.74, 0.90$) were generated, with each set having different echo bandwidths (1 kHz, 2 kHz, and 5 kHz, corresponding to echo durations of roughly 1.0 ms, 0.5 ms, and 0.2 ms, respectively). The textures of the stimuli were varied so that granularity ranged from sparse to smooth, while the other factors such as duration, loudness, and bandwidth, were kept constant.

Procedure. There were three sections in the experiment, one section for each of the three sets of the stimuli. Each section consisted of a practice phase and an experimental phase.

The task of the participants was to rate the perceived texture dissimilarity of the presented pair, and listen to the sounds, played in sequence with a short intervening silence. They then enter their perceived dissimilarity using a 0 to 10 scale, with zero indicating that the presented sounds were identical, and 10 indicating that the two sounds in the presented pair were the most different within the section.

The participants press the "Play" button of the experiment GUI using a slider. In order to facilitate the judgment, the pair having maximal texture difference in the section (i.e., the pair of lowest and highest echo density sequences, defined to have a dissimilarity of 10) is available as a reference pair throughout the practice and experimental phases. Participants were allowed to listen to the testing pair and the reference pair as many times as they want, but were advised not to repeat too many times, before making their final decision on scaling, and proceeding to the next pair.

In the practice phase, five sample pairs were presented for rating. In the experimental phase, twenty-five pairs per section (all the possible pairs from five stimuli) were presented in a random order. The order of presenting the sections was randomized as well.

It should be pointed out that the maximally dissimilar pair used as a reference employed different sequences than those presented for rating. Also, so as to distinguish the ability to discern different sequences having identical echo densities from the ability to recognize identical sounds, two different sequences were generated at each echo density, and each pair presented drew one sound from each generation.

3.1.2. Analysis

The dissimilarity judgments were analyzed using linear regression (also known as least squares estimation) [12], with absolute NED differences as the independent variable, and their reported perceived dissimilarities as the dependent variable.

The mean of the coefficient of determination (R2, R^2, or R-squared, which represents the goodness of fit) among participants is used to judge the linear relationship between the NED distance and perceived dissimilarity. We first applied individual linear regression for each section and each participant. The R2 values of one section from all the participants were then averaged to find the mean degree of fit (mean R2) of each section.

In addition to the NED-based analysis, the same analyses were repeated using distances based on

AED and on log AED, as independent variables. Figure 3 shows mean R2 values from the linear regression analyses based on these three independent variables.

Absolute difference in NED is a good model for perceived texture dissimilarity, having a mean R2 of 93%. The log AED is a reasonable indicator of texture dissimilarity, with a mean R2 of 88%. AED, however, fails as a usable model.

3.2. Texture Categorization

This experiment inherits the framework described in our earlier paper [11]. The basic idea is to understand if there are any commonly perceived anchors in the perception of gradually changing texture, e.g. if there are clear boundaries to divide the texture clusters when the texture is changing from smooth to rough.

In this experiment, we asked the participants to divide the static echo noises into three groups, and observed the trend in the reported boundaries. Also of interest is whether a boundary point in NED is consistent among echo patterns with various bandwidths.

3.2.1. Method

Participants. Nine normal-hearing participants, musicians, recording engineers, and staff from the Department of Music and Sound at the Banff Centre volunteered for the experiment.

Stimuli. Four sets of echo patterns having 19 different static echo densities (NED $= 0.05, 0.10, 0.15, ..., 0.95$) were generated at each of four bandwidths (1 kHz, 2 kHz, 5 kHz, and 10 kHz). The textures of the stimuli were varied so that granularity ranged from sparse to smooth, while the other factors such as duration, loudness, and bandwidth, were kept constant.

Procedure. Buttons allowing the subject to listen to the nineteen static noise patterns are presented in ascending NED order. The participants were instructed to listen to the noise patterns as many times as they wished and in whatever order. They were asked to select two breakpoints, grouping the noise sequences into three texture regions, e.g., rough, medium, and smooth. The sections were organized by bandwidth, and the order of section presentation was randomized.

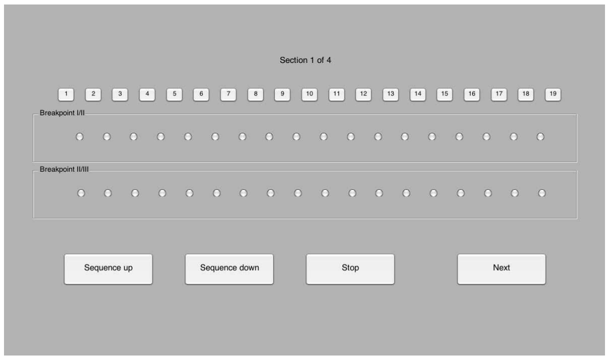

Fig. 4: Texture categorization experiment graphical user interface.

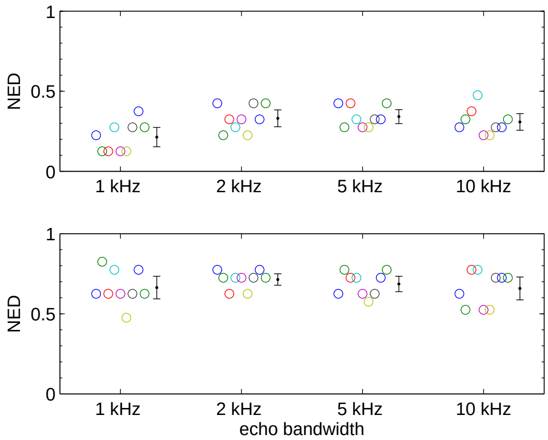

Fig. 5: Breakpoint 1 (top) and breakpoint 2 (bottom) separating three texture regions along a continuum of low to high static echo density. Response means and 95 % confidence intervals (.) are plotted to the right of individual subject responses (o).

3.2.2. Analysis

The NED values of the reported breakpoints are shown in Figure 5 along with the mean NED and 95% confidence intervals for each of the experiment sections. The subject responses are seen to cluster around an NED of 0.3 for the first breakpoint, and an NED of about 0.7 for the second breakpoint irrespective of the stimulus bandwidth. These breakpoints are in good agreement with perceived noise texture groups described in [11].

units	texture breakpoint	echo bandwidth (kHz) 1	2	5	10
NED	1	0.21	0.33	0.34	0.31
	2	0.66	0.71	0.69	0.66
AED	1	182	717	1884	2496
	2	1595	4007	8468	12670

Table 1: Mean texture breakpoints across echo bandwidths, expressed in normalized echo density and absolute echo density (echoes/second).

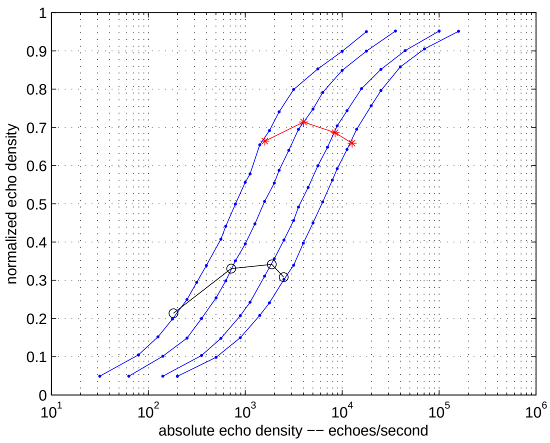

Fig. 6: Mean texture breakpoints (o,*) for echo patterns (.) having static echo densities and bandwidths of 1, 2, 5, and 10 kHz (left to right).

Mean breakpoint values were also computed for each of the sections in terms of the absolute echo density (AED). These and the mean NED values appear in Table 1, and are plotted in Figure 6. Figure 6 also shows the NED-AED pair associated with each of the static echo sequences presented. The breakpoints are seen to occur at NED values across bandwidth, whereas they occur at different AED values, roughly exponentially increasing with increasing bandwidth.

3.3. Texture Matching

This experiment was conducted to study the perceived texture quality across bandwidth. A reference echo pattern having a bandwidth of 3.16 kHz was provided, and participants were asked to select an echo pattern with the closest texture from echo patterns having a different bandwidth. The idea is to form a sort of "equal texture contour" across echo bandwidth and to test the stability of NED across bandwidth.

3.3.1. Method

Participants. Ten normal-hearing participants, musicians, recording engineers, and staff from the Department of Music and Sound at the Banff Centre volunteered for the experiment.

Stimuli. Four sets of echo patterns having 17 different static echo densities (NED $= 0.1, 0.15, ..., 0.90$) were generated, with each set having a different echo bandwidth (1 kHz, 2 kHz, 5 kHz, and 10 kHz). The textures of the stimuli were varied so that granularity ranged from sparse to smooth, while other factors such as duration, loudness, and bandwidth, were kept constant. In addition, three echo patterns of bandwidth 3.16 kHz and NED $= 0.25, 0.5, 0.75$ were used as reference echo patterns.

Procedure. The experiment had 12 sections (three reference patterns, for each of four test sets). Within a section, pairs of the reference sound and one of the seventeen test sounds were prepared and presented with icons on the computer display. Participants were asked to listen to reference/sound pairs thoroughly as many times as they desired, and to select one of the nineteen test sounds which had the most similar perceived texture.

3.3.2. Analysis

NED values of the static echo patterns perceived to match the texture of a 3.16 kHz-bandwidth reference pattern are shown in Figure 8 for each of three reference pattern NEDs. The corresponding mean NED and mean AED values appear in Table 2. The mean matching NED values are all close to the reference NED values, indicating that, as a predictor of perceived texture, NED is insensitive to bandwidth. By contrast, AED produces bandwidth-dependent equal texture contours, taking on an exponential curve.

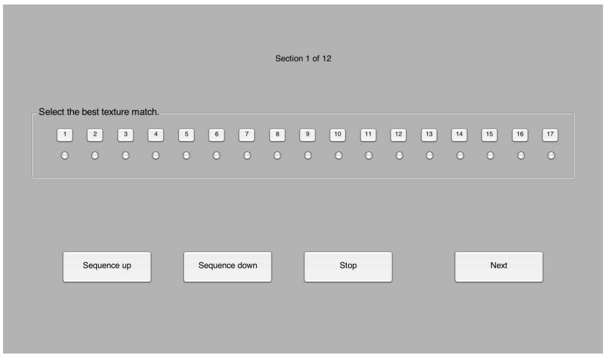

Fig. 7: Texture matching experiment graphical user interface.

units	ref	echo bandwidth (kHz) 1	2	*3.16*	5	10
NED	1	0.34	0.25	*0.25*	0.26	0.21
	2	0.51	0.43	*0.50*	0.39	0.38
	3	0.77	0.78	*0.75*	0.71	0.58
AED	1	423	508	*794*	1511	1667
	2	897	1212	*2512*	2689	5911
	3	3617	6219	*7943*	11859	11122

Table 2: Mean texture match across echo bandwidths to three reference patterns having an echo bandwidth of 3.16 kHz, expressed in normalized echo density and absolute echo density (echoes/second).

4. TIME-VARYING ECHO DENSITY

Reverberation impulse responses were synthesized using absolute echo density profiles which increased quadratically over time, similar to the time behavior of echo density in three-dimensinal spaces. A set of echo patterns having echo bandwidths of 10 kHz were generated which reached an NED of 0.5 at times ranging from 20 to 2000 ms (see Figure 9).

Frequency-dependent decay was implemented by sending the echo patterns through an octave-band filter bank, imposing appropriate amplitude envelopes on the bandpassed signals, and summing the windowed signals. Three frequency-dependent decay

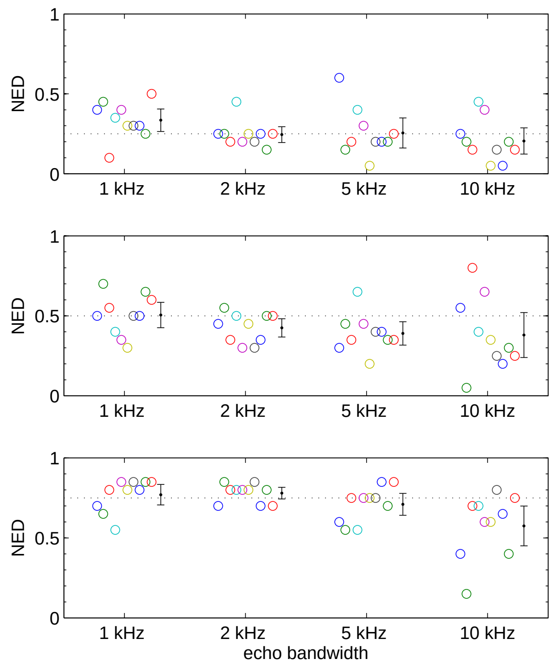

Fig. 8: Static echo pattern NEDs most similar to a reference static echo pattern havinga NED of 0.25, 0.5, or 0.75 (top to bottom). Response means and 95 % confidence intervals (.) are plotted to the right of individual subject responses (o). The associated reference NED is represented by a dotted line.

and equalization profiles were designed in order to have a wide range of realistic-sounding reverberation impulse responses (see Table 3). The echo patterns were then normalized relative to the energy in the first 0.25 seconds in order to maintain similar volume levels between patterns having different reverberation times.

4.1. Mixing Time Dissimilarity

In this experiment, we investigated the perception of time-varying echo patterns, focusing on the relationships between perceived dissimilarity and mixing time, defined as the time at which a specified NED is achieved.

4.1.1. Method

Participants. Eighteen normal-hearing participants,

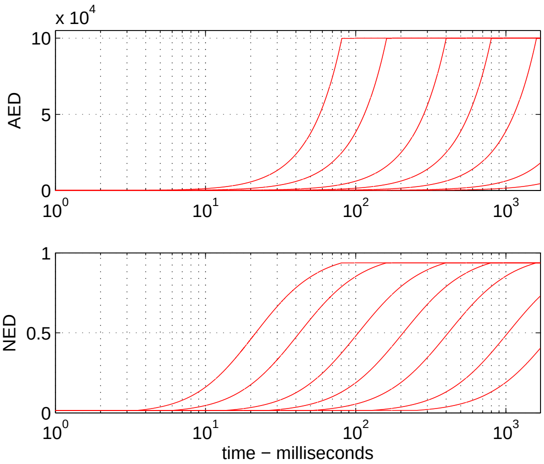

Fig. 9: Absolute echo density profiles (top) and normalized echo density profiles (bottom) for the set of echo patterns used to generate the stimuli in the mixing time dissimilarity experiment.

musicians, recording engineers, and staff from the Department of Music and Sound at the Banff Centre volunteered for the experiment.

Stimuli. Four sets of stimuli were generated with different decay rates and equalizations as listed in Table 3. Within each set the decay and equalization were constant, but the mixing time was varied across seven values.

Procedure. The framework of this experiment is the same as that of the static echo pattern dissimilarity experiment. There were four sections, associated with the four sets of stimuli. As before, each section consisted of a practice phase and an experimental phase, with the subjects rating dissimilarity on a scale ranging from 0 to 10.

In the practice phase, five sample pairs were presented for rating. In the experimental phase, twenty-four pairs per section (about half of all possible pairwise combinations of the seven stimuli) were presented in a random order. The order of section presentation was randomized as well.

Only 24 out of the possible 49 pairs were presented in order to prevent subject fatigue. The 24 pairs selected were chosen from the symmetric half of all possible combinations, with the pair orders randomized. Out of seven possible identical pairs, three of them were selected for testing in the experiment. This approach is justified in view of the fact that the dissimilarity judgment of symmetric pairs (pairs of sounds, AB and BA) are known to be statistically equivalent [13].

set	T_{60} at 1 kHz	decay equalization	stimulus duration
1	∞	none	0.85 s
2	2.91	lowpass	1.7 s
3	1.37	slightly lowpass	1.3 s
4	0.65	highpass	0.85 s

Table 3: Reverberation time, frequency-dependent decay characteristic, and sound duration for each stimulus set in the mixing time dissimilarity experiment.

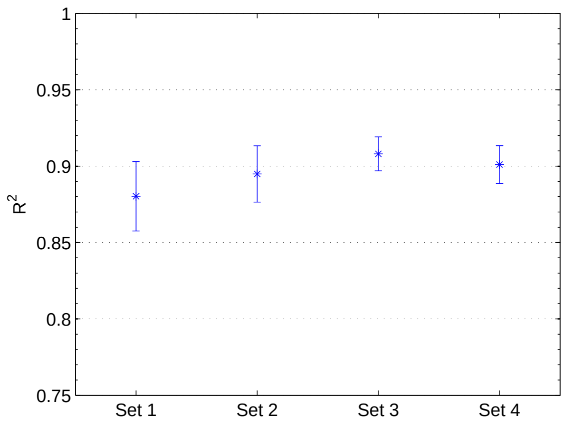

Fig. 10: Mean R^2 and 95% confidence intervals of linear regression on dissimilarity ratings of time-varying echo patterns with different equalization and decay characteristics.

4.1.2. Analysis

The dissimilarity judgments were analyzed using linear regression, with the absolute log ratio of the mixing times selected as the independent variable, and the reported perceived dissimilarities as the dependent variable.

The R2 mean was computed for each section for each subject. The average subject R2 means for each section are shown in Figure 10 with 95% confidence intervals. It is clear that the log mixing time ratio predicts perceived dynamic texture dissimilarity.

Note that the mean R2 is high irrespective of the decay and equalization characteristics tested. Based on this, we expect the absolute log mixing time ratio to indicate perceived texture dissimilarity over a wide range of reverberation impulse responses.

5. SUMMARY AND FUTURE WORK

The static echo pattern dissimilarity experiment provides the quantitative evidence that normalized echo density is a good predictor of texture perception. NED has a strong linear correlation with perceived textures for echo patterns having a wide range of static echo densities, regardless of the echo duration. This time-domain analysis method can be a powerful tool for developing quantitative texture descriptors related in a straightforward way to human perception of sound texture.

The texture categorization and texture matching experiments show that NED as a perceptual descriptor of texture is consistent and robust across bandwidths—static echo patterns having similar NED values will be close in perceived texture regardless of echo bandwidth.

The mixing time dissimilarity experiment showed that the log ratio of mixing time defined at a specified NED successfully predicts the dissimilarity perception of time-varying echo patterns, irrespective of the decay and equalization characteristics imposed on them. Mixing time predicted by analyzing the boundary between early reflections and late reverberation, using such methods as described in [7, 8, 9], can have similar perceptual properties since the concept of mixing time is common regardless of its analytical method.

In studying the perceived mixing time textures, it was noted that the perceived distances between echo patterns having 20 ms and 40 ms mixing times were smaller than would otherwise be predicted by the log ratio mixing times. This suggests further work to explore a "mixing time threshold"—the mixing time below which it is not possible to perceive texture differences.

6. ACKNOWLEDGEMENTS

We thank all the subjects from CCRMA and the Banff Center who participated in our listening tests. Special thanks to Theresa Leonard who invited us to the Banff Center to conduct these listening tests, Chris Segnitz for equipment loans, Allyson Rogers for logistical assistance.

7. REFERENCES

[1] L. L. Beranek, "Concert hall acoustics—1992," *Journal of the Acoustical Society of America* vol. 92, no. 1, pp. 1-39, July 1992.

[2] V. P. Sivonen, "Directional loudness and binaural summation for wideband and reverberant sounds," *Journal of the Acoustical Society of America*, vol. 121, no. 5, pp. 2852-2861, May 2007

[3] S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki, "Psychoacoustic cues in room size perception", in *Proceedings of the 126th AES Convention*, preprint 6084, Berlin, October 2004.

[4] A. Cellier and H. Bagenal, "Acoustic Experiences with Comments," *Music and Letters*, vol. 7, no. 2, pp. 165-176, April 1926.

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

[6] J. Abel, and P. Huang, "A simple, robust measure of reverberation echo density," in *Proceedings of the 121st AES Convention*, preprint 6985, San Francisco, October 2006.

[7] T. Hidaka, Y. Yamada, and T. Nakagawa, "A new definition of boundary point between early reflections and late reverberation in room impulse responses," *Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 326-332, July 2007.

[8] R. Stewart and M. Sandler, "Statistical measures of early reflections of room impulse responses," in *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*, pp. 59-62, Bordeaux, France, September 2007.

[9] G. Defrance and J. D. Polack, "Measuring the mixing time in auditoria," in *Proceedings of Acoustics '08 Paris*, pp. 3869-3974, Paris, France, July 2008.

[10] H. Terasawa, P. Huang, J. Abel, and J. Berger, "A Hybrid Model for Timbre Perception—Part2: The Texture of Sound," in *Proceedings of the 10th International Conference on Music Perception and Cognition*, Sapporo, Japan, August 25-29, 2008.

[11] P. Huang and J. Abel, "Aspects of Reverberation Echo Density," in *Proceedings of the 121st AES Convention*, preprint 7163, New York, October 2007.

[12] W. Mendenhall, and T. Sinich, *Statistics for Engineering and the Sciences*, Prentice Hall, Upper Saddle River, NJ, 1995.

[13] J. Rogers, K. Howard, and J. Vessey, "Using Significance Tests to Evaluate Equivalence Between Two Experimental Groups," *Psychological Bulletin*, vol. 113, no. 3, pp. 553-565, 1993.