PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Room acoustic texture: a methodology for its quantification

Alejandro Bidondo[1]; Leonardo Pepino[2]

[1] UNTREF University, Argentina

[2] UNTREF University, Argentina

ABSTRACT

Room acoustic texture is defined by Beranek as the subjective impression listeners derive from the temporal and amplitude patterns of early reflections, at the receiver's locations inside the room. Traditionally, room acoustic texture was qualified by visual inspection of the room impulse responses (RIR) or counting reflection´s peaks. Taking into account that the reflections in the later part of the RIR follow a Gaussian probability distribution and are totally mixed, we define the early reflections as every amplitude outlier present in a RIR, and mixing time as the instant when their cumulative energy reaches 99% of their total energy. To find outliers amplitudes, we proceeded to cancel the decay of the energy time curve (ETC) under analysis using a mobile median filter. Then, the particular echo density function (*edf*) is defined as the decay-cancelled outliers cumulative energy, over time. From this processing, a group of descriptors were defined that jointly describe the room acoustic texture, at one point in the sound field. Among these descriptors are the mixing time, expected texture and distance between models. Applications of these descriptors and their spatial standard deviation in rooms seem to be very broad, describing their temporal fine-structure.

Keywords: Acoustic texture, Early reflections.

1. INTRODUCTION

The acoustic texture of a room was defined by Beranek: "Texture is the subjective impression that listeners derive from the patterns in which the sequence of early sound reflections arrive at their ears. In an excellent hall those reflections that arrive soon after the direct sound follow in a more-or-less uniform sequence. In other halls there may be a considerable interval between the first and the following reflections. Good texture requires a large number of early reflections, uniformly but not precisely spaced apart, and with no single reflection dominating the others" (Beranek, 1996).

"Sound radiated in a reverberant environment will interact with objects and surfaces in it 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 of reflections 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" (Jot, 1997).

Regardless of being able to assess the importance of early reflections, separating them from a Room Impulse Response (RIR) has been the subject of many investigations, sometimes questioning the need to clearly identify the room in which the listener is (Kahle, 2018).

The aim of this research is to develop a method to efficiently separate early reflections

[1] abidondo@untref.edu.ar
[2] leonardodpepino@gmail.com

(ER) from the rest of a RIR reflections, and study their amplitude and temporal distribution to develop a set of parameters describing the acoustic texture at one point inside a sound field.

2. PREVIOUS STUDIES

As expressed in (O´Donovan, 2008), room acoustics is generally evaluated in terms of various subjective characteristics that expert musicians / listeners assign to sound received at a location in space such as liveness, intimacy, fullness / clarity, warmth / brilliance, texture, blend, and ensemble. Most of these criteria are related to the room impulse response produced with the excitation of sound sources (usually from speakers on stage, or distributed in the hall), registered at several receiver or listener locations.

Several studies proposed to quantify the acoustic texture of a RIR: In (Hidaka, 2008) an acoustic texture parameter is defined as the number of peaks with amplitude higher than the threshold of absolute perceptibility (aWs curve) of a single reflection vs. delay time for music (Schubert, 1961), in the first 80 ms of RIR. This method uses Hilbert transformation for envelope extraction. In (Paskaš, 2010), the author used fractal dimension to quantify texture.

The acoustic texture would express the degree of direct sound coloration, sound source localization modification and, if reflections are coming from lateral directions, acoustic source width (ASW) changes. Regardless of considering these phenomena as acoustic distortion or not, they should be able to be identified and quantified, to finally analyze their uniformity in different areas of a sound field, all these from a RIR. From these previous statements, can be deduced that: a) there could be an "ideal" texture, b) that texture is a matter of early reflections, and c) a criteria for deciding which reflections are considered "early" is needed.

As described by J. D. Polack (Polack, 1988), impulse responses are Gaussian processes, provided that global analysis is carried out on hand of a proper model of impulse responses. In this process, it is essential to discard the early part with strong reflections, and the very late part which simply is background noise. Finally, the reverberation tail exhibits a gaussian distribution of amplitudes in function of time, decaying exponentially.

Abel (Abel, 2006) mentions room impulse response texture as a descriptor for reverberation quality and proposed methods to visualize the echo density profile (EDP) of a RIR, and detect outliers from a gaussian distribution, considering every outlier as an early reflection. He also expressed that the temporal quality of artificial reverberators, analysed through their impulse responses, is strongly correlated to the diffusion settings used to generate them. From Abel´s work (Abel, 2007), the reverberation resulting from a diffusion setting of 0.2 was described as "crackly" or "sputtery," while the reverberation resulting from the highest diffusion setting showed to has a smooth acoustic texture, described subjectively as "smooth" or "windy".

Abel *et al* also said: "Traditional acoustical parameters for reverberation (ISO 3382, 2012) 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."

Considering the lack of metrics to quantify room acoustic texture, a set of descriptors are proposed. In order to find meaningful descriptors, early reflections are identified, isolated and processed from the room impulse response information.

3. ROOM IMPULSE RESPONSE TEXTURE MODEL

3.1 Acoustic Texture

As defined in (11), *texture* is:

- the physical feel of something — smooth, rough, fuzzy, slimy, and lots of descriptions something in between.
- the physical composition of something (especially with respect to the size and shape of the small constituents of a substance).
- the essential quality of something.

As time evolves, every RIR is composed by the direct sound, a group of reflections named “early reflections”, and after those, the “late reflections”. The instant of separation between both is named Mixing time (Mt). After Mt, the reverberation tale can be considered an exponentially decay of gaussian white noise. The particular temporal distribution and amplitudes of this early reflections reflect the room acoustic texture. In this research we quantify the texture of a RIR as the comparison of the shape of the temporal evolution curve of the room´s dynamic system early reflections cumulative summation, with the shape of an “ideal” or “expected” case. This calculation is made between the initial time delay gap (ITDG) and Mt. Around these considerations, a set of descriptors can be defined - globally and over third octave bands -, which describe the temporal evolution of the early sound field.

3.2 Echo density function - *edf*

In this study, an efficient method to compute the energy density function (*edf*), detecting every amplitude outlier in a RIR and constructing a function from the cumulative sum of their energies over time was developed.

For this, considering every amplitude outlier as an ER, a new definition of Mt is proposed: the instant when the cumulative outliers function has reached 99% of its maximum. In this sense, while we cannot talk about ergodicity (12), this proposal complies with the mathematical definition of mixing - time required by a Markov chain for the distance to stationarity to be small (Levin, 2017) - and, consequently of course, with a memoryless state - because all direction information is carried out just in the outliers, which have fade out or disappeared after Mt - (Lindau, 2010). As the analysis process takes place in a RIR, instead of referring to local temporal diffusion, we propose the quantification of the *texture* of the RIR, analized in the 0 : Mt time interval.

To detect all outliers reflections, a median moving filter (MMF) was used with a window length related to the minimum frequency of the first third octave frequency band under analysis, as described by eq. 1. The median is just the middle value in a distribution, and it is less influenced by the outliers, as opposed to the mean that can be "dragged" up or down by extreme values (13).

For a reflection to be outlier in our case, its amplitude has to stand out with respect to values close in time. The method includes a decay subtraction to the RIR under analysis and normalization, relative to the total summation of the outliers energies.

$$Wd = \frac{1}{f_{min}} \cdot f\,s \qquad \text{eq. 1.}$$

Where:
Wd is the MMF window duration,
f_{min} is the minimum frequency of the first third octave band under analysis,

fs is the sampling frequency.

The median moving filter was applied to the energy time curve (ETC), as is described by eq. 2 (also see Figure 1).

$$RIR_{Median}(t) = MMF\left(10 \cdot \log_{10}\left(RIR(t)^2 + \xi\right)\right) \qquad \text{eq. 2.}$$

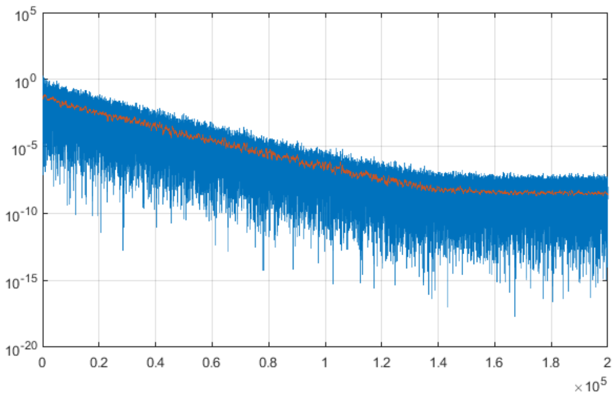

Figure 1. Blue line: Room's Energy Time Curve (ETC) of a generic RIR. Red line: median moving filter result applied to the ETC.

Afterwards, the Decay - cancelled Early Reflections (DcER) was obtained as described by eq. 3.

$$DcER(t) = 10 \cdot \log_{10}\left(RIR(t)^2\right) - RIR_{Median}(t) \qquad \text{eq. 3.}$$

And the echo density function, *edf*, from the RIR under analysis, is obtained by eq. 4.

$$Actual\ edf\ (t) = \frac{cumsum\left(RIR_{Outliers}(t)\right)}{Max\left(cumsum\left(RIR_{Outliers}(t)\right)\right)} \qquad \text{eq. 4.}$$

Where:
RIR (t) is the room impulse response.
RIR_{Median} is the room impulse response after the MMF processing.
DcER: are the Decay-cancelled Early Reflections or outliers, over time.
Actual edf(t): is the calculation applied on the actual RIR under analysis.
edf (t): generally speaking, is the echo density function.

Synthetic RIRs were generated from exponentially decaying gaussian white noise with different RT60s. These signals were devised with constant and evolving echo density, and were used to test the proposed method. Both cases implied an absence of outlier reflections, resulting in a smooth *edf*. It was observed the cumulative energy of the outliers follows eq. 5, and can be thought as a capacitor charging over time, as can be seen in figure 2. The generalized and *ideal* equation governing this behaviour is eq 5.

$$edf\ (t) = \left(1 - a \cdot e^{b \cdot t}\right) \qquad \text{eq.5.}$$

Three edf's were calculated for every RIR: One *actual edf* and two "reference" *edf's.*

Actual edf: is the direct application of the eq. 3 on the actual RIR under analysis.

Ideal edf: For the first "reference" *edf* of eq. 4, *a* and *b* constants are adjusted using two known values taken from the actual *edf*: the initial value of the function, t_0, which corresponds to the initial time delay gap (ITDG) and *Mt*, where the *actual edf* (t) reaches an amplitude of 0.99 from its final value. Also, third octave frequency filtering was applied to the actual RIR, finding Mt values over third octave frequency bands. This way, the *ideal edf*, is established through the ITDG and *actual Mt*.

Expected edf: A second "reference" edf is calculated by best fitting eq.4 to the *actual edf*.

Once the models are attained, the curves are displayed in a log(t) scale.

In figure 2, resulting curves at 1 KHz frequency band are shown with some of the associated texture descriptors.

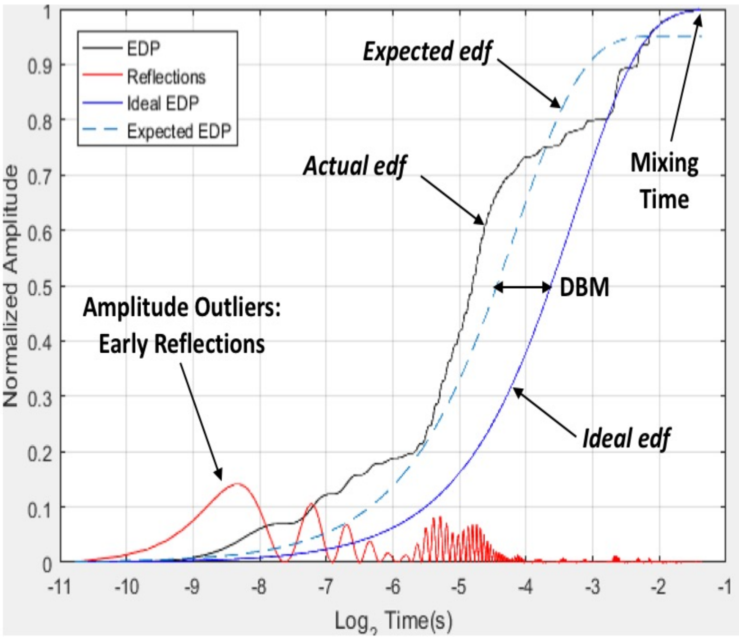

Figure 2. Resulting *edf* curves from the texture calculation for 1 KHz frequency band of a very renowned local Opera Theater. RIR (audience area). Observing the *actual edf* curve, deviation from the smooth growth of the *expected edf* can be seen. Expected texture (ETx) is the Pearson correlation coefficient between the *actual edf* and the *expected edf*. DBM is the Bhattacharya distance between *expected edf* and *ideal edf* curves. Associated numerical results are: ETx = 0.1815, DBM = -2.93, Ctr = 44.67 ms, Mt = 176 ms.

For the case of synthesized gaussian white noise RIRs, the *ideal edf*, *expected edf* and *actual edf* are coincident, as can be seen in figure 3. No considerably irregularities appear in the *actual edf* due to non-existence of early reflections. We refer to this type of cases, the perfectly distributed ER over time, with outliers amplitudes not disturbing the sound field.

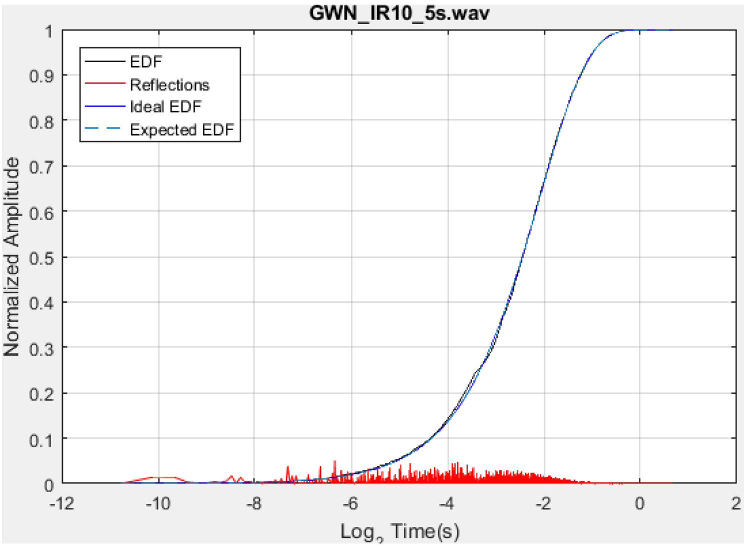

Figure 3. Actual edf (black line), *Ideal edf* (blue line) and *Expected edf* (dashed line) computed on a full bandwidth gaussian white noise synthesized RIR. Early reflections in red. In this case, Expected Texture = 0.999 and DBM ≅ 0.

3.3 Texture of a RIR

As acoustic coatings, particularly diffusers, are added, placed and distributed inside a room, the RIR's *actual edf* curve changes, and tends to move to the origin (t=0); simultaneously an *expected mixing time* (EMt) is estimated through the *expected edf*. If its change is uniform throughout this process - constructing a curve without irregularities from t = ITDG to Mt, resulting in a high value of ETx -, which sometimes means a uniform spatial distribution of acoustic coatings (because this means a uniform spreading of the acoustic energy in time), Mt finally gets effectively reduced compared to the less acoustically treated state.

As this "building process" of the *actual edf* happens, it is possible to measure the distance between the *ideal edf* and the *expected edf*, DBM, the autocorrelation duration of the RIR, Ctr [ms] (Bidondo, 2016), as well as the instantaneous texture that the *actual edf* is showing, compared with the *ideal* and the *expected edf´s*.

Expected Texture (ETx) is defined as the Pearson correlation coefficient between the *actual edf* and the *Expected edf*, as stated by eq. 6.

$$r = \frac{cov(\, real\ edf\ ,\ expected\ edf\ )}{\sigma_{real\ edf} \cdot \sigma_{expected\ edf}} \qquad \text{eq. 6}$$

Where:
r: is the Pearson's coefficient of correlation.
cov: the covariance between variables.
σ: standard deviation of the variable.

The same concept is valid for the Texture (Tx), defined as the Pearson correlation coefficient between the *actual edf* and the *ideal edf*.

The distance between models (DBM) is the Bhatthacharyya distance between the *ideal* and the *expected edf* models. A positive sign in the result means the *ideal edf* curve is after the expected edf curve. A negative sign in the result means the *ideal edf* curve is before the *expected edf* curve. In both cases, it is desirable that DBM be small.

As ETx and Tx values were found to be mainly between 0.8 and 1, it was decided to display r^{20} to get a spread between 0 and 1. An $r^{20} \cong 1$ is obtained for the synthetic gaussian white noise RIR, indicating that the regression predictions almost perfectly fit the data, as can be seen in Figure 3.

4. RESULTS

Third octave calculation over real RIRs showed different Mt, ETx and DBM results.

By modifying the median moving filter window size, evidence showed it has little or no effect in the results.

By analyzing synthesized RIRs through a neural net model, it was found ETx is directly affected by diffusion and RT, and inversely affected by room's volume [m^3]. On the other hand, DBM is inversely affected by diffusion and EDT, and directly affected by room's volume [m^3].

Maximum correlation between the real edf and the expected edf means an almost ideal diffusion for the exhibited expected mixing time, which in turns should coincide with the ideal edf mixing time.

5. CONCLUSIONS AND FUTURE WORK

A simple and efficient calculation method to find the early reflections and the mixing time was presented. It was also presented an ideal temporal evolution curve for the diffusion process of a room considering it as a dynamic system, and the description and quantification of the room acoustic texture through several descriptors. Global and third octave frequency bands results of the proposed descriptors could give information about the need of acoustic diffusers and / or EDT and / or RT modification, and may allow to know those bands outside of "targeted" (ETx, DBM, Mt) values.

Future work includes studying how diffusers surface and their location inside a room modify the *actual edf* function and their associated parameters. Also, a study of the acoustic texture is needed for every type of room.

6. DISCUSSION

Evidence curiously shows high similarity between this dynamic acoustic process (temporal evolution of early reflections carrying directional information) and the facilitated diffusion process in biology, and diffusion processes in chemistry.

In preliminary studies it was observed that both ETx and DBM are sensitive to RT, EDT, room volume, sound field diffuseness and location of the sound source, in different proportions, being able to use them to evaluate the acoustic texture characteristics of the measurement point. This descriptor´s behaviour may led to measure the sound field diffuseness with DBM and to evaluate the balance between room volume [m^3], EDT [s] and sound field diffuseness with ETx. May this balance show that large values of scattering coefficient are not necessarily reflected in the texture of the sound field?

The acoustic expected texture (ETx) is inversely proportional to the Volume [m^3], and directly proportional to EDT [s], RT [s] and the diffuseness of the sound field. Acoustic expected texture is also sensitive to the spatial distribution of the acoustic coating. It was also observed that as diffusion of the sound field increases, for constant room volume, EDT and RT, ETx gets maximized, Mt diminishes, and DBM tends to zero. Also, when *ideal* and *expected edf* curves tend to coincide, acoustic "distortion" would diminish, and could be a numeric, objective target for room acoustic designs.

Faced with this evidence, it would seem appropriate to propose that (temporal) diffusion would not be a state but a process; for that reason it would not be correct to look for a certain amount of diffusion, but a certain development of it in time.

ACKNOWLEDGEMENTS

Authors may acknowledge financial and infrastructure support to UNTREF University.

REFERENCES

(Jot, 1997): 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.

(Kahle, 2018): Kahle, E. "Halls without qualities - or the effect of acoustic diffusion". Proceedings of the Institute of Acoustics: Auditorium Acoustics, Hamburg, Germany, October 4-6, 2018.

(Beranek, 1996): Beranek, L. "Concert Halls, how they sound". p-25. 1996. Acoustical Society of America.

(O´Donovan, 2008): O'Donovan, A., Duraiswami, R., Zotkin, D. "Imaging concert hall acoustics using visual and audio cameras". Perceptual Interfaces & Reality Lab., Computer Science & UMIACS, Univ. of Maryland, College Park. 1-4244-1484-9/08. 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, USA. 2008.

(Hidaka, 2008): Hidaka, T. "On the objective parameter of texture". Forum Acusticum 2008. Paris. France. 2008.

(Schubert, 1961): Schubert, V. P. (1961). "Die Wahrnehmbarkeit von Rückwürfen bei Musik," Z. Hochfrequenztechn. u. Elctroakust., 78, 230-245.

(Paskaš, 2010): Paskaš, I. P., Gavrovska, A. M., Mijić, M. M., Reljin, B. D.. "Qualitative Analysis of Texture of Room Impulse Response using Fractal Dimension". 18th Telecommunications forum TELFOR 2010. Serbia, Belgrade, November 23-25, 2010.

(Polack, 1988): Polack, J. D. "Latransmission del'énergie sonore dans les salles". Ph.D. dissertation, Université du Maine, 1988.

(Abel, 2006): Abel, J. S., Huang, P. "A Simple, Robust Measure of Reverberation Echo Density". AES Convention paper. 2006.

(Abel, 2007): Abel, J., Huang, P. "Aspects of Reverberation Echo Density" 123rd AES Convention. 2007.

(ISO 3382, 2012): ISO 3382, "Acoustics—measurement of reverberation time of rooms with reference to other acoustical parameters," 2012.

(11): https://www.vocabulary.com/dictionary/texture

(12): https://en.wikipedia.org/wiki/Ergodicity

(Levin, 2017): Levin, D. A., Peres, Y. "Markov Chains and Mixing Times", second edition. Ch. 4. American Mathematical Society. 2017. ISBN-13: 978-1470429621

(Lindau, 2010): Lindau, A., Kosanke, L., Weinzierl, S. "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses". 128th Audio Engineering Convention paper, London, UK. 2010.

(13): https://www.researchgate.net/deref/https%3A%2F%2Fwww.youtube.com%2Fwatch%3Fv%3DoaFKTrD_fZk

(Bidondo, 2016): Bidondo, A., Vazquez, S., Vazquez, J., Arouxet, M., Heinze, G. "A new and simple method to define the time limit between the early and late sound fields". 141st AES Convention paper. Los Angeles, USA. 2016.