Audio Engineering Society

Convention Paper 10612

Presented at the 152nd Convention
2022 May, In-Person and Online

Time-Frequency Adaptive Room Optimization of Audio Signals

Samuel Maurer[1,2] and Christof Faller[2]

[1]*Graz University of Technology, Graz, Austria*
[2]*Illusonic GmbH, Bahnstrasse 23, 8610 Uster, Switzerland*

Correspondence should be addressed to Christof Faller (christof.faller@illusonic.com)

ABSTRACT

Room equalization (REQ) is a common method to adapt audio signals to the room in which they are reproduced in. REQ for example attenuates the audio signal at the room resonance frequencies, to reduce negative effects at those frequencies, when the signal is played back. REQ is a time-invariant method. Recently a time-frequency adaptive method to adapt audio signals to rooms has been proposed [1]. The results of a subjective evaluation are presented in this paper. Amount of room reverb and quality are assessed in a blank room, same room with absorbers, and blank room with time-frequency adaptive processing.

1 Introduction

When installing an audio system into a living room, usually it requires acoustical treatment to get the best quality out of it. A fashionably designed and decorated room does not always sound great, but on the other hand, an acoustically treated room often does not look very attractive, because different physical modifications have to be made. In most cases, the reverberation time and room resonances are considered for optimization, which can be achieved with different types of absorbers. But there are rooms, where it is not possible to make modifications due to different limitations, often visual, or lack of space.

Typical negative effects when reproducing sound in a sub-optimal room are coloration [2], masking from the reverberant decay [3], and overly slow reverberant decay at room resonance frequencies. The latter, depending on listening position, results in amplification or attenuation of sound at the room resonance frequencies.

By means of applying room equalization (REQ) to audio signals prior to loudspeaker reproduction, some negative effects of rooms can often be reduced. Minimum phase parametric equalizers [4] are suitable for this task, preventing pre-echoes and being computationally efficiently implementable as digital infinite-impulse-response (IIR) filters [5].

In theory, one can apply the inverse of the impulse response from a loudspeaker to a single position of the room. But as room impulse responses (RIRs) are constantly changing with temperature and humidity [6], RIR inversion would need periodic re-measurement and inversion of RIR, which is practically not feasible.

Here we are limiting room adaptation to REQ and explore how on top of that time-frequency adaptive processing can improve the result further. The considered time-frequency adaptive processing [1] aims at removing the amount of reverb in the loudspeaker signal, which the room will add again. Even if the source has no reverb, its time-frequency-envelopes are modified, e.g. decays shortened, such that after room effect these reflect more those of the source signal. In this paper, we are denoting the described time-frequency adaptive processing "room adaptive processing" (RAP).

The subjective test described in this paper has the goal to assess to which degree RAP can improve sound reproduction on top of REQ. For this purpose, an untreated room was compared to an acoustically treated room. We were wondering to which degree REQ and RAP applied in the untreated room could achieve the performance of the acoustically treated room.

Comparing sound in a room with modifications (without and with absorbers) is difficult, because adding and removing absorbers takes time. Therefore we decided to compare the room without and with absorbers and various processing options via binaural headphones rendering. We conducted a MUSHRA test [7] with instant switching between the different conditions.

The paper is organized as follows. Section 2 reviews the investigated RAP. Section 3 describes the subjective test and results. The conclusions are in Section 4.

2 Time-Frequency Adaptive Room Adaptation

On the left in Figure 1 an example loudspeaker spectrogram X and corresponding spectrogram Y in room is shown. The effect of the room on the spectrogram can clearly be seen. One effect is that the time decays become longer.

With RAP, a modified loudspeaker signal with spectrogram $\tilde{X}$ is reproduced. The goal is that the corresponding spectrogram in room $\hat{X}$ resembles more the original signal's spectrogram X. An example is illustrated on the right of Figure 1.

RAP first determines the spectrogram $\tilde{X}$. Based on signal's original spectrogram X and target spectrogram $\tilde{X}$, a gain filter (time-frequency adaptive suppression filter) is determined. Example spectrograms and corresponding gain filter are shown in Figure 2.

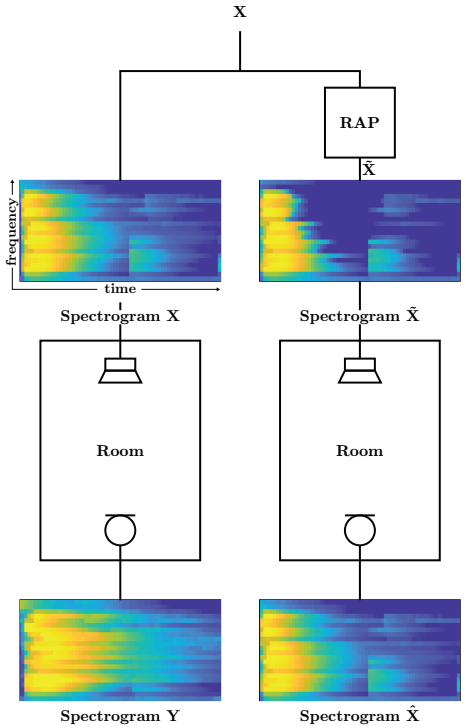

Fig. 1: Left: loudspeaker signal spectrogram X and corresponding spectrogram in room Y. Right: RAP modified loudspeaker spectrogram $\tilde{X}$ and corresponding spectrogram in room $\hat{X}$.

2.1 Computation of target spectrogram

The spectrograms are power spectra with time and frequency indices k and i. The relation between a signal's spectrogram $X(k,i)$ and corresponding spectrogram in room $Y(k,i)$ is modeled with an FIR filter per frequency

$$Y(k,i) \approx H(k,i) \star X(k,i)\,,$$

where $\star$ denotes convolution with respect to time index k and

$$H(k,i) = \delta(k)(1-g) + g r^k\,, \qquad (1)$$

with

$$\begin{aligned} g &= 10^{\frac{G}{10}} \\ r &= 10^{\frac{-60kH}{10 f_s \mathrm{RT60}}}\,, \end{aligned}$$

where G is reverb gain in dB and RT60 is reverberation time. G and $RT60$ can have different values for different frequencies i. The top panel of Figure 3 shows an example P_H with $G = -4$ dB and RT60 $= 0.6$ s.

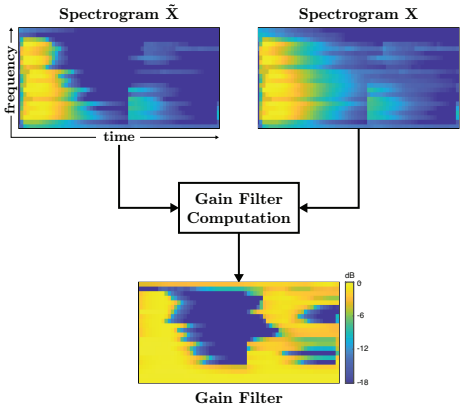

Fig. 2: A time-frequency adaptive gain filter, applied to the audio signal, is computed based on spectrograms $\tilde{X}$ and X.

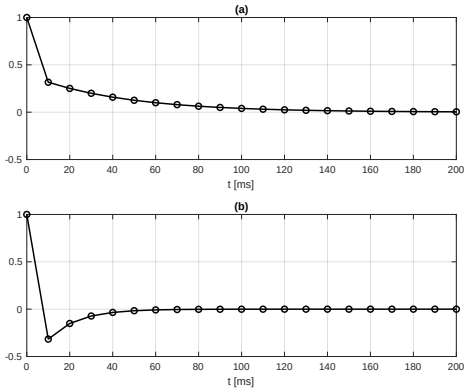

Fig. 3: (a) Example modeling filter H at one frequency and (b) its inverse H^{-1}.

The inverse of (1) is causal,

$$H^{-1}(k,i) = \begin{cases} 1 & \text{if } k = 0 \\ -gr & \text{if } k = 1 \\ P_H^{-1}(k-1,i)(1-g)r & \text{if } k > 1 \end{cases} .$$

The corresponding inverse of the example shown in the top panel of Figure 3 is shown in the bottom panel.

RAP computes the target spectrogram $\tilde{X}$ as

$$\tilde{X}(k,i) = H^{-1}(k,i) \star X(k,i) .$$

Details on how to numerically obtain $H^{-1}(k,i)$, based on room impulse responses, are described in [1].

3 Subjective Evaluation

In order to asses the effectivity of RAP, a MUSHRA test [7] was conducted, comparing

- **A**: Music playback in free-field.
- **B**: Music playback in the same room, but acoustically treated, with REQ.
- **C**: Music playback in the untreated room with REQ and RAP.
- **D**: Music playback in the untreated room with REQ.
- **E**: Music playback in the untreated room.

As mentioned in the introduction, modification of a room (adding and removing absorbers) is not possible quickly. Thus, the test was carried out with headphones and binauralization of the room without and with absorbers and the various processing options.

In terms of clarity and quality of headphones reproduction, A clearly is the best, and is used as reference in the subjective test. The degraded quality of untreated room, E, is used as an anchor. D will show improvement that can be achieved with REQ. B is another reference, the quality of the acoustically treated room (also in the acoustically treated room, REQ is needed for neutral timbre). C is the untreated room with REQ and RAP. How close will this case perform compared to B, that is, untreated room with REQ and RAP, compared to acoustically treated room with REQ?

3.1 Room and Absorbers

A room at the office of Illusonic GmbH was chosen for the investigations. An image of the room is shown in Figure 4. This room was considered without and with absorbers. Figure 5 shows a floor plan of the room, loudspeaker (PSI-Audio A21) position, and microphone array position, and absorber positions. An array with 6 measurement microphones was used.

The left-right distance of the two chosen microphones of the array was 20 cm, such that the measurements

Fig. 4: Room at Illusonic GmbH that was used for the investigations.

[Hz]	100	125	160	200	250	315	400	500	630	800
[α]	0.14	0.19	0.39	0.64	0.84	1.02	1.07	1.07	1.05	1.05
[Hz]	1000	1250	1600	2000	2500	3150	4000	5000		
[α]	1.03	1.02	1.01	0.98	0.99	0.98	0.96	0.94		

Table 1: Absorption coefficients α of the used absorbers for the different frequency bands.

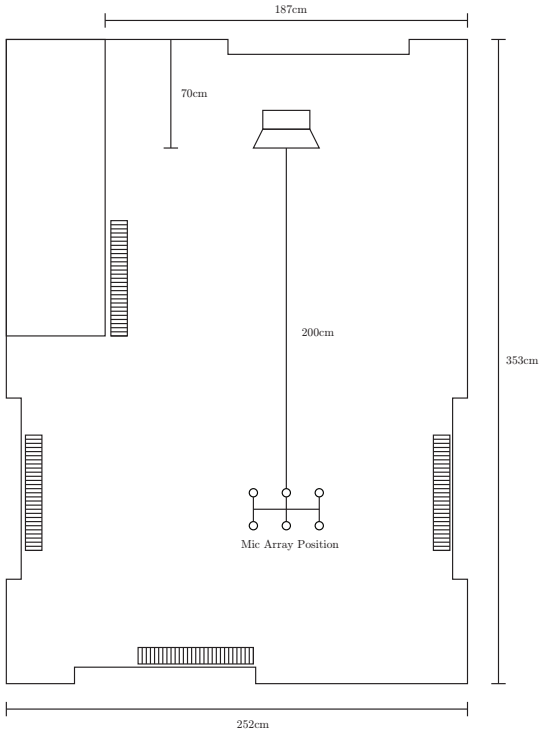

Fig. 5: Room plan with loudspeaker and 6-microphone array position.

could be used for a simple binaural simulation, by filtering audio signals with left and right impulse responses. Measurement with two microphones and corresponding simple binauralization are illustrated on the left and right of Figure 6, respectively.

Glass wool absorbers were used with the dimensions 150 cm x 60 cm x 5 cm. The absorption coefficients in the different frequency bands are shown in Table 1. Note that the values greater then 1 are caused by the diffraction effect [8]. The placement was done temporarily, so the absorbers were not fixed at the walls, but placed on the floor and tilted to the walls. Figure 7 shows an image with the room with absorbers.

With these absorbers, according to Table 1, the absorption does not cover the low end of the frequency spectrum well. Lower frequencies were not specifically treated and this limitation is taken into account in the later discussion of the results. Considering the tilted placement of the absorbers, where the bottom is 20 cm away from the wall and the top 5 cm, we can calculate the frequencies, which are treated the most effectively by the absorber because of the maximum velocity which occurs at $\frac{\lambda}{4}$ away from the wall [9]. Frequencies between f_{Bottom} and f_{Top},

$$f_{Bottom} = \frac{c}{4 \cdot d_{Bottom}} \approx \frac{343[\frac{\mathrm{m}}{\mathrm{s}}]}{0.8[\mathrm{m}]} = 429 \text{ Hz}$$
$$f_{Top} = \frac{c}{4 \cdot d_{Top}} \approx \frac{343[\frac{\mathrm{m}}{\mathrm{s}}]}{0.2[\mathrm{m}]} = 1715 \text{ Hz} .$$

are therefore treated most efficiently.

3.2 Subjective evaluation

Subjects and playback setup Nine experienced listeners between the age of 25 and 52 participated in the subjective test. Due to the large travel distances and the Covid pandemic, some of the subjects did the test on their own with detailed instructions. This was not really problematic because the test was designed for

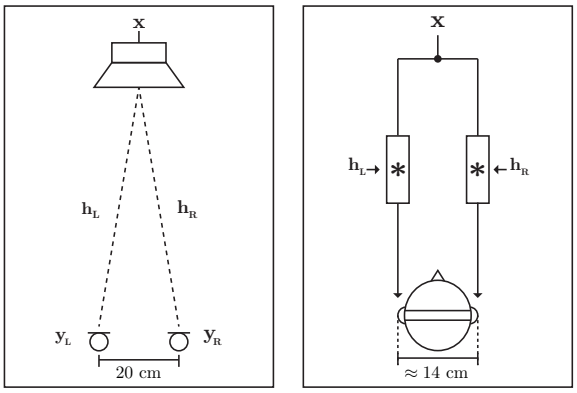

Fig. 6: Left: room measurement. Right: corresponding simple binauralization.

Fig. 7: Room with absorbers.

headphones. The instructions included that high quality over ear headphones and an external audio interface should be used for the test. The subjects who did the test at our office used Sennheiser HD 650 headphones and a RØDE AI-1 Interface.

Testing method A MUSHRA test method was used, comparing a number of modified items to the reference. Besides the reference, a hidden reference and an anchor was used, as well as three items with different modifications. Reference A, anchor E, and different modified versions B-D were described in the beginning of this section.

An internal software of the Institute of Electronic Music and Acoustics (IEM) Graz was used, which acts as a remote control for the DAW Reaper. The subjects could listen to the reference and the other items as many times as they desired until the button was clicked to get to the next music signal. The switching between the stimuli was possible at any time, not only after one loop has finished. Nonetheless, it was recommended, that at least one loop should be listened to before switching between them.

Stimuli Prior to the listening experiment, a large selection of musical signals, ranging from rock, jazz, funk, percussive signals to pop music from different time periods were tested and rated by the first author. All signals were first down-mixed to mono.

The key factors to select the songs were the ability to create short loops that are not varying much over time, so it is easier to compare them, as well as different genres. Two of the selected tracks were used for training purposes only. Table 2 lists the chosen music for training and subjective tests. Each signal was cut into a loop of length 5 to 10 seconds. All clips were sampled at 44.1 kHz.

Use	Song	Original Release
Training	3 Doors Down - I Feel You	2002
	Fat Freddy's Drop - Ernie	2005
Test	Daft Punk - Get Lucky	2013
	Gorillaz - Clint Eastwood	2001
	Michael Jackson - Billie Jean	1982
	Rock Candy Funk Party - Octopus-E	2013
	Khruangbin - Mr. White	2015

Table 2: Excerpts from these songs were used for training and subjective tests.

REQ The spectrum of the room impulse response was calculated and smoothed in one-third octave bands. After inverting this curve, it was applied to the RIRs using a minimum-phase-approximating FIR-filter. Figure 8 shows the spectrum of an unprocessed and smoothed RIR in the top panel, and the corresponding equalization curve in the bottom panel. Note that REQ uses a flat target curve, such that the equalized items would timbre-match the reference A. An effective room curve is part of the non-free-field tuned headphones.

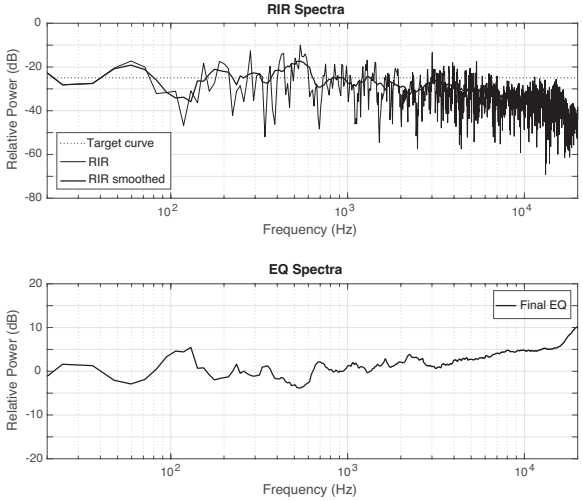

Fig. 8: Top: room impulse response spectrum and 1/3-octave smoothed spectrum. Bottom: corresponding REQ spectrum.

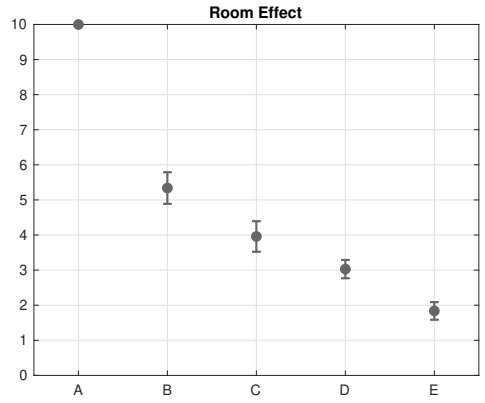

Fig. 9: Test results for room effect with 95% intervals, averaged over listeners and music items.

Level alignment To eliminate the factor of varying levels between the stimuli, which would make it more difficult to rate the actual attributes, a level alignment of all stimuli was made. Because of the large amount of test signals prior to the selection, this step was automated and is therefore also consistent over all items and easy to reproduce. First, the reference items A were normalized in loudness with K-Weighting [10] to make sure that it is not necessary to modify the headphone level during the experiment. Then, the other items B-E were level-aligned with K-Weighting to the corresponding normalized items A.

3.3 Results

Methods for the statistical evaluation The nine subjects rated a total of five songs for "Room Effect" and "Clarity/Quality" compared to the reference. This resulted in a data-set of 45 values for every stimulus (A - E). Using this data, the mean value and confidence interval of 95% was calculated. To investigate the strength of the results, a Wilcoxon signed-rank test [11] was done, since not all sets came from a normal distribution according to Lilliefors [12] tests.

Room Effect Figure 9 shows the result for "room effect" averaged over all listeners and music items. The hidden reference A was recognized by all test subjects and the signal of the unmodified room E is rated significantly ($p < 0.01$) worst. Adding REQ to the unmodified room (D) shows a significant ($p < 0.01$) improvement and additionally adding RAP (C) gave further improvement ($p < 0.01$). When comparing the RAP processed signal with the room which was physically altered (B), there is also a significant ($p < 0.01$) difference.

Clarity/Quality Overall, the results of the second part of the test, where clarity/quality was rated, are shown in Figure 10, again averaged over all listeners and music items. The hidden reference A was again recognized by all test subjects and the signal of the unmodified room E is rated significantly ($p < 0.01$) worst. Adding REQ to the unmodified room (D) shows a significant ($p < 0.01$) improvement and additionally adding RAP (C) gave further improvement ($p < 0.01$). When comparing the RAP processed signal with the room which was physically altered (B), there is also a significant ($p < 0.01$) difference.

Inspection of the individual songs The results which are presented in Figure 11 show, that the overall rating for the different music signals was consistent. For one song (Khruangbin - Mr. White) we can see that the version where RAP was applied was rated even higher than the version of the modified room. This

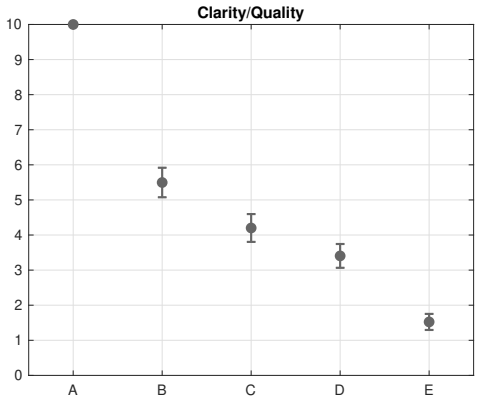

Fig. 10: Test results for clarity/quality with 95% intervals, averaged over listeners and music items.

song contains prominent transients, which could explain the effectiveness of RAP for this specific case. Observations and informal listening showed, that more transient signals had an overall better performance than more stationary signals.

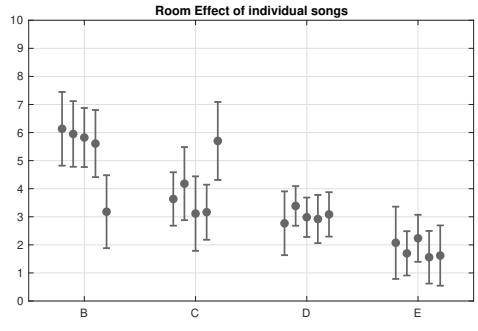

Fig. 11: Test results for room effect with 95% intervals, averaged over listeners for each music item.

Inspection of the individual subjects Figure 12 shows the average ratings of all songs for each subject. The descending rating from B to E is also visible when looking at the individual ratings.

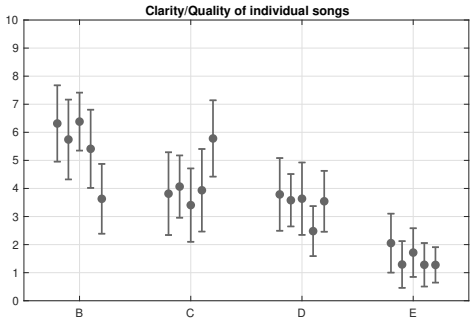

Fig. 12: Test results for clarity/quality with 95% intervals, averaged over listeners for each music item.

3.4 Discussion

For both parts of the test, we can see a large difference from the reference A to the signal that represents the room with absorption B. This difference was expected, because the reference is free-field (direct signal to headphones) versus a binaural signal that was generated using the impulse responses of the measurement in the room with absorption.

When comparing only the signals without the reference (B-E), they look almost equally distributed in descending order from B to E for both experiments, with a slight deviation between D and E when looking at the results of the part where clarity/quality was graded. The goal to achieve similar results by using RAP, as with physically modifying the room was about halfway reached for the tested signals. An aspect that may impact the results, was the fact, that the modified room did not cover the low frequencies well, whereas RAP is more effective in these lower frequency bands.

Because testing multiple listening positions would have increased the length of the listening experiment by the factor of the number of positions, only one listening position was tested. The verification of the effectiveness of RAP throughout a broader listening area was part of informal listening, which indeed indicated that RAP is not very sweet spot sensitive. This is plausible, as a room's reverb decays do not depend on listener position.

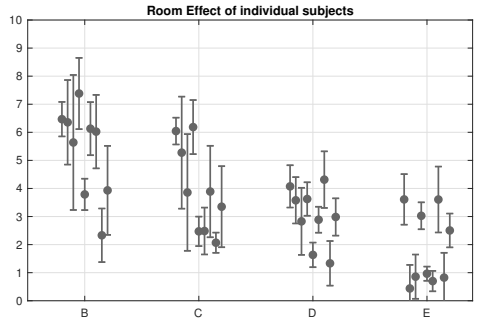

Fig. 13: Test results for room effect with 95% intervals, averaged over all songs for each listener.

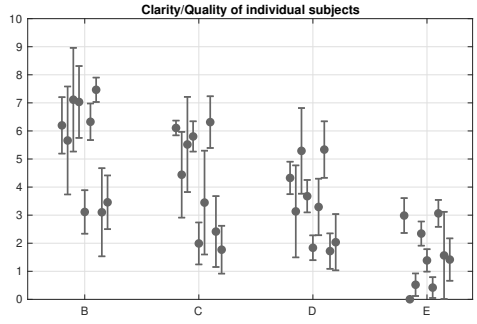

Fig. 14: Test results for clarity/quality with 95% intervals, averaged over all songs for each listener.

4 Conclusion

The goal of the investigation presented in this paper was to assess to which degree time-frequency adaptive room adaptation processing (RAP) can improve results beyond room equalization (REQ).

Via binauralization, listeners compared a blank room to the same room with absorbers. The blank room was tested with REQ and RAP. The reference and anchor for the MUSHRA test were free-field reproduction and blank room without REQ, respectively.

REQ of the blank room gave a significant improvement in terms of perceived amount of reverb and clarity/quality. On top of that RAP gave another significant improvement. These results indicate that indeed RAP has the potential for sound quality improvement in rooms beyond REQ.

References

[1] Faller, C., "Modifying Audio Signals for Reproduction with Reduced Room Effect," in *Preprint 147th Conv. Aud. Eng. Soc.*, 2019.

[2] Faller, C., "Signal Processing for Speech, Audio, and Acoustics," 2006, course Notes, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

[3] Zarouchas, T. and Mourjopoulos, J., "Modeling perceptual effects of reverberation on stereophonic sound reproduction in rooms," *J. Acoust. Soc. Am.*, 126(1), 2009.

[4] Massenburg, G., "Parametric Equalization," in *Audio Engineering Society Convention 42*, 1972.

[5] Oppenheim, A. V. and Schaefer, R. W., *Discrete-Time Signal Processing*, Signal Processing Series, Prentice Hall, 1989.

[6] Elko, G. W., Diethorn, E., and Gänsler, T., "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. Intl. Workshop on Acoust. Echo and Noise Control (IWAENC), Kyoto, Japan*, 2003.

[7] ITU, *Method for the subjective assessment of intermediate quality level of audio systems*, International Telecommunication Union, Geneva, Switzerland, rec. itu-r bs.1534-3 edition, 2015.

[8] Sengpiel, "Absorptionsgrad größer 1," 2005, http://www.sengpielaudio.com/AbsorptionsgradGroesserEins.pdf, Accessed on 23.06.2021.

[9] Dickreiter, M., Dittel, V., Hoeg, W., and Wöhr, M., editors, *Handbuch der Tonstudiotechnik*, De Gruyter Saur, 2014, doi:doi:10.1515/9783110316506.

[10] ITU, *Algorithms to measure audio programme loudness and true-peak audio level*, International Telecommunication Union, Geneva, Switzerland, rec. itu-r bs.1770-4 edition, 2015.

[11] Wilcoxon, F., "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Biometrics Bulletin*, 1(6), pp. 80–83, 1945.

[12] Lilliefors, H. W., "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, 62(318), pp. 399–402, 1967.