Prawda, Karolina; Välimäki, Vesa; Serafin, Stefania

Evaluation of Accurate Artificial Reverberation Algorithm

Published in:
SMC 2020 - Proceedings of the 17th Sound and Music Computing Conference

Published: 01/06/2020

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Prawda, K., Välimäki, V., & Serafin, S. (2020). Evaluation of Accurate Artificial Reverberation Algorithm. In S. Spagnol, & A. Valle (Eds.), *SMC 2020 - Proceedings of the 17th Sound and Music Computing Conference* (pp. 247-254). (Proceedings of the 17th Sound and Music Computing Conference). CERN - European Organization for Nuclear Research. https://smc2020torino.it/adminupload/file/SMCCIM_2020_paper_144.pdf

Evaluation of Accurate Artificial Reverberation Algorithm

Karolina Prawda, Vesa Välimäki
Acoustics Lab,
Dept. Signal Processing and Acoustics
Aalto University, Espoo, Finland
karolina.prawda@aalto.fi

Stefania Serafin
Multisensory Experience Lab,
Dept. Architecture, Design and Media Technology
Aalborg University, Copenhagen, Denmark
sts@create.aau.dk

ABSTRACT

Artificial reverberation algorithms aim at reproducing the frequency-dependent decay of sound in a room that is perceived as plausible for a particular space. In this study, we evaluate a feedback delay network reverberator with a modified cascaded graphic equalizer as an attenuation filter in terms of accurate reproduction of measured impulse responses of three rooms with different decay characteristics. First, the late reverb is synthesized by the proposed method and mixed with the early reflections separated from the original signal. The synthesized and measured signals are compared in terms of their decay characteristics and reverberation time values. The experiment shows that the proposed reverberator design reproduces real impulse responses well, although the decay-rate error exceeds the just noticeable difference of 5% in many cases. Additionally, perceptual qualities of the synthesized sounds were assessed through a listening test. Four qualities were tested for three room impulse responses and three kinds of stimuli. The results show that for the qualities reverberance, clarity, and distance, on average 75–79% of participants noticed only a slight or no difference between the measured and synthetic reverbs. Similar results were obtained for the speech and signing voice stimuli and the reverberation of lecture room and concert hall.

1. INTRODUCTION

Reverberation is considered to be one of the most important sound qualities of physical spaces. It is also used in virtual environments to make them sound and feel more real. Therefore, many algorithms that aim at synthesizing reverberation are used nowadays, with Feedback Delay Networks (FDNs) being among the most popular due to flexibility in design and computational efficiency [1, 2].

To make the artificial reverberation sound perceptually plausible, i.e., logical and probable for the particular space, the energy decay must be frequency-dependent, which can be achieved by inserting attenuation filters into the algorithm. Over time, various types of such filters have been proposed, starting from a first-order low-pass infinite impulse response (IIR) filter [3], to biquadratic filters, which allowed to control the decay in few frequency bands [4], to high-order filter designs [5].

Advanced control over frequency-dependent reverberation is achieved by using a proportional graphic equalizer (GEQ). This idea was first proposed by Jot [6] and later improved by Schlecht and Habets [7] to enhance the system's accuracy while ensuring its stability. Recent work by Prawda et al. [8] suggests to use a modified cascaded GEQ with shifted and scaled frequency response and a first-order high-shelf filter inserted at high frequencies to further increase the accuracy of the reverberation approximation.

The artificial reverberation should primarily be perceptually plausible. Therefore, various types of perceptual evaluation techniques were proposed to assess different qualities of synthesized reverberation. Czyżewski [9] proposed several criteria for assessing concert hall reverberation. The listening tests together with the objective evaluation comparing synthetic late reverberation and a measured impulse response were described in [5, 10–12]. The two studies, [11, 12], are of special interest to the present work since they focus on evaluating FDN reverberators.

This paper presents an objective as well as a perceptual evaluation of the accurate reverberation design proposed in [8]. The study compares impulse responses measured in rooms with various reverberation characteristics with synthesized versions of the same signals. The evaluation includes objective measures, such as frequency-dependent energy decay, and listening tests examining the perceptual qualities of signals for different types of sound stimuli.

The paper is organized as follows. Section 2 describes the algorithm used to synthesize room impulse responses. Section 3 presents the target signals and shows the results of an objective evaluation of the artificial reverberation algorithm. Section 4 describes the listening test and reports on its results and their statistical analysis. Section 5 discusses the results. Finally, Section 6 summarizes the work presented in the paper, concludes on the findings, and proposes ideas for future research.

2. ARTIFICIAL REVERBERATION ALGORITHM

The FDN algorithm with the modified GEQ that was presented in [8] proved to work well for a simplified case in which one delay line (single-delay-line absorptive feedback comb filter) was analyzed. However, that type of

FDN is unusual in practice, as it does not provide enough echo and modal density. Usually, a high-order system is required to obtain smooth reverberation without audible artifacts [13]. Therefore, a commonly used FDN comprising of 16 delay lines was adopted in this work.

Another decision concerning the design of the algorithm was the choice of the feedback matrix. In principle, the stability of an FDN is achieved, when the matrix is unilossless, i.e., it does not cause any loss of energy for any type of delay when no attenuation is introduced in the system [14]. To fulfill the above-mentioned requirement, in the present work a 16th-order Householder matrix is used. It is created by recursively embedding the fourth-order Householder matrix

$$\mathbf{A_4} = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix} \tag{1}$$

for each entry in a matrix of identical structure [4, 12]:

$$\mathbf{A_{16}} = \frac{1}{2}\begin{bmatrix} \mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 \\ -\mathbf{A}_4 & \mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 \\ -\mathbf{A}_4 & -\mathbf{A}_4 & \mathbf{A}_4 & -\mathbf{A}_4 \\ -\mathbf{A}_4 & -\mathbf{A}_4 & -\mathbf{A}_4 & \mathbf{A}_4 \end{bmatrix}. \tag{2}$$

2.1 Attenuation filters

To obtain a frequency-dependent reverberation, attenuation filters must be inserted either at the beginning or at the end of every delay line. They should be designed to approximate the same target gain-per-sample in dB, which is given by

$$\gamma_{\mathrm{dB}}(\omega) = \frac{-60}{f_{\mathrm{s}} T_{60}(\omega)}, \tag{3}$$

where $T_{60}(\omega)$ is the reverberation time in seconds, $\omega = 2\pi f/f_{\mathrm{s}}$ is the normalized frequency, f is the frequency in Hz, and f_{s} is the sampling rate in Hz. For all the delay lines to approximate the same reverberation time, the attenuation should be proportional to the number of unit delays in samples L, such that the attenuation filter's magnitude response in dB is expressed as

$$A_{\mathrm{dB}}(\omega) = L\gamma_{\mathrm{dB}}(\omega). \tag{4}$$

The attenuation filter controls the decay rate of the synthetic response in a broad enough frequency range to make it perceptually the same as the measured room impulse response (RIR). To obtain such similarity, this study used a cascaded GEQ that can regulate the reverberation time in ten octave bands, having their center frequencies from 31.5 Hz to 16 kHz [15].

To smooth the equalizer's magnitude response and to avoid any unwanted increase in $T_{60}(\omega)$ below and above the frequency range of interest, the gains for all the frequency bands were first shifted up (boosted) by their median value and then scaled down (attenuated) by the same number. The more detailed explanation of those operations is presented by Prawda et al. [8]. The equalizer's final response in dB is given by:

$$\widetilde{H}_{\mathrm{dB}}(e^{j\omega}) = g_0 + \sum_{m=1}^{M}(H_{\mathrm{dB,m}}(e^{j\omega}) - \frac{g_0}{M}), \tag{5}$$

where $m = 1, 2, ..., M$ is the number of controlled frequency bands, g_0 is the broadband gain factor (set to the median of all gains), and $H_{\mathrm{dB,m}}$ are the frequency responses of equalizing filters.

Additionally, the first-order high-shelf filter was inserted in the GEQ above 16 kHz to ensure a correct energy decay for the high frequencies. The gain of the filter was set to the gain of the highest peak-notch filter and the crossover frequency, i.e. the frequency at which the gain is the arithmetic mean of the extreme gains in dB [16], fixed at 20.2 kHz, as proposed in [8].

2.2 Early reflections

In this study, the FDN was intended to synthesize the late part of the impulse response and therefore the decision was made to obtain the early reflections from a suitable portion of the measured RIR and mix it with the FDN's output. In this approach, the challenge is to find the correct truncation point to capture the right amount of early reflections.

There exist a few approaches discussing the issue of RIR truncation that suggest studying the skewness and kurtosis of the windowed signal [17], the changes in the RIR's phase over time [18], or the echo density profile [19–21]. In this study the method suggested by Stewart and Sandler [22] was adopted. It assumes that the RIR has an approximately Gaussian distribution in the time domain, with a standard deviation of the group of samples being the measure of the spread of samples defined as

$$\sigma = \sqrt{\mathbf{E}(x^2) - \mathbf{E}(x)^2}, \tag{6}$$

where $\mathbf{E}(x)$ is the expected value of x. In a normal distribution one-third of the samples lies outside and two-thirds inside of one standard deviation from the mean, but in case of early reflections, more samples lie within one standard deviation. Therefore, the truncation point is found by analyzing the ratio of samples outside and inside the standard deviation.

To observe the ratio of samples, the RIRs of interest were windowed with a 20-ms rectangular window with no overlap. The length of the window was described in [12, 23] as favorable, since a shorter one would not provide enough samples to perform reliable calculations, whilst a longer window could lead to choosing the truncation point too early or too late. The ratio's threshold was set to 30%, meaning that when at least 30% of samples lie outside one standard deviation, the point after the window is the correct truncation point. After truncating, the right half of a 32-sample long Hanning window was applied to the early reflections part of the RIR to fade the signal energy smoothly to zero, as suggested in [12].

3. OBJECTIVE EVALUATION

In order to test the proposed method of producing reverberation, an evaluation was performed. In the objective part,

Center frequency	31.5 Hz	63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
Meas. RIR RT (s)	1.62	0.81	0.76	0.73	0.63	0.48	0.41	0.36	0.27	0.19
Synth. RIR RT (s)	1.13	1.00	0.79	0.74	0.67	0.56	0.43	0.42	0.33	0.30
Error (%)	**30.25**	**23.46**	3.95	1.37	**6.35**	**16.67**	4.87	**16.67**	**22.22**	**57.89**

(a) Office room, short reverberation, cf. Fig. 1a.

Center frequency	31.5 Hz	63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
Meas. RIR RT (s)	1.20	0.95	0.71	0.78	0.85	0.88	0.87	0.87	0.62	0.39
Synth. RIR RT (s)	1.10	0.89	0.75	0.78	0.86	0.93	0.87	0.88	0.67	0.53
Error (%)	**8.33**	**6.32**	**5.63**	0.00	1.18	**5.68**	0.00	1.15	**8.06**	**35.90**

(b) Lecture room, medium reverberation, cf. Fig. 1b.

Center frequency	31.5 Hz	63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
Meas. RIR RT (s)	2.09	2.08	2.03	2.06	2.03	2.10	1.98	1.6	0.68	0.18
Synth. RIR RT (s)	2.33	2.10	2.11	2.06	2.12	2.21	1.99	1.77	0.83	0.27
Error (%)	**11.48**	0.96	3.94	0.00	4.43	**5.23**	0.51	**10.63**	**20.06**	**50.00**

(c) Concert hall, long reverberation, cf. Fig. 1c.

Table 1: Reverberation time for measured and synthetic RIRs and the modeling error in octave frequencies for the three tested reverbs. Errors exceeding JND of 5.0% are highlighted. Cf. Fig. 1.

three RIRs measured in different venues were chosen to be reproduced with the proposed algorithm:

1. Short reverberation – RIR of an office room [24];
2. Medium reverberation – RIR of a lecture room [24];
3. Long reverberation – RIR of a concert hall in Pori, Finland.

The reverberation time (RT) values of the chosen RIRs were calculated in ten octave bands with center frequencies from 31.5 Hz to 16 kHz with the energy decay curve evaluation method [25, 26]. To ensure that all the delay lines create a meaningful contribution to the synthesized reverberation, the delay-line lengths were randomized over the range between 10 ms and 100 ms.

The RT values of RIRs produced with the FDN were compared to the target values to check whether the differences between them were exceeding 5%, which is the just noticeable difference (JND) for the RT of an acoustic impulse response [25, 27]. The spectrograms of both sets of impulse responses were also analyzed.

3.1 Results

Figure 1 presents the spectrograms of the measured (top) and synthetic (bottom) RIRs, respectively, for all the test cases, and the Table 1 shows the RT values for center frequencies of the octave bands and the error between the target and obtained RT.

For all three RIRs, the errors are biggest for the highest frequencies, 8-16 kHz, where the reverberation synthesized with the proposed algorithm is longer than the measured values. The differences are also considerable in the 31.5-Hz band. For the short and medium reverbs the obtained values are lower than the target ones, whilst for the concert hall the RT is longer.

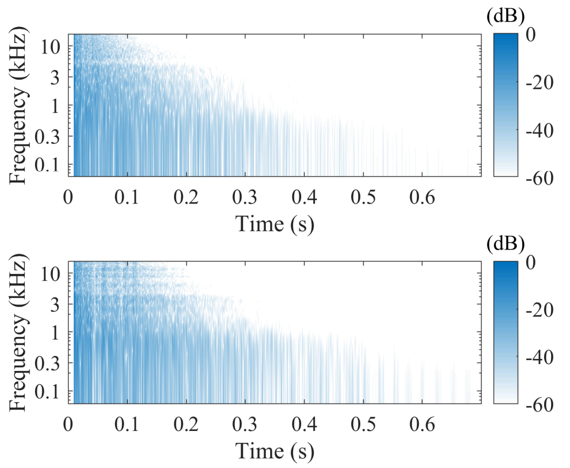
The short reverberation of the office proved to be the most problematic case in the objective evaluation. The JND of 5% is not exceeded in only a few bands, as shown in the Table 1a. The error for 500 Hz, however, is higher than the threshold by less than two percentage points.

The best results were obtained for the medium reverberation of the lecture room. Although the differences between the target and obtained values are lower than the JND for only a few bands, as presented in the Table 1b, the error exceeds 10% only at 16 kHz. In two bands, 125 Hz and 1 kHz, the JND is passed by less than one percentage point, whilst in 63 Hz the difference is only 1.32 percentage points larger than the threshold of noticeability. The spectrograms of the measured and synthesized RIRs shown in Fig. 1b. The target decay is modeled accurately for low and mid frequencies from around 100 Hz up to 6 kHz. The only exception is the slight lack of energy at about 300 Hz, from 0.55 s on.

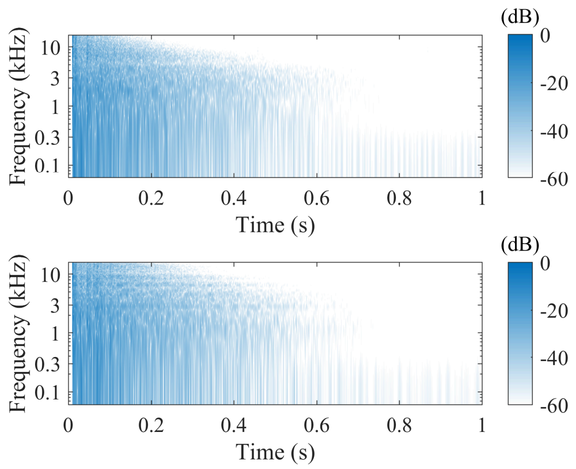
In the case of the concert hall, that the error between target and obtained values is smaller than 5% in five frequency bands, 63-500 Hz and 2 kHz whilst for 1 kHz the JND is exceeded only by 0.23%, as shown in the Table 1c. As depicted in the Fig. 1c, the decay of the measured signal is well reproduced by the synthetic one in mid frequencies, however, there is a visible overshoot in energy over 5 kHz. The modeled decay is noticeably shorter than the original one at about 300 Hz.

4. SUBJECTIVE EVALUATION

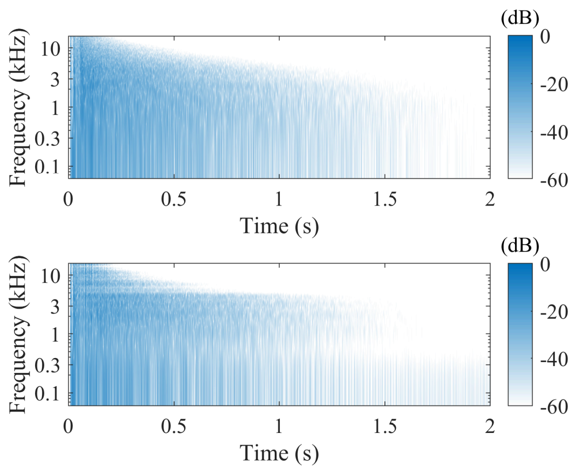
In addition to the objective evaluation of the algorithm performance, a subjective evaluation in a form of a listening test was conducted in order to assess the perceptual qualities of the synthetic reverberation.



(a) Office



(b) Lecture room



(c) Concert hall

Figure 1: Spectrograms of the (top panes) measured and (bottom panes) synthetic RIRs of the three test cases.

4.1 Stimuli and test setup

Prior to the listening test, the assumption was made that assessing the perceptual qualities of raw impulse responses would prove difficult for participants. Therefore, the test sounds were created by adding reverberation to the following anechoic recordings:

1. Speech – A 3-s sample of a female saying "The juice of lemons makes fine punch" [28];

2. Singing voice – A 49-s sample of a male singing [29], which was truncated to 10 s;

3. Guitar – A 49-s sample of guitar music [29], which was truncated to 4 s.

The truncation was performed in order to avoid tiring participants with overly long stimuli.

Every sample was convolved with each of the measured RIRs and also reverberated using the approach described in previous sections of this paper. The samples convolved with the measured RIRs were used as references, whilst the remaining sounds were the test items. Each question of the test comprised of one reference sound and a respective test sample. The task was to determine how big the audible differences in the test sounds in comparison to the reference sound were within the scope of four qualities: reverberance, clarity, distance, and coloration. The answers were presented in the form of a Likert scale, with "No difference", "Slight difference", "Clear difference", and "Strong difference" as possible responses.

The listening test was conducted in the anechoic chamber of the Aalborg University Multisensory Experience Lab, using Sennheiser HD-600 headphones. The test was carried out by using the web-based experiment software webMUSHRA developed by International Audio Laboratories Erlangen [30]. Before the test, the subjects were allowed to adjust the volume of the sound, which remained the same during the experiment. Since it is known that the loudness affects the perception of reverberance [31], the subjects were advised to keep the volume at a high level, and not reduce it below the default setting. It was also confirmed that all participants knew and understood the terminology used during the evaluation. They were familiarized with the task in a short training session, which was not included in the results.

Twelve people participated in the test. The answers of two of them, however, were dismissed due to hearing impairment reported in the post-test questionnaire. The average age of the participants whose results were analyzed was 29.8 years (SD = 5.6). All the participants were either students or employees of Aalborg University in Copenhagen. Many of them had previously participated in similar listening tests.

4.2 Listening test results

The results of the listening test are presented separately for the three reverbs (short, medium, and long) with the further division based on the type of stimulus (speech, singing

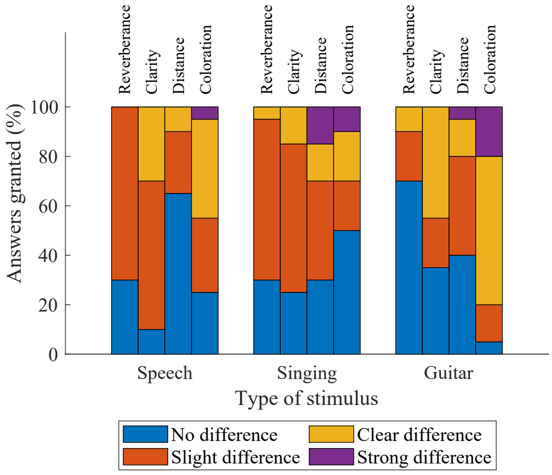


Figure 2: Listening test results for the short reverberation.

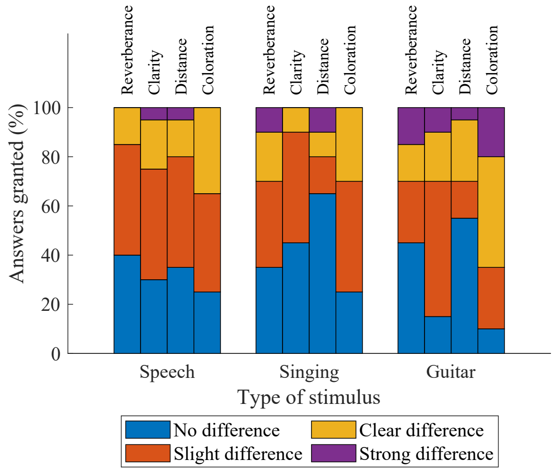


Figure 4: Listening test results for the long reverberation.

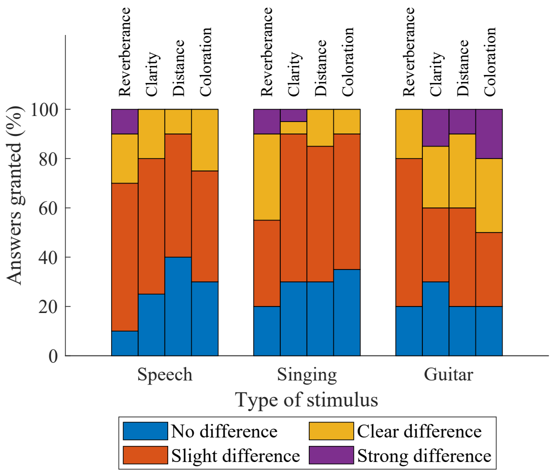


Figure 3: Listening test results for the medium reverberation.

voice, and guitar music) and assessed quality (reverberance, clarity, distance, and coloration). To ensure the ease of interpretation of the results the number of responses given to each question was converted to a percentage.

Figure 2 shows the distribution of answers for the short reverberation. Most subjects perceived the differences between the reverberance of the test and reference sounds as slight or did not notice any difference at all, regardless of the type of stimulus. In the guitar music sample, 70% of the participants chose the “No difference” answer and only a small percent of “Clear difference” responses. Clarity was evaluated similarly, however, it received fewer “No difference” and more “Clear difference” answers.

For the speech stimulus, over 60% of the subjects chose the “No difference” answer. In cases of the singing voice and guitar music more “Clear difference”, and “Strong difference” answers were given. For speech stimulus, distance received over 60% of “No difference” responses and no “Strong difference” ones. Coloration received the largest number of “Clear difference” responses for all stimuli. When accessing the reverberated singing voice, however, over 50% of participants picked the “No difference” answer.

The listening test results for the medium reverberation are presented in Fig. 3. The distribution of answers is similar for each stimulus, with the answer “Slight difference” being chosen most of the time – between 30% and 60% for each quality. The “No difference” response was granted almost as frequently, from 20% to 40% times in each question, except for reverberance in case of speech stimuli. For this reverberation, the “Clear difference” grades were granted no more than 30% of the time. Strong differences between the test and reference sounds were noticed mostly for guitar music but by no more than 20% of the participants. Three qualities for the speech stimulus – distance, clarity, and coloration - two for singing voice – distance and coloration – and reverberance for the guitar music did not receive any “Strong difference” answers.

The results for the long reverberation of the concert hall are presented in Fig. 4. For the singing voice, the “No difference” answer is the most prominent, being chosen between 30% and 60% of the time, depending on the assessed quality. For speech the percentage of the “No difference” and the “Slight difference” grades was similar, between 20% and 40%. Two out of the four qualities, reverberance and coloration for speech, and clarity and coloration for singing voice, did not receive any “Strong difference” responses.

In the case of guitar music, the distribution of the “No difference” and the “Slight difference” responses was the most uneven. Reverberance and distance were many times granted the “No difference” response, but both of them also received “Strong difference” answers. 50% of participants reported slight differences between the clarity of the assessed samples. The most “Clear difference” and “Strong

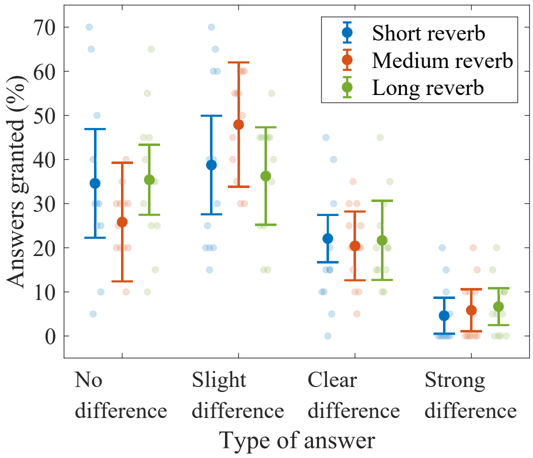


Figure 5: Listening test results analyzed for the RT values.

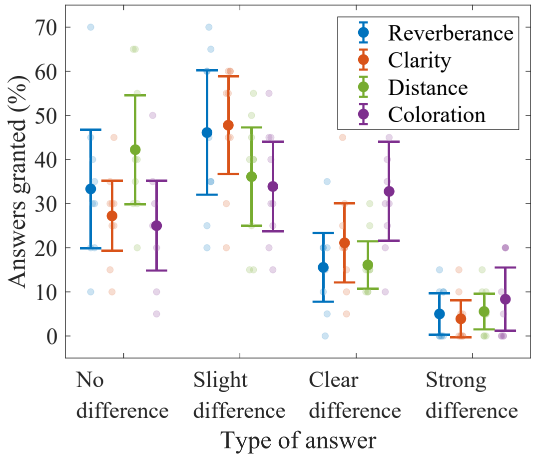


Figure 6: Listening test results analyzed for the type of perceptual sound quality.

difference" answers were given to the coloration.

4.3 Statistical analysis

The results of the listening test were further analyzed according to the RT, quality, and stimulus. Figs. 5, 6, and 7 present the average percentage of each type of answer granted to the sounds depending on the factors mentioned above. The mean grade is marked with a dot of the respective color with bars showing the 95% confidence intervals. The smaller dots with less opacity present the percentage of each type of response granted in each question of the listening test.

4.3.1 Reverberation time

Figure 5 presents the average percentage of each type of answer granted to the sounds depending on their RT values. It shows that the participants chose the "Slight difference" and "No difference" answers most frequently. The combined average percentages for those two answers were 73%, 74%, and 72% for short, medium, and long reverberation, respectively.

The analysis reflects the tendencies observed in the objective evaluation, where the algorithm performed better for long and medium reverberation than for the short one. It shows that the more accurately modeled RIRs are perceived as more similar to the measured ones even when the JND between target and modeled RT values is exceeded in some frequency bands.

4.3.2 Quality type

The analysis based on the quality type is given in Fig. 6. In the case of reverberance, clarity, and distance there are significant differences between the mean percentage of "Slight difference" and "Clear difference" answers, and between the "Clear difference" and the "Strong difference". For coloration, only the mean percentage of the "Strong difference" answers was significantly different from the others.

The analysis shows that in most cases, the "Slight Difference" or the "No difference" responses were chosen.

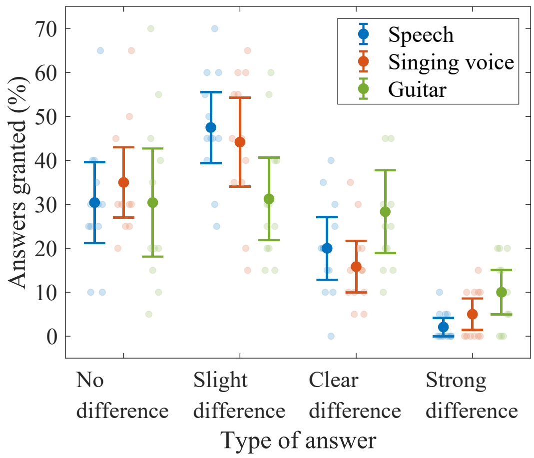


Figure 7: Listening test results analyzed for the type of the stimulus.

The combined average percentage of those two types of responses was 79%, 75%, and 78% for reverberance, clarity and distance, respectively. Distance was perceived as the quality that was the most similar in the test and reference sounds, with the "No difference" response granted in 42% questions on average. Fig. 6 shows that the results concerning coloration are the most equivocal, indicating that the algorithm still needs improvement to accurately reproduce that quality of sound.

4.3.3 Stimulus type

The results analyzed according to the type of stimulus are presented in Fig. 7. The subjects gave the most consistent grades for questions concerning speech, which resulted in all adjacent means for all types of responses for that stimuli to be significantly different. In the case of the singing voice, the mean percentage of answers "No difference" and "Slight difference" are similar, whilst for the guitar, only the mean percentage of "Strong difference" answers are significantly different from the other three. For speech the “No Difference” answer was picked 30% of the time, whilst slight dissimilarities between the references and test sounds were noticed in 48% of questions. The respective values for singing voice are 35% and 44%. For guitar music the “No difference”, “Slight difference”, and “Clear difference” were chosen around 30% of the time.

The analysis proves that the algorithm works well when reproducing the frequency range of the human voice, however, in the case of a guitar, with a lower frequency range, the improvement is needed to obtain better accuracy.

5. DISCUSSION

There are a few explanations as to why the proposed artificial reverberation algorithm reproduces the target RT most inaccurately in the high frequencies. One of them may be that the GEQ used as the attenuation filter works best when the gains are within the range of ± 12 dB. Many times the attenuation required to obtain target RT goes beyond the lower bound of that range, especially when long delay lines are used to approximate short reverbs, as presented in Fig. 8.

Another reason is that the high-shelf filter, which reduces overshoot in the decay for frequencies above 16 kHz, also introduces ripple in the filter’s magnitude response, as was reported in [8]. This phenomenon is shown in Fig. 8, where the drop and rise in the magnitude are present at high frequencies in all three attenuation filters.

Similarly, the undershoot in the RT values observed in the 31.5 Hz band may be the consequence of shifting and scaling of the attenuation filter’s magnitude response. Those operations create a decrease in the magnitude response for very low frequencies, as shown in Fig. 8, which may lower the algorithm’s accuracy.

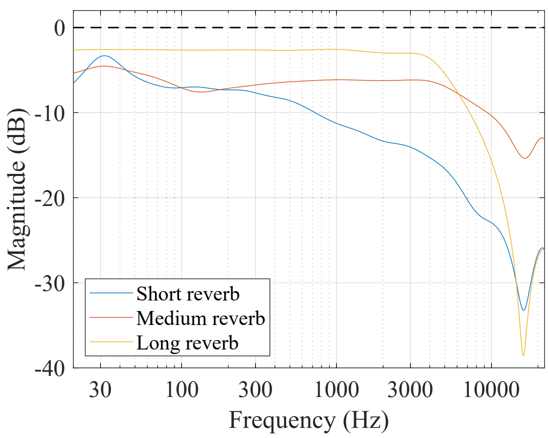


Figure 8: The attenuation filter’s magnitude response for the three test cases of short, medium, and long reverberation and a delay-line length of 90 ms.

6. CONCLUSIONS

The present work studied the ability of the FDN with the modified cascaded GEQ as the attenuation filter to reproduce the measured RIRs accurately. The evaluation was conducted both by comparing the RT values and decay characteristics of the original and synthetic RIRs and by the means of the listening test.

In the objective assessment the RIRs produced with the proposed method replicated the target decay best in the mid-frequency range between 125 Hz and 2 kHz. For each RIR type, the biggest dissimilarities between the target and obtained RT values occurred in the 16 kHz band. The results show that the proposed design performs best when long reverberation times are modeled, and that small differences between the RT values in the neighboring frequency bands produce more accurate approximation.

The listening test showed that for the three types of reverberation and stimuli, when the four qualities of sound were assessed, the subjects mostly perceived only slight differences between the sounds convolved with measured and synthetic RIRs. Many times the “No difference” answer was chosen as well. The differences were easiest to notice for the coloration in case of the sound quality and the guitar music in terms of the stimulus. However, further testing is needed to establish whether the test sounds were truly indistinguishable from the references.

Accurately reproducing impulse responses of real spaces with parametric artificial reverberation is still a difficult task. The present study shows that with the proposed FDN design in many cases it is possible to trick the human perception into not noticing dissimilarities between the original and artificially produced signals. However, this study suggests that the FDN reverberator can still be improved by further developing the accuracy of the attenuation filter. Additionally, future work may look into the choice of the feedback matrix and delay lengths, which were not considered in this work.

Acknowledgments

This work was supported by the “Nordic Sound and Music Computing Network—NordicSMC”, NordForsk project number 86892. This work was initialized when Karolina Prawda made a Short-Term Scientific Mission to the Aalborg University Copenhagen on October 28 – November 15, 2019.

7. REFERENCES

[1] J. M. Jot and A. Chaigne, “Digital delay networks for designing artificial reverberators,” in *Proc. 90th Audio Eng. Soc. Convention*, Paris, France, Feb. 19–22, 1991.

[2] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, “Fifty years of artificial reverberation,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 5, pp. 1421–1448, Jul. 2012.

[3] J. Moorer, “About this reverberation business,” *Computer Music J.*, vol. 3, no. 2, pp. 13–28, 1979.

[4] J. M. Jot, “Efficient models for reverberation and distance rendering in computer music and virtual audio

reality," in *Proc. Int. Computer Music Conf.*, Thessaloniki, Greece, Sept. 1997.

[5] T. Wendt, S. van de Par, and S. D. Ewert, "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *J. Audio Eng. Soc.*, vol. 62, no. 11, pp. 748–766, Nov. 2014.

[6] J.-M. Jot, "Proportional parametric equalizers–Application to digital reverberation and environmental audio processing," in *Proc. 139th Audio Eng. Soc. Conv.*, New York, USA, Oct. 29–Nov. 1, 2015.

[7] S. J. Schlecht and A. P. Habets, "Accurate reverberation time control in Feedback Delay Networks," in *Proc. Digital Audio Effects (DAFx-17)*, Edinburgh, UK, Sept. 5–9, 2017, pp. 337–344.

[8] K. Prawda, S. J. Schlecht, and V. Välimäki, "Improved reverberation time control for feedback delay networks," in *Proc. Int. Conf. Digital Audio Effects*, Birmingham, UK, Sept. 2019.

[9] A. Czyżewski, "A method of artificial reverberation quality testing," *J. Audio Eng. Soc.*, vol. 38, no. 3, pp. 129–141, Mar. 1990.

[10] B. Katz, D. Poirier-Quinot, B. Postma, D. Thery, and P. Luizard, "Objective and perceptive evaluations of high-resolution room acoustic simulations and auralizations," in *Proc. Euronoise 2018*, Heraklion, Crete, Greece, 27–31 May 2018, pp. 2107–2114.

[11] P. Stade and J. M. Arend, "Perceptual evaluation of synthetic late binaural reverberation based on a parametric model," in *Proc. Audio Eng. Soc. Int. Conf. Headphone Technology*, Aalborg, Denmark, Aug. 2016.

[12] M. Steimel, "Implementation of a hybrid reverb algorithm. Parameterizing synthetic late reverberation from impulse responses," Master's thesis, Aalborg University, Denmark, 2019.

[13] B. Alary, A. Politis, S. J. Schlecht, and V. Välimäki, "Directional feedback delay network," *J. Audio Eng. Soc.*, vol. 67, no. 10, pp. 752–762, Oct. 2019.

[14] S. J. Schlecht and A. P. Habets, "On lossless Feedback Delay Networks," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1554–1564, Mar. 2017.

[15] V. Välimäki and J. Liski, "Accurate cascade graphic equalizer," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 176–180, Feb. 2017.

[16] V. Välimäki and J. Reiss, "All about audio equalization: Solutions and frontiers," *Appl. Sci.*, vol. 6, no. 5, May 2016.

[17] A. Primavera, S. Cecchi, J. Li, and F. Piazza, "Objective and subjective investigation on a novel method for digital reverberator parameters estimation," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 441–452, Feb. 2014.

[18] G. Defrance and J.-D. Polack, "Measuring the mixing time in auditoria," *J. Acoust. Soc. Am.*, vol. 123, no. 5, p. 3499, May 2008.

[19] J. Abel and P. Huang, "A simple, robust measure of reverberation echo density," in *Proc. AES 121st Conv.*, San Francisco, CA, USA, Oct. 2006.

[20] P. Huang and J. Abel, "Aspects of reverberation echo density," in *Proc. AES 123rd Conv.*, New York, NY, USA, Oct. 2007.

[21] P. Huang, J. S. Abel, H. Terasawa, and J. Berger, "Reverberation echo density psychoacoustics," in *Proc. AES 125th Conv.*, San Francisco, CA, USA, Oct. 2009.

[22] R. Stewart and M. Sandler, "Statistical measures of early reflections of room impulse responses," in *Proc. Int. Conf. Digital Audio Effects (DAFx-06)*, Bordeaux, France, Sept. 2007, pp. 213–218.

[23] R. Stewart, "Hybrid convolution and filterbank artificial reverberation algorithm using statistical analysis and synthesis," Master's thesis, The University of York, York, UK, 2006.

[24] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proc. Int. Conf. Digital Signal Process. (DSP)*, Santorini, Greece, Jul. 2009, pp. 1–4.

[25] ISO, "ISO 3382-2, Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces," International Organization for Standardization, Geneva, Switzerland, Tech. Rep., 2009.

[26] ——, "ISO 3382-2, Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms," International Organization for Standardization, Geneva, Switzerland, Tech. Rep., 2009.

[27] M. Karjalainen and H. Järveläinen, "More about this reverberation science: Perceptually good late reverberation," in *Proc. 111th Audio Eng. Soc. Conv.*, New York, USA, Sept. 21–24, 2001.

[28] P. Kabal, "TSP speech database," Department of Electrical and Computer Engineering, McGill University, Montreal, Canada, Tech. Rep., 2018.

[29] B. Bernschütz, "Anechoic recordings," Cologne University of Applied Sciences, Institute of Communication Systems, Cologne, Germany, Tech. Rep., 2013.

[30] M. Schoeffler, S. Bartoschek, F.-R. Stöter, M. Roess, S. W. B. Edler, and J. Herre, "WebMUSHRA. A comprehensive framework for web-based listening tests," *J. Open Research Software*, vol. 6, no. 1, p. 8, Feb. 2018.

[31] D. Lee, D. Cabrera, and W. L. Martens, "The effect of loudness on the reverberance of music: Reverberance prediction using loudness models," *J. Acous. Soc. Am.*, vol. 131, no. 2, pp. 1194–1205, Feb. 2012.