Audio Engineering Society

Convention Paper 10226

Presented at the 147th Convention
2019 October 16 – 19, New York

This convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Modifying Audio Signals for Reproduction with Reduced Room Effect

Christof Faller[1]

[1]*Illusonic GmbH, Bahnstrasse 23, 8610 Uster, Switzerland*

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

ABSTRACT

Conventionally, equalizers are applied when reproducing audio signals in rooms, to reduce coloration and effect of room resonances. Another approach, filtering audio signals with an inverse of the room impulse response (RIR), can theoretically eliminate the effect of the room in one point. But practical issues arise, such as impaired sound at other positions, a need to update when RIRs change, and loudspeaker-challenging signals. A technique is presented, which modifies the time-frequency envelopes (spectrogram) of audio signals, such that the corresponding spectrogram in the room is more similar to the original signal's spectrogram, i.e. room effect is attenuated. The proposed technique has low sensitivity on RIR and listener position changes.

1 Introduction

When sound is reproduced in rooms, it arrives at a listener position spectrally and temporally modified. Some effects of early reflections and reverb are desirable for music reproduction, such as increased spatial impression [1]. Other effects are undesirable, such as excessive coloration from early reflections [2], masking from the reverberant decay [3], and overly slow reverberant decay at resonance frequencies. Usually rooms in private homes are not sufficiently acoustically treated and the undesirable room effects weigh high. This motivates adaptation of audio signals to rooms in a way to reduce the undesirable effects.

A simple way of adapting audio signals to rooms is the use of an equalizer to modify the magnitude spectrum of the signal, such that at listening position(s) spectral distortions are 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].

But the strategy on how to determine the equalization curve can be complex. Shall a single sweet spot be optimized or an average over multiple positions [6, 7]? What are reasonable limits in spectral detail and equalization gain [8, 9]?

Acoustic treatment reduces a need for equalization. Also, loudspeaker and listening positions can be optimized to reduce spectral variation due to the room [8, 7, 10].

A more complex way of adapting audio signals to rooms is to filter them with the inverse of the loudspeaker-sweetspot room impulse response (RIR). In theory, this approach eliminates the room effect entirely in one point. RIRs are changing constantly as temperature and humidity is changing [11]. Any movement of a person in the room also gives rise to RIR changes. Thus, an inverse filter has to be adapted all the time, which is unpractical. Determination of more robust and magnitude-limited inverse filters, allowing some inversion error, has been the topic of many investigations, e.g. [12, 13]. An inherent limitation of all approaches based on filter-inversion is, that they essentially only work in the sweetspot. At other positions, sound is more colored and reverberated than without inverse filtering.

Also, with inverse RIR filtering, the loudspeaker reproduces highly modified signals, potentially leading to an increase in harmonic and modulation distortions [14].

RIR filtering can be designed to achieve a less ambitious goal, e.g. inverting the minimum phase component of RIR [15] and shortening RIR at modal frequencies [16]. But also these approaches are highly dependent on listening position.

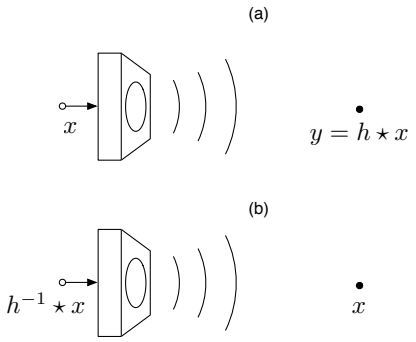


Fig. 1: (a) A loudspeaker reproducing signal x which arrives at the listening position as signal y. (b) Ideal corresponding reproduction, eliminating effect of the room.

In the following, we are motivating the approach proposed in this paper. Figure 1(a) illustrates a setup reproducing signal x, which arrives at the listening position as signal $y = h \star x$ (x filtered with the corresponding RIR).

An ideal sound reproduction with inverse-RIR-filtering is illustrated in Figure 1(b). The inverse-RIR-filtered signal $h^{-1} \star x$ is given to the loudspeaker and the corresponding listening position signal is x, i.e. there is no room effect.

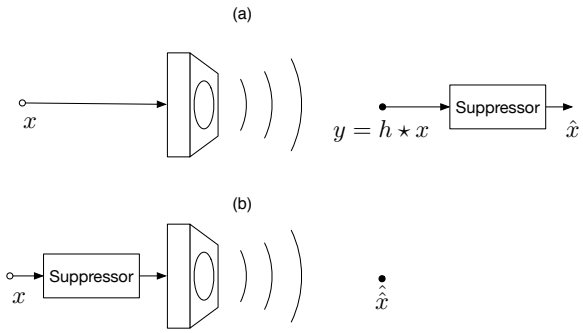


Fig. 2: (a) A reverb suppressor applied to signal with room effect y yields a signal $\hat{x}$ with reduced room effect. (b) The proposed method applies a suppressor to signal x prior to reproduction with the goal that the corresponding signal at the listener position $\hat{x}$ has reduced room effect.

As mentioned above, this approach is highly sensitive to RIR and listener position changes. Consider the setup shown in Figure 2(a). If a time-frequency adaptive reverb suppressor [17, 18, 19, 20] is applied to the listener position signal y, a signal with reduced room effect $\hat{x}$ is generated. A main challenge and limiting factor for performance of reverb suppressors is estimation of the reverb [21, 22, 23, 24] to be suppressed. The reverb suppressor used here is expected to work well, since the RIR can be measured and reverb is thus known.

The characteristics of the reverb (direct-to-reverberant ratio, reverberation time) are changing much less as one changes the listener position than the waveform details of the RIR. Thus a reverb suppressor optimized for the listening position gives useful results in a larger listening area than RIR-inversion.

On the other hand, the setup of Figure 2(a) is not useful for listening, since it is not possible to apply reverb suppression before sound enters the listener's ears.

Therefore, we are considering the setup shown in Figure 2(b). A time-frequency adaptive suppressor is applied to signal x before it is given to the loudspeaker. The suppressor shall operate such that the corresponding listener position signal $\hat{x}$ has reduced room effect. The reverb suppressor used in Figure 2(a) is not directly suitable, since signal x has the room reverb not yet added. The reverb is added later as x travels through the room to the listener position. This paper is about the design of this suppressor, applied to signals before given to loudspeakers.

The paper is organized as follows. Section 2 describes the proposed suppressor based on "spectrogram inversion". Computation of the spectrogram inversion filters, based on measured RIRs, is described in Section 3. The proposed techniques are illustrated with a music signal and a measured living room RIR in Section 4. Section 5 discusses equalizing versus proposed processing, sensitivity to listener position, stereo audio reproduction, and informal testing. Conclusions are in Section 6.

2 Optimizing Audio Signals for Room Reproduction

2.1 RIR Spectrogram Inversion

Consider an audio signal x which is reproduced over a loudspeaker in a room. The corresponding signal y at a specific listening position is

$$y(n) = h(n) \star x(n)\,, \tag{1}$$

where n is a time index, h is a room impulse response (RIR), and $\star$ is linear convolution. The time index n relates to time in seconds by n/f_s, where f_s is the sampling frequency of the time discrete signals.

A signal x in a short-time Fourier transform (STFT) domain is denoted $X(k,i)$, where k and i are time and frequency indices. Indices k and i relate to time in seconds and frequency in Hertz, respectively, by kH/f_s and if_s/F, where H and F are STFT hopsize and FFT size, respectively.

The power spectra (spectrograms) of x, y, and h are

$$\begin{aligned} P_X(k,i) &= |X(k,i)|^2 \\ P_Y(k,i) &= |Y(k,i)|^2 \\ P_H(k,i) &= |H(k,i)|^2\,. \end{aligned} \tag{2}$$

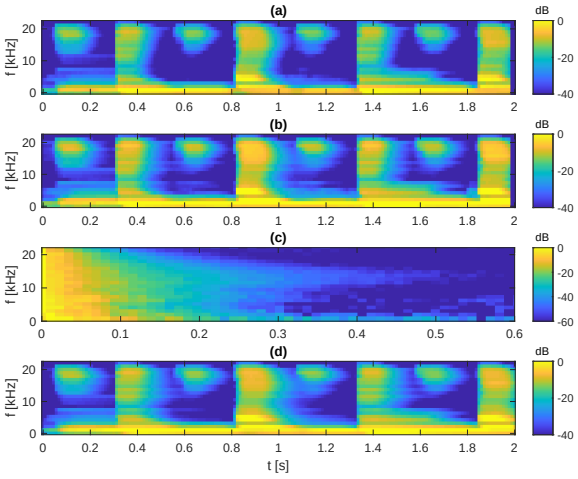


Fig. 3: Spectrograms of (a) loudspeaker signal P_X, (b) listener position signal P_Y, (c) RIR P_H, and (d) estimated spectrogram at listener position $P_H \star P_X$.

The RIR spectrogram $P_H(k,i)$ is causal, i.e. $P_H(k,i) = 0$ for $k < 0$. Panels (a) to (c) in Figure 3 show example signal and RIR spectrograms.

The spectrogram of y can be approximated by

$$P_Y(k,i) \approx P_H(k,i) \star P_X(k,i)\,, \tag{3}$$

where $\star$ is convolution with respect to k. Figure 3(d) shows an example of estimated P_Y. For the purpose of designing the proposed suppressor, we found that the spectrogram approximation (3) is sufficiently precise[1].

The inverse spectrogram P_H^{-1} is defined such that

$$P_H^{-1}(k,i) \star P_H(k,i) = P_H(k,i) \star P_H^{-1}(k,i) = \delta(k)\,, \tag{4}$$

where δ is the Kronecker-Delta. When instead of x, a signal x' with spectrogram

$$P_{X'}(k,i) = P_H^{-1}(k,i) \star P_X(k,i)\,, \tag{5}$$

is reproduced, then the corresponding spectrogram at listener position $P_{Y'}$ (3) approximates P_X.

[1] We are using a listener position spectrogram estimation filter P_H, directly derived from the corresponding RIR (2). This method gives useful P_H filters which reasonably approximate spectrograms at listener positions from loudspeaker signal spectrograms. It may be possible to compute P_H for approximating listener position spectrograms more precisely, e.g. determining P_H as least-mean-squares estimators.

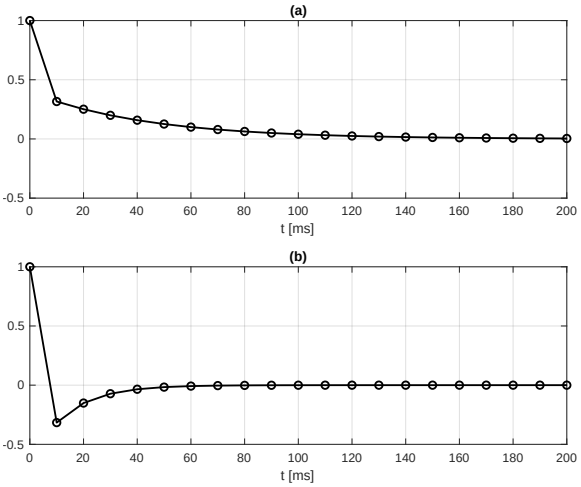


Fig. 4: (a) Example RIR spectrogram P_H at one frequency and (b) its inverse P_H^{-1}.

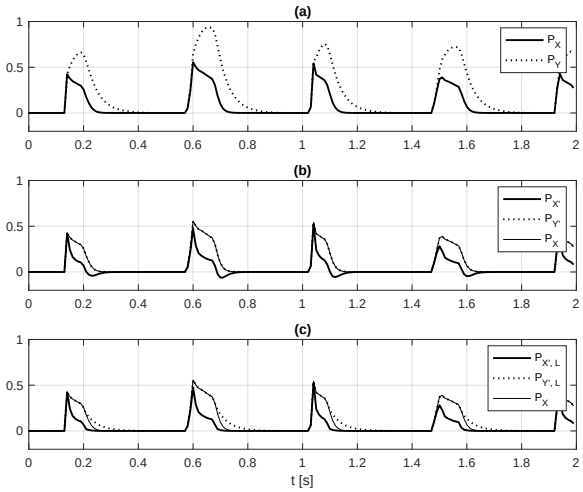


Fig. 5: (a) Loudspeaker and listener position power P_X and P_Y; (b) modified $P_{X'}$, corresponding $P_{Y'}$, and target (P_X); (c) same as (b) but using an attenuation limit of $L = -10$ dB.

Figure 4 shows for purpose of example a RIR spectrogram P_H at one frequency and its inverse P_H^{-1}, corresponding to a reverberation time of RT60 $= 0.6$ s. Using these, the data in Figure 5 has been computed. Panel (a) shows an example of loudspeaker signal spectrogram P_X at 2 kHz and corresponding listener position spectrogram P_Y. The middle panel shows the modified loudspeaker signal spectrogram $P_{X'}$ (8) and the corresponding listener position spectrogram $P_{Y'}$. It can be seen that $P_{Y'} = P_H \star P_{X'}$ is equal to P_X, as desired.

Note that $P_{X'}$ can be negative, meaning that negative signal power would need to be reproduced to achieve spectrogram P_X at the listener position. All that can be done at such a time and frequency, is, to emit zero signal power.

Signal x' is computed in the STFT domain as

$$X'(k,i) = W(k,i)X(k,i)\,, \tag{6}$$

where W are the time-frequency adaptive gains of the suppressor [2],

$$W(k,i) = \max\left\{\frac{P_{X'}^{\beta}(k,i)}{P_X^{\beta}(k,i)},\ 10^{\frac{L}{10}}\right\}, \tag{7}$$

[2] In noise suppression literature W is denoted "gainfilter".

L is an attenuation limit in dB, and β controls the aggressivity of the suppression [25]. Here and in the following we use $L = -10$ dB and $\beta = 1$.

The spectrogram of signal x' (6), computed with the attenuation limit L, is

$$P_{X',\,L}(k,i) = \max\left\{P_{X'}(k,i),\ P_X(k,i)10^{\frac{L}{10}}\right\}. \tag{8}$$

The bottom panel of Figure 5 shows the loudspeaker $P_{X',\,L}$ with $L = -10$ dB and corresponding listener position $P_{Y',\,L}$, which now slightly deviates from P_X.

Purposely an extreme transient case of signal x has been chosen for the data shown in Figure 5. The same data for a less extreme case is shown in Figure 6. Here, with attenuation limit, Panel (c), $P_{Y',\,L}$ is virtually identical to P_X.

2.2 Target RIR

The target RIR spectrogram does not have to be unity, as above considered. Other targets $P_{H'}$ may be used. For example

$$P_{H'}(k,i) = \begin{cases} P_H(k,i) & \text{if } k \le \text{round}(\frac{kH\tau}{f_s}) \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

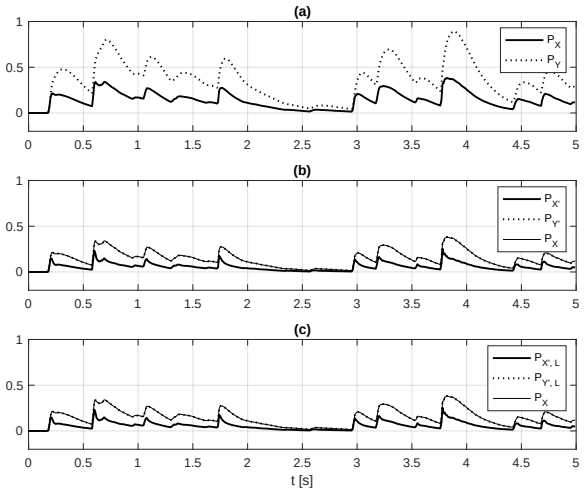


Fig. 6: (a) Loudspeaker and listener position power P_X and P_Y; (b) modified $P_{X'}$, corresponding $P_{Y'}$, and target (P_X); (c) same as (b) but using an attenuation limit of $L = -10$ dB.

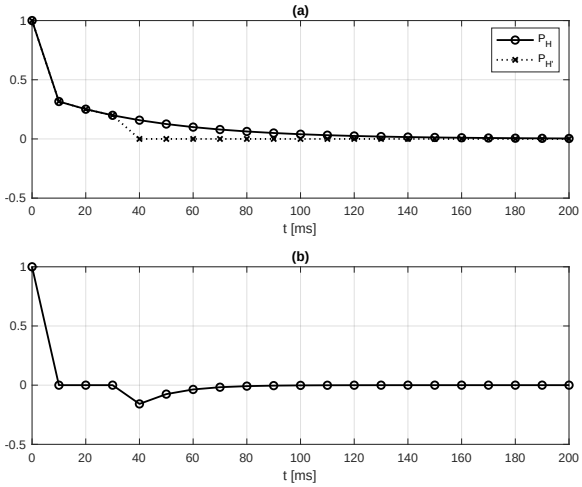


Fig. 7: Example (a) RIR spectrogram P_H at one frequency, (b) RIR target $P_{H'}$, and (c) P_I.

contains early reflections up to τ seconds.

The spectrogram filter P_I removing reverberation beyond τ in P_H, i.e.

$$P_I \star P_H = P_{H'}\,, \tag{10}$$

is

$$P_I = P_H^{-1}(k,i) \star P_{H'}(k,i)\,. \tag{11}$$

When a signal x' with spectrogram

$$P_{X'}(k,i) = P_I(k,i) \star P_X(k,i)\,, \tag{12}$$

is reproduced, then the spectrogram at listener position (3) is approximately equal to $P_{H'}(k,i) \star P_X$, i.e. only reverb beyond τ seconds is attenuated.

Figure 7 shows an example of P_H, $P_{H'}$, and P_I. The example of Figure 5 is shown again with $\tau = 30$ ms in Figure 8, also corresponding to the responses in Figure 7.

3 RIR Spectrogram Modeling and Inversion

3.1 Modeling

When requiring that P_H should be inverted with a causal filter, then P_H^{-1} is in many cases instable, i.e. its impulse response magnitude grows with time. Non-causal P_H^{-1} are more stable but require delay in the signal processing.

Precise inversion of P_H goes against the goal of optimizing sound over a larger listening area. Therefore, at the cost of being less effective at an optimal listening point we model RIR spectrograms without detailed early reflection pattern. In this sense, the spectrogram of a RIR is modeled as

$$P_H(k,i) = \delta(k)(1-g) + g r^k\,, \tag{13}$$

with

$$\begin{aligned} g &= 10^{\frac{G}{10}} \\ r &= 10^{\frac{-60kH}{10 f_s \mathrm{RT60}}}\,, \end{aligned} \tag{14}$$

where G is reverb gain in dB and RT60 is reverberation time. The top panel of Figure 4 shows an example P_H with $G = -4$ dB and RT60 $= 0.6$ s.

The inverse of (13) is causal,

$$P_H^{-1}(k,i) = \begin{cases} 1 & \text{if } k = 0 \\ -gr & \text{if } k = 1 \\ P_H^{-1}(k-1,i)(1-g)r & \text{if } k > 1 \end{cases}\,. \tag{15}$$

The corresponding inverse of the example shown in the top panel of Figure 4 is shown in the bottom panel.

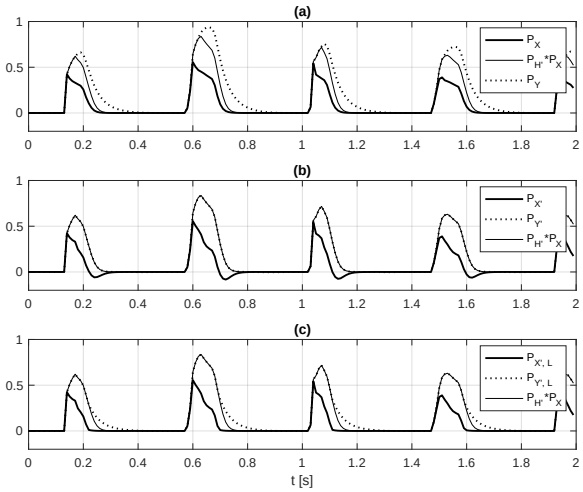


Fig. 8: (a) Loudspeaker P_X, target $P_{H'}(k,i) \star P_X$, and listener position P_Y; (b) modified $P_{X'}$, corresponding $P_{Y'}$, and target $(P_{H'}(k,i) \star P_X)$, (c) same as (b) but using an attenuation limit of $L = -10$ dB.

3.2 Inversion

The goal is to find a filter P_I (11), which effectively attenuates the tail of a measured P_H. This filter P_I is then applied to the audio signal spectrogram P_X to obtain the suppressor target spectrogram $P_{X'}$ (12), from which the suppressor gains W (7) are computed.

P_I is computed based on a modeled RIR (13) with parameters early reflections time τ (9), reverb gain G, and reverberation time RT60. The question is, which values of these parameters effectively attenuate the tail of the measured P_H.

In the following experiments, τ, G, and RT60 are determined at each frequency i as:

1. RIRs are shifted to the left, when needed, to remove unnecessary delay.

2. A simple rule to determine early reflections time τ, such that they capture a good portion of RIR power, is

$$\tau(i) = \arg\max_k P_H(k,i) kH/f_s + 0.02 \text{ s}. \quad (16)$$

That is, τ is chosen to be 20 ms after the position of the maximum of the RIR power at each frequency.

3. RT60 is determined from P_H using the Schroeder method [26].

4. Using the previously determined τ and RT60, G is determined such that the L_1 norm considered along time index k of

$$P_\delta(k,i) = |P_I(k,i) \star P_H(k,i)| \quad (17)$$

is minimized, where $|.|$ is the absolute value of a number. This is implemented by comparing this norm computed for different G in 0.1 dB steps. The G with the smallest norm is used to model the measured RIR spectrogram.

4 Experiments

4.1 Inversion of Measured RIR

Panel (a) in Figure 9 shows a RIR spectrogram P_H of a living room. The corresponding modeled P_H (13) is shown in Panel (b), whereas τ, G, and RT60 were determined at each frequency as described above.

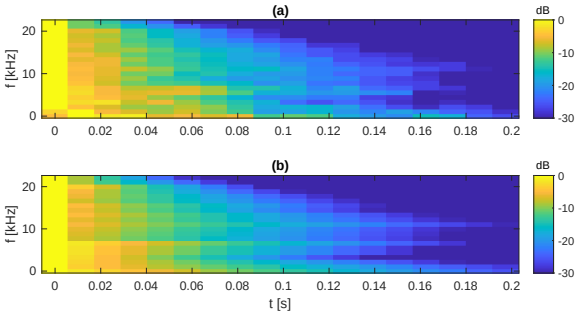


Fig. 9: (a) Measured RIR spectrogram P_H and (b) corresponding modeled RIR.

Panel (a) in Figure 10 shows the same measured RIR spectrogram P_H again and Panel (b) the corresponding inverted spectrogram $P_I \star P_H$. The inverted spectrogram has negative values. These have been set to to -30 dB in the figure. Panel (c) shows the magnitude of these negative values.

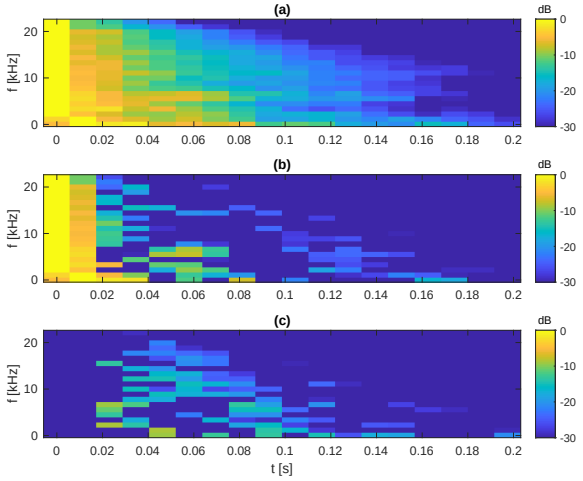


Fig. 10: (a) Measured RIR spectrogram P_H, (b) corresponding inverted spectrogram $P_I \star P_H$, (c) magnitude of negative values of $P_I \star P_H$.

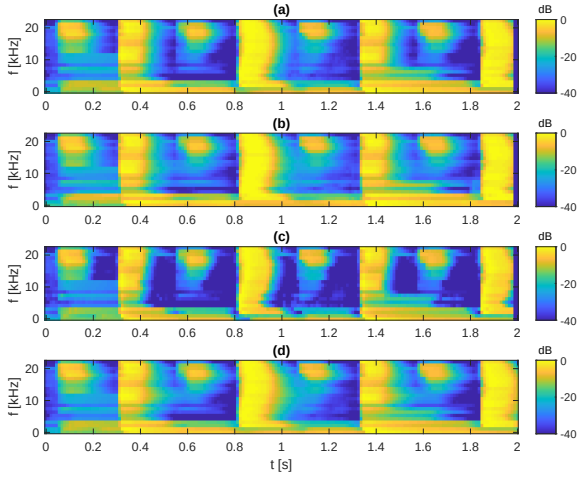


Fig. 11: (a) Spectrograms of loudspeaker signal P_X, (b) listener position signal P_Y, (c) modified loudspeaker signal $P_{X'}$, and (d) corresponding listener position signal $P_{Y'}$.

The inversion of the measured RIR based on the inverse of a modeled RIR effectively works. The reverb decay is clearly reduced in $P_I \star P_H$ compared to P_H. The residual of the RIR tail beyond τ is expected to be zero-mean and so obviously positive and negative values occur.

4.2 Adapting music signal to RIR

The music signal and RIR, which were used for the data in Figures 3, 9, and 10, were used again for generating the data in Figure 11. The spectrogram of the music signal P_X is shown in Panel (a). The corresponding spectrogram at the listener position P_Y is shown in Panel (b). Comparing these data, the effects of the room reverb are slower signal decays and amplified stationary signal portions.

We are interested in the temporal features of the signals and to make comparison in this respect clearer, we normalized these and following spectrograms such that their maximum at each frequency is 0 dB.

Panel (c) of Figure 11 shows the spectrogram $P_{X'}$ (8) of the audio signal X' (6) adapted to the acoustic. The time-frequency suppressor gains W (7) to generate X' from X are shown in Figure 12.

Finally, Panel (d) of Figure 11 shows the spectrogram at listener position when signal X' is reproduced by the loudspeaker. Comparing spectrograms of Panels (b) and (d) to the music signal spectrogram in Panel (a), indicates that spectrogram (d) is more similar. That is, the proposed processing reduces room effect in this case.

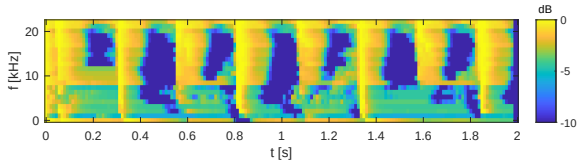


Fig. 12: Time-frequency gains W of suppressor to generate room adapted loudspeaker signal x'.

5 Discussion

5.1 Equalizing versus Proposed Suppressor

The spectrogram inversion filter P_I (11) is determined in Section 3.2 based on a normalized RIR model (13)

(direct/early sound $= 0$ dB). Therefore, the proposed processing does not modify direct/early sound level at each frequency, but merely attenuates reverb.

Additionally, if τ is chosen such that most of the RIR power is captured as early reflections and thus maintained, as the method in Section 3.2 does in most cases, then the attenuation of the reverb tail does modify overall frequency response of the audio reproduction only modestly. That is, the proposed scheme, when used like this, can be used in addition to conventional equalizing.

5.2 Sensitivity to listener position

The proposed time-frequency adaptive suppressor preprocesses signals such that reverb at a specific listener position is attenuated. When the listener is moving, as long as direct-to-reverberant ratio is not significantly changed, the suppressor stays effective.

When the listener is moving farther away, where direct-to-reverberant ratio is smaller, the suppressor still reduces reverb, just less as it would if the suppressor would have been optimized for this case.

When the listener is closer and the direct-to-reverberant ratio is larger, then the suppressor acts on the signal more than necessary. When the music signal itself contains reverberation, the listener will hear it as slightly dereverberated. For dry music signals, some unnatural modifications may occur, such as unnaturally decaying crash cymbals.

5.3 Stereo

For stereo reproduction, the RIR spectrogram is computed as the sum (of power) of left and right RIR spectrograms. From this data, P_I is computed. The suppressor gains are computed based on the sum of left and right signal spectrograms. These gains are then applied separately to the left and right signals.

5.4 Informal evaluations

We have tested the proposed scheme extensively, but without yet a subjective test. After several years of development, it runs on a DSP in a digital HiFi preamplifier. A measurement system supplies the needed RIR data to the algorithm. Feedback from professionals and ordinary listeners has been very positive. The more a listener is trained or focused, the more he perceives the sometimes subtle attenuation of room reverb. Most noticeable, in many cases, bass becomes obviously more precise and dry. The effect in large venues is often more dramatic, because explicitly perceptible room reverb is attenuated.

6 Conclusions

Equalizing can be used to reduce coloration of sound reproduction in rooms. Besides the coloration, other temporal room effects are not reduced by equalizing. A time-frequency adaptive suppressor has been described, which modifies the reproduced signal spectrogram such that the spectrogram in the specific room is more similar to the signal spectrogram. A linear estimator, estimating room signal spectrogram from original signal spectrogram is inverted and applied to signal spectrogram to obtain the spectrogram that shall be reproduced. The time-frequency adaptive suppressor modifies the signal to be reproduced to have this spectrogram and so reduces temporal room effects not addressed by equalizing.

Acknowledgments

The author would like to thank Mihailo Kolundzija for fruitful discussions on early stage of this project.

References

[1] Mason, R., *Elicitation and measurement of auditory spatial attributes in reproduced sound*, Ph.D. thesis, University of Surrey, 2002, a review of existing measurements that relate to spatial impression.

[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] Elliott, S. J. and Nelson, P. A., "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *J. Audio Eng. Soc*, 37(11), pp. 899–907, 1989.

[7] Toole, F., "The Measurement and Calibration of Sound Reproducing Systems," *J. Audio Eng. Soc*, 63(7/8), pp. 512–541, 2015.

[8] Groh, A. R., "High-Fidelity Sound System Equalization by Analysis of Standing Waves," *J. Audio Eng. Soc*, 22(10), pp. 795–799, 1974.

[9] Fielder, L. D., "Practical Limits for Room Equalization," in *Audio Engineering Society Convention 111*, 2001.

[10] Welti, T. and Devantier, A., "Low-Frequency Optimization Using Multiple Subwoofers," *J. Audio Eng. Soc*, 54(5), pp. 347–364, 2006.

[11] 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.

[12] Kirkeby, O. and Nelson, P. A., "Digital Filter Design for Inversion Problems in Sound Reproduction," *J. Audio Eng. Soc*, 47(7/8), pp. 583–595, 1999.

[13] Kyriakakis, C. and Bharitkar, S., "Signal Processing Methods for Room Equalization," in *Audio Engineering Society Conference: UK 23rd Conference: Music Everywhere*, 2008.

[14] Klipsch, P. W., "Modulation Distortion in Loudspeakers," *J. Audio Eng. Soc*, 17(2), pp. 194, 196, 198, 200, 202, 204, 206, 1969.

[15] Mourjopoulos, J. N., "Digital Equalization of Room Acoustics," *J. Audio Eng. Soc*, 42(11), pp. 884–900, 1994.

[16] Karjalainen, M., Antsalo, P., and Makivirta, A., "Modal Equalization by Temporal Shaping of Room Response," in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, 2003.

[17] Langhans, T. and Strube, H., "Speech enhancement by nonlinear multiband envelope filtering," in *Proc. IEEE ICASSP*, 1982.

[18] Lebart, K. and Boucher, J. M., "A New Method Based on Spectral Subtraction for Speech Dereverberation," *Acta Acustica*, 87, pp. 359–366, 2001.

[19] Wu, M. and Wang, D., "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, 14(3), pp. 774–784, 2006.

[20] Habets, E., *Single and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Technische Universiteit Eindhoven, Netherlands, 2007.

[21] Habets, E. A. P., Gannot, S., and Cohen, I., "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Lett.*, 16(9), pp. 770–773, 2009, doi:10.1109/LSP.2009.2024791.

[22] Habets, E. A. P., *Speech dereverberation using statistical reverberation models*, chapter 3, pp. 57–93, Springer London, 2010, doi:10.1007/978-1-84996-056-4_3.

[23] Schwartz, O., Gannot, S., and Habets, E. A. P., "Joint estimation of late reverberant and speech power spectral densities in noisy environments using frobenius norm," in *Proc. European Signal Processing Conf. (EUSIPCO)*, pp. 1123–1127, IEEE, 2016.

[24] Braun, S., Kuklasinski, A., Schwartz, O., Thiergart, O., Habets, E. A. P., Gannot, S., Doclo, S., and Jensen, J., "Evaluation and comparison of late reverberation power spectral density estimators," *IEEE Trans. Audio, Speech, Lang. Process.*, 26(6), pp. 1052–1067, 2018.

[25] Vary, P., "Noise suppression by spectral magnitude estimation-mechanism and theoretical limits," *Signal Processing*, 8, pp. 387–400, 1985.

[26] Schroeder, M. R., "Integrated-Impulse method measuring sound decay without using impulses," *J. Acoust. Soc. Am.*, 66(2), 1979.