IMPULSE RESPONSE SHORTENING FOR ACOUSTIC LISTENING ROOM COMPENSATION

Markus Kallinger and Alfred Mertins

University of Oldenburg, School of Mathematics and Natural Sciences
D-26111 Oldenburg, Germany
{markus.kallinger, alfred.mertins}@uni-oldenburg.de

ABSTRACT

The objective of this paper is to investigate the usability of channel shortening approaches known from data transmission for the equalization of acoustic systems. In setups for data transmission, the equalizing filter usually succeeds the channel, whereas in systems for the compensation of room acoustics it is placed in the signal path in front of the loudspeaker, which then acts as an acoustic source in the room. In both data transmission and room equalization, the channel impulse response is usually assumed to be known. In this paper we investigate both setups under the more realistic assumption of imperfect channel knowledge, and we show under which conditions the designs are equivalent. In particular, taking imperfect channel knowledge into account leads to robust designs that allow for more coarse, but faster channel estimation techniques.

1. INTRODUCTION

Approaches for listening room compensation (LRC) are based on a setup with an equalization filter in front of the loudspeaker [1]. The filter is designed with respect to one or more microphone positions in the room. In the present work, only a single microphone is considered. The room impulse response (RIR) is denoted by $c(n)$, and its z-transform is given by $C(z)$. In general, $C(z)$ is a mixed-phase system, having zeros inside and outside the unit circle. Therefore, only its minimum-phase component can be inverted by a standard causal IIR filter [2]. More recent proposals [3] stress the importance of equalizing the remaining allpass component, too. Alternative approaches are based on minimizing the mean squared error (MSE) between the output of a reference system with impulse response $\tilde{d}(n) = d(n - n_0)$ and the concatenation of the equalization filter, denoted as $h(n)$, and the RIR $c(n)$ [4, 5]. The parameter n_0 is an explicitly introduced system delay. The choice of the reference system is quite arbitrary, and in all known approaches for acoustical applications, a delayed discrete pulse or a bandpass filtered version of such a pulse is used as the desired target system.

Aiming at perfect equalization is quite intuitive and straightforward, but the concept can cause practical problems when the channel $C(z)$ has zeros close to, or even on the unit circle. For such channels, in data transmission, the method of channel shortening has been developed [6, 7]. It has originally been proposed to reduce the implementation cost of maximum likelihood detection via the Viterbi algorithm [6], and it is now also widely used in orthogonal frequency division multiplex (OFDM) and discrete multitone (DMT) systems to reduce the effective channel order to the length of the guard interval [7]. In this paper, we investigate the channel shortening concept for the use in listening room compensation. Thus, we look at the joint optimization of the FIR prefilter $h(n)$ and the FIR target system $d(n)$ with impulse response length L_d. The optimization of n_0 is a separate problem which is not considered here. In addition to the arguments used in data transmission, this approach is also motivated by the fact that a comparable relaxed requirement can be found in psychoacoustics: here one uses, for example, the D50-measure, which is defined as the ratio of the energy within 50 ms after the first peak of a RIR versus the complete impulse response's energy [8]. For this measure, which is related to the speech intelligibility within a room, the actual form of the impulse response $d(n)$ is not too relevant. Only the energy distribution is of interest. Thus, by choosing a target system with an optimized impulse response of 50 ms duration, we can directly maximize the D50-measure.

Known approaches for channel shortening assume perfect channel knowledge. However, in real-world applications where the channel has to be estimated first, this perfect knowledge is not always available. To address this problem, in our approach, we consider some measurement noise on the channel estimate. That is, the channel $c(n)$ is replaced by a model $\tilde{c}(n) = c(n) + p(n)$ where the sequence $p(n)$ is a random perturbation that is statistically independent of the input signal and other possible noise components. Figs. 1 and 2 show the two setups considered in this paper, including the random channel perturbation. In the configuration in Fig. 2 for data transmission, an additive channel noise $\eta(n)$ is present. The LRC system includes no additive noise, but

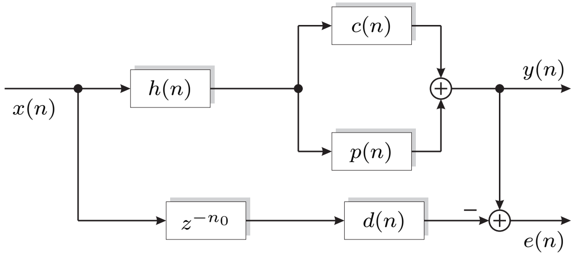

Fig. 1. Single-channel setup for room equalization. $p(n)$ is a random perturbation of the assumed channel impulse response $c(n)$.

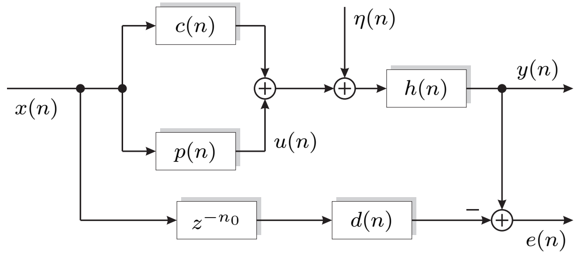

Fig. 2. Setup for memory truncation in data transmission.

the hypothetical noise amplification of the prefilter $h(n)$ is still of interest and can be considered in the design.

In the next section we will analyze the setups in Figs. 1 and 2 and show under which conditions the respective optimal solutions are equivalent. In Section 3, we then carry out the joint optimization of the prefilter and the target system while considering imperfect channel knowledge. Our impulse response shortening approach is based on the method by Kammeyer [9]. Simulation results are given in Section 4, and Section 5 concludes the paper.

Notation. Vectors and matrices are printed in boldface. The superscripts ^{T}, *, and ^{H} denote transposition, complex conjugation, and Hermitian transposition, respectively. $\Re\{\cdot\}$ returns the real part of a complex value, and δ_{ik} is the Kronecker symbol. The asterisk $*$ denotes convolution.

2. ANALOGIES OF EQUALIZER SETUPS IN DATA TRANSMISSION AND ACOUSTICAL SYSTEMS

We first consider the system in Fig. 1 and define the vectors

$$\begin{aligned}
\mathbf{x}_c &= \left[x(n), x(n-1), \ldots, x(n-L_c-L_h+2)\right]^T \\
\mathbf{x}_p &= \left[x(n), x(n-1), \ldots, x(n-L_p-L_h+2)\right]^T \\
\mathbf{x}_d &= \left[x(n-n_0), \ldots, x(n-n_0-L_d+1)\right]^T \\
\mathbf{h} &= \left[h(0), h(1), \ldots, h(L_h-1)\right]^T \\
\mathbf{c} &= \left[c(0), c(1), \ldots, c(L_c-1)\right]^T \\
\mathbf{p} &= \left[p(0), p(1), \ldots, p(L_p-1)\right]^T \\
\mathbf{d} &= \left[d(0), d(1), \ldots, d(L_d-1)\right]^T .
\end{aligned} \tag{1}$$

Note that for the signal vectors, the discrete time index n has been omitted. The terms L_h, L_c, L_p, and L_d denote the lengths of $h(n)$, $c(n)$, $p(n)$, and $d(n)$, respectively. We assume that $L_c \leq L_p$, which means that the random channel perturbation $p(n)$ can be longer than the assumed impulse response $c(n)$.

The error signal $e(n)$ can be described as

$$e(n) = \mathbf{x}_c^T \mathbf{C}\mathbf{h} + \mathbf{x}_p^T \mathbf{P}\mathbf{h} - \mathbf{x}_d^T \mathbf{d} \tag{2}$$

where $\mathbf{C}$ and $\mathbf{P}$ are convolution matrices of size $(L_c + L_h - 1) \times L_h$ and $(L_p + L_h - 1) \times L_h$, respectively.

In addition, we define the vector

$$\mathbf{v} = \left[v(n), v(n-1), \ldots, v(n-L_h+1)\right]^T \tag{3}$$

where $v(n)$ is a hypothetical noise process that would result in the filtered noise $\varepsilon(n) = \mathbf{v}^T\mathbf{h}$ when fed into the prefilter $h(n)$. The power of $\varepsilon(n)$ then gives us an indication of the noise amplification of the system $h(n)$. We assume that the three random processes $x(n)$, $p(n)$, and $v(n)$ are mutually uncorrelated and that at least $p(n)$ and $v(n)$ have zero mean and that $x(n)$ and $v(n)$ are wide-sense stationary.

An objective function is now defined as the weighted sum of the powers of the output error $e(n)$ and the hypothetical noise process $\varepsilon(n)$:

$$Q_1 = \mathrm{E}\left\{|e(n)|^2\right\} + \beta \mathrm{E}\left\{|\varepsilon(n)|^2\right\}, \quad \beta > 0. \tag{4}$$

We have

$$\begin{aligned}
Q_1 = {} & \mathbf{h}^H\mathbf{C}^H \mathrm{E}\left\{\mathbf{x}_c^*\mathbf{x}_c^T\right\}\mathbf{C}\mathbf{h} - 2\Re\{\mathbf{h}^H\mathbf{C}^H\mathrm{E}\left\{\mathbf{x}_c^*\mathbf{x}_d^T\right\}\mathbf{d}\} \\
& + \mathbf{h}^H\mathrm{E}\left\{\mathbf{P}^H\mathbf{x}_p^*\mathbf{x}_p^T\mathbf{P}\right\}\mathbf{h} + \mathbf{d}^H\mathrm{E}\left\{\mathbf{x}_d^*\mathbf{x}_d^T\right\}\mathbf{d} \\
& + \beta\mathbf{h}^H\mathrm{E}\left\{\mathbf{v}^*\mathbf{v}^T\right\}\mathbf{h} + 2\Re\{\mathbf{h}^H\mathbf{C}^H \underbrace{\mathrm{E}\left\{\mathbf{x}_c^*\mathbf{x}_p^T\mathbf{P}\right\}}_{=0}\mathbf{h}\} \\
& - 2\Re\{\mathbf{h}^H \underbrace{\mathrm{E}\left\{\mathbf{P}^H\mathbf{x}_p^*\mathbf{x}_d^T\right\}}_{=0}\mathbf{d}\}.
\end{aligned} \tag{5}$$

Next, we will investigate the setup in Fig. 2 with the filter $\mathbf{h}$ succeeding the channel. Here, the error signal results in

$$e(n) = \mathbf{x}_c^T\mathbf{C}\mathbf{h} + \mathbf{x}_p^T\mathbf{P}\mathbf{h} - \mathbf{x}_d^T\mathbf{d} + \boldsymbol{\eta}^T\mathbf{h} \tag{6}$$

with $\boldsymbol{\eta} = \left[\eta(n), \eta(n-1), \ldots, \eta(n-L_h+1)\right]^T$, where $\eta(n)$ is zero-mean channel noise that is uncorrelated to $x(n)$

and $p(n)$. An objective function is defined as

$$Q_2 = \mathrm{E}\{|e(n)|^2\} = \mathbf{h}^H\mathbf{C}^H\mathrm{E}\{\mathbf{x}_c^*\mathbf{x}_c^T\}\,\mathbf{C}\mathbf{h}$$
$$-2\Re\{\mathbf{h}^H\mathbf{C}^H\mathrm{E}\{\mathbf{x}_c^*\mathbf{x}_d^T\}\,\mathbf{d}\} + \mathbf{h}^H\mathrm{E}\{\mathbf{P}^H\mathbf{x}_p^*\mathbf{x}_p^T\mathbf{P}\}\,\mathbf{h}$$
$$+\mathbf{d}^H\mathrm{E}\{\mathbf{x}_d^*\mathbf{x}_d^T\}\,\mathbf{d} + \mathbf{h}^H\mathrm{E}\{\boldsymbol{\eta}^*\boldsymbol{\eta}^T\}\,\mathbf{h} \quad (7)$$
$$+2\Re\{\mathbf{h}^H\mathbf{C}^H\underbrace{\mathrm{E}\{\mathbf{x}_c^*\mathbf{x}_p^T\mathbf{P}\}}_{=0}\mathbf{h}\} - 2\Re\{\mathbf{h}^H\underbrace{\mathrm{E}\{\mathbf{P}^H\mathbf{x}_p^*\mathbf{x}_d^T\}}_{=0}\mathbf{d}\}.$$

By comparing (5) and (7) we see that for $\eta(n) = \sqrt{\beta}v(n)$ both objective functions are the same. Therefore, the solutions derived in the next section are valid for both setups.

3. IMPULSE RESPONSE SHORTENING WITH STOCHASTIC CHANNEL ESTIMATION ERROR

We follow the notation for Q_1. By setting the derivative of Q_1 with respect to $\mathbf{h}$ equal to zero and solving the resulting linear system, we obtain

$$\mathbf{h} = \big(\mathbf{C}^H\mathbf{R}_{\mathbf{x}_c\mathbf{x}_c}\mathbf{C} + \mathrm{E}\{\mathbf{P}^H\mathbf{x}_p^*\mathbf{x}_p^T\mathbf{P}\} + \beta\mathbf{R}_{vv}\big)^{-1}\mathbf{C}^H\mathbf{R}_{\mathbf{x}_c\mathbf{x}_d}\mathbf{d}. \quad (8)$$

The autocorrelation matrices $\mathbf{R}_{\mathbf{x}_c\mathbf{x}_c}$, $\mathbf{R}_{vv}$, and $\mathbf{R}_{\mathbf{x}_c\mathbf{x}_d}$ in (8) are defined according to their indices and the associated signals in the expectation operators in equation (5).

The expression $\mathrm{E}\{\mathbf{P}^H\mathbf{x}_p^*\mathbf{x}_p^T\mathbf{P}\}$ is the autocorrelation matrix of the process $u(n) = \sum_{i=0}^{L_p-1} p(i)x(n-i)$ that results from the convolution of the random input $x(n)$ with the random perturbation $p(n)$. For its autocorrelation sequence we obtain

$$\mathrm{E}\{u^*(n)u(n+\kappa)\} =$$
$$\mathrm{E}\Big\{\sum_{i=0}^{L_p-1} p^*(i)x^*(n-i)\sum_{j=0}^{L_p-1} p(j)x(n+\kappa-j)\Big\} \quad (9)$$
$$= \sum_{i=-(L_p-1)}^{L_p-1} r_{xx}(\kappa-i)\rho_{pp}(i)$$

with

$$r_{xx}(\kappa) = \mathrm{E}\{x^*(n)x(n+\kappa)\}$$
$$\rho_{pp}(i) = \sum_{n=0}^{L_p-1} r_{pp}(n,i)$$
$$r_{pp}(n,i) = \mathrm{E}\{p^*(n)p(n+i)\}\,.$$

The correlation matrix is given by

$$\mathbf{R}_{uu} = \mathrm{E}\{\mathbf{u}^*\mathbf{u}^T\} = \mathrm{E}\{\mathbf{P}^H\mathbf{x}_p^*\mathbf{x}_p^T\mathbf{P}\} \quad (10)$$

with $\mathbf{u} = \big[u(n), u(n-1), \ldots, u(n-L_h+1)\big]^T$.

Finally, the equalizer's coefficient vector becomes

$$\mathbf{h} = \underbrace{\big(\mathbf{C}^T\mathbf{R}_{\mathbf{x}_c\mathbf{x}_c}\mathbf{C} + \mathbf{R}_{uu} + \beta\mathbf{R}_{vv}\big)}_{=\mathbf{A}}{}^{-1}\mathbf{C}^T\mathbf{R}_{\mathbf{x}_c\mathbf{x}_d}\mathbf{d}. \quad (11)$$

From (11) we see the following. If the stochastic estimation error $p(n)$ is temporally not correlated, i.e. $r_{pp}(n,i) = \sigma_p^2\delta_{ni}$, and if $x(n)$ and $v(n)$ have the same statistical properties, then the perturbation $p(n)$ and the hypothetical noise $v(n)$ have the same quality apart from the scalar factor β.

Equation (11) gives us the optimal prefilter $h(n)$ for a given target system $d(n)$. Instead of choosing the target system in an ad hoc manner, we will now consider the choice of the optimal length-L_d target system $d(n)$ for a given channel $c(n)$. For this, we follow the method in [9], which, unlike the ones in [6, 7], avoids solving large eigenvalue problems and results in a linear system of equations.

We first formulate the homogeneous linear system:

$$\begin{bmatrix} \mathbf{R}_{\mathbf{x}_d\mathbf{x}_d} & -\mathbf{R}_{\mathbf{x}_c\mathbf{x}_d}^H\mathbf{C} \\ -\mathbf{C}^T\mathbf{R}_{\mathbf{x}_c\mathbf{x}_d} & \mathbf{A} \end{bmatrix}\begin{bmatrix} \mathbf{d} \\ \mathbf{h} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \quad (12)$$

The upper part expresses the fact that the first L_d coefficients of the impulse response $c(n) * h_{\mathrm{opt}}(n)$ should be equal to $d(n)$. The lower part equals (11). By setting $d(\ell) = 1$ in the vector $\mathbf{d}$ in (12) for some value of ℓ with $0 \leq \ell \leq L_d - 1$ and removing the ℓth row of the resulting linear system, we obtain an inhomogeneous system that can be easily solved for the remaining coefficients of $d(n)$ and the filter $h(n)$. Altogether, the method yields the optimal filters $\mathbf{d}_{\mathrm{opt}}$ and $\mathbf{h}_{\mathrm{opt}}$ under the condition that $d(\ell) = 1$.

4. SIMULATION RESULTS

A room impulse response (RIR) with a reverberation time of $\tau_{60} = 100\,\mathrm{ms}$, sampled at a frequency of 8 kHz, was generated with the well-known image method [10]. The length of the RIR and its perturbation was set to 800 taps, the equalizer's length accounted to 512 taps, and the target system consisted of 20 coefficients. The delay in front of the target system was set to $n_0 = 50$.

Fig. 3 depicts the original and the shortened impulse response. We see that the method is quite successful in reducing the effective impulse response length. Clearly, a better reduction of the tail of $c(n) * h_{\mathrm{opt}}(n)$ can be achieved with a longer prefilter, but even with prefilters of much shorter length, a significant reduction of the effective impulse response length can be achieved.

A measure of interest is the early-to-late ratio (ETLR)

$$\mathrm{ETLR} = \frac{\sum_{n=0}^{L_d+k_d-1}\mathrm{E}\{|g(n)|^2\}}{\sum_{n=L_d+k_d}^{L_p+L_h-1}\mathrm{E}\{|g(n)|^2\}}, \quad (13)$$

with $g(n)$ being the random overall response given by

$$g(n) = \sum_i h_{\mathrm{opt}}(i)c(n-i) + \sum_i h_{\mathrm{opt}}(i)p(n-i). \quad (14)$$

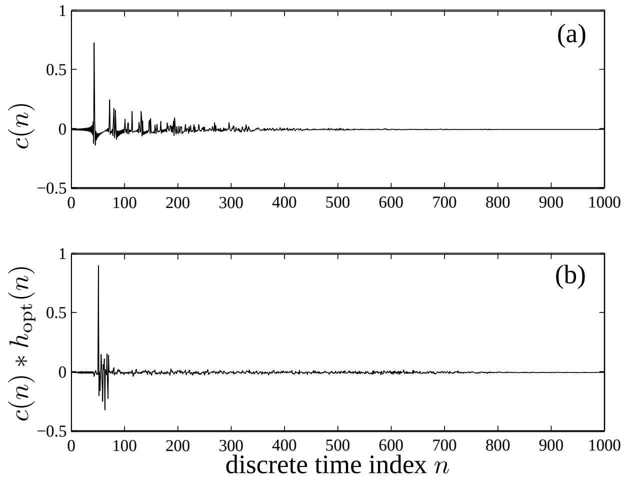

Fig. 3. Impulse responses. (a) Original. (b) Shortened.

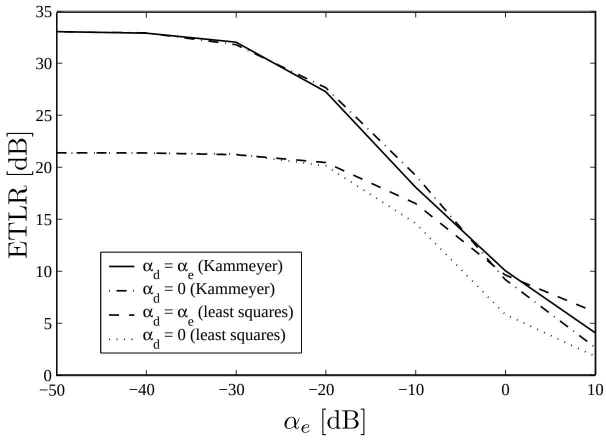

Fig. 4. Early-to-late-ratio ETLR as a function of α_e.

In our experiments, the perturbation process was defined as $p(n) = \alpha p_0(n)$, where $p_0(n)$ is a white random process with $\sum_{n=0}^{L_p-1} \mathrm{E}\{|p_0(n)|^2\} = \sum_{n=0}^{L_c-1} |c(n)|^2$, and α is used to adjust the average power of $p(n)$. We employed two different scaling factors, α_d and α_e, for the design and evaluation, respectively. Fig. 4 shows the ETLR measure for different choices of α_d and α_e. As expected, we see that it is optimal to have $\alpha_d = \alpha_e$, but even for $\alpha_d = 0$ and high perturbation, the results do not differ from the constrained design too heavily.

However, we can observe a distinct advantage of the chosen shortening approach by Kammeyer compared to a conventional least-squares equalizer

$$\mathbf{h} = \mathbf{A}^{-1}\mathbf{C}^T\mathbf{R}_{\mathbf{x}_c\mathbf{x}_d}\mathbf{d} \tag{15}$$

with a discrete pulse as a target system $\mathbf{d}$.

For very low perturbation, the ETLR measure saturates because of the finite prefilter length. Improvements are possible by increasing L_h.

5. CONCLUSIONS

In this paper, we have shown a method for the joint optimization of the prefilter and the target system for acoustic listening room compensation. To increase the robustness of the design, we introduced a possible perturbation of the previously measured room impulse response. It could be shown that assuming such a perturbation allows us to obtain better early-to-late ratios in scenarios where there is a mismatch between the measured and the true room impulse response.

6. REFERENCES

[1] J. N. Mourjopoulos, "Digital Equalization of Room Acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, pp. 884–900, Nov. 1994.

[2] S. T. Neely and J. B. Allen, "Invertibility of a Room Impulse Response," *Journal of the Acoustical Society of America (JASA)*, vol. 66, pp. 165–169, July 1979.

[3] B. D. Radlovic and R. A. Kennedy, "Nonminimum-Phase Equalization and its Subjective Importance in Room Acoustics," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 6, pp. 728–737, Nov. 2000.

[4] S. J. Elliott and P. A. Nelson, "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *Journal of the Audio Engineering Society*, vol. 37, no. 11, pp. 899–907, Nov. 1989.

[5] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast Deconvolution of Multichannel Systems using Regularization," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, März 1998.

[6] D. D. Falconer and F. R. Magee, "Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation," *The Bell System Technical Journal*, vol. 52, no. 9, pp. 1541–1562, Nov. 1973.

[7] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse Response Shortening for Discrete Multitone Transceivers," *IEEE Trans. on Computers*, vol. 44, no. 12, pp. 1662–1672, Dec. 1996.

[8] International Organization for Standardization (ISO), "ISO Norm 3382: Acoustics – Measurement of the Reverberation Time of Rooms with Reference to other Acoustical Parameters," .

[9] K. D. Kammeyer, "Time Truncation of Channel Impulse Responses by Linear Filtering: A Method to Reduce the Complexity of Viterbi Equalization," *Archiv für Elektronik und Übertragungstechnik (AEÜ) – Int. Journal of Electronics and Communications*, vol. 48, no. 5, pp. 237–243, May 1994.

[10] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.