# Estimation of the Optimum Delay for Speech Dereverberation by Inverse Filtering

Stefan Goetze[1], Markus Kallinger[2], Alfred Mertins[3], and Karl-Dirk Kammeyer[4]

## Introduction

Equalization of room impulse responses is an attractive approach for dereverberation of speech signals in a hands-free scenario. In this contribution we address the choice of the delay which has to be introduced in least-squares equalization approaches for a maximum amount of dereverberation. Since designing one equalizer (EQ) for each possible delay and choosing the best one is computationally inefficient we evaluate the dependence of the optimum equalizer delay of various measures characterizing room impulse responses (RIRs). A high correlation was found between the so-called central time of the room impulse response and the optimum equalizer delay. Since the central time can be determined based on estimates of the initial peak of the RIR and the room reverberation time, we propose to use a very short filter for system identification and an estimate of the room reverberation time to identify the optimum equalizer delay. The proposed approach prevents a low performance of the equalizer which may occur for an improperly chosen delay by automatically estimating the optimum delay.

## Listening-Room Compensation

The equalization of (acoustic) channels has been research topic for several years now [1, 2]. However, due to the nature of usual room impulse responses which are mixed-phase systems having a length of several thousand taps it still remains a challenging problem [3, 4]. The choice of an appropriate system delay for the equalization filter is unaddressed in most of the contributions and will be analyzed in this paper since it has a strong influence on the performance of the equalizer.

Fig. 1 shows the common setup for listening-room compensation with the equalization filter $\mathbf{c}_{\mathrm{EQ}}$ preceding the room impulse response (RIR) $\mathbf{h}$. To remove reverberation caused by the convolution with the RIR the equalizer $\mathbf{c}_{\mathrm{EQ}}$ tries to minimize the system distance between the concatenated system of $\mathbf{c}_{\mathrm{EQ}}$ convolved with $\mathbf{h}$ and a desired target system $\mathbf{d}$ [2].

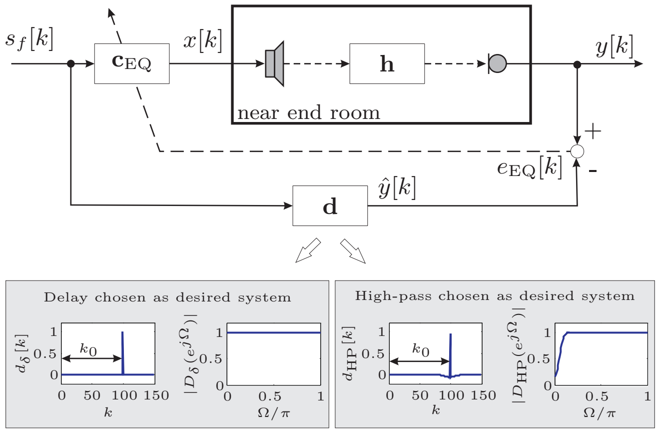


**Figure 1:** Least-squares equalizer $\mathbf{c}_{\mathrm{EQ}}$ for listening-room compensation and two possible desired systems $d_\delta[k]$ (delay) and $d_{\mathrm{HP}}[k]$ (delayed high-pass) in time and frequency domain.

The minimization of the error signal $\mathrm{E}\left\{|e_{\mathrm{EQ}}[k]|^2\right\} = \mathrm{E}\left\{|\mathbf{s}_f^T[k]\mathbf{H}\mathbf{c}_{\mathrm{EQ}} - \mathbf{s}_f^T[k]\mathbf{d}|^2\right\}$ leads to the well known least squares equalizer [3, 4]

$$\mathbf{c}_{\mathrm{EQ}} = \mathbf{H}^{+}\mathbf{d} \tag{1}$$

under the assumption of a white input signal. Here, $\mathbf{H}^{+}$ is the Moore-Penrose pseudoinverse of the channel convolution matrix $\mathbf{H}$ and $\mathbf{c}_{\mathrm{EQ}}$ is a vector containing the filter coefficients of the equalizer. The vector $\mathbf{d}$ contains the delayed desired system and can be chosen as a delayed unit impulse ($\mathbf{d}_\delta$) or a delayed high pass ($\mathbf{d}_{\mathrm{HP}}$) to account for the frequency responses of imperfect transfer characteristics of loudspeakers and microphones, e.g.

$$\mathbf{d}_\delta = [\,\mathbf{0}_{1\times(k_0-1)},\, 1,\, \mathbf{0}_{1\times(L_h+L_{c,\mathrm{EQ}}-k_0-3)}]^T \tag{2}$$

$$\mathbf{d}_{\mathrm{HP}} = [\,\underbrace{0,\,...,\,0}_{\tilde{k}_0},\, d_0,\,...,\, d_{\lfloor L_d/2\rfloor},\,...,\, d_{L_d-1},\, \underbrace{0,\,...,\,0}_{L_h+L_{c,\mathrm{EQ}}-1-L_d-\tilde{k}_0}\,]^T \tag{3}$$

$L_h$, $L_{c,\mathrm{EQ}}$ and $L_d$ are the lengths of the RIR, of the equalizer filter and the desired system, respectively. The delay introduced by the equalizer is denoted as $k_0$ as depicted in Figure 1. It corresponds directly to the position of the one for the delayed impulse in (2) and for desired systems of length $L_d > 1$ as in (3) the delay $k_0$ corresponds to the middle position of the desired system $k_0 = \tilde{k}_0 + \lfloor L_d/2 \rfloor$. Many contributions in the literature suggest to use a *good guess* for the parameter $k_0$. In this paper we will try to find a better way to determine an optimum $k_0$.

Since the equalizer performance depends on the specific RIR that has to be equalized different measures characterizing a RIR are briefly introduced in the following to determine if one of these measures can be used to estimate an optimum $k_0$ properly.

[1]Stefan Goetze is with Fraunhofer Institute for Digital Media Technology (IDMT), Project group Hearing Speech and Audio Technology (HSA), 26129 Oldenburg, Germany, [2]Markus Kallinger is with Fraunhofer Institute for Integrated Circuits (IIS), 91058 Erlangen, Germany, he contributed to this work while he was with University of Oldenburg, Signal Processing Group, 26111 Oldenburg, Germany, [3]Alfred Mertins is with University of Lübeck, Institute for Signal Processing, 23538 Lübeck, Germany, [4]Karl-Dirk Kammeyer is with University of Bremen, Dept. of Communications Engineering, 28334 Bremen, Germany. Work was supported in parts by the German research foundation DFG under grant Ka841-17.

# Objective Measures Characterizing RIRs

Room impulse responses can be characterized by several objective measures (see e.g. [5]). The following list contains some of them:

**Room reverberation time $\tau_{60}$ [5]:** An important measure characterizing an RIR is the so-called room reverberation time $\tau_{60}$. It is defined as the time after which the energy of the RIR is decayed by 60dB.

**Delay of direct path of the RIR:** The delay of the direct path of the RIR $k_{h_{max}}$ directly corresponds to the distance between source and microphone. Here, $k_{h_{max}}$ is defined as the discrete-time index of the maximum of $|h[k]|$.

**Direct-to-Reverberation-Ratio (DRR) [6]:** $\mathrm{DRR} = 10\log_{10}\frac{h^2[k_{h_{max}}]}{\sum_{k\neq k_{h_{max}}} h^2[k]}$. The DRR is the ratio between direct path to reverberation (all other pathes) in dB.

**Definition [5]:** $D_{50} = \frac{\sum_{k=0}^{k_{50}-1} h^2[k]}{\sum_{k=0}^{L_h} h^2[k]}$. Here, $k_{50} = 50\mathrm{ms}\cdot f_s$ is the discrete-time index corresponding to a time of 50ms. Thus, the definition measure is defined as the ratio between the energy of the first 50ms to the overall energy of the RIR.

**Clarity Index (CI) [5]:** $\mathrm{CI} = 10\log_{10}\frac{\sum_{k=0}^{k_{80}} h^2[k]}{\sum_{k=k_{80}}^{L_h} h^2[k]}$. Here, $k_{80} = 80\mathrm{ms}\cdot f_s$ is the discrete time index corresponding to a time of 80ms.

**Central Time (CT) [5]:** $\mathrm{CT} = \frac{\sum_{k=0}^{L_h} k\cdot h^2[k]}{\sum_{k=0}^{L_h} h^2[k]}$. The central time of an RIR can be interpreted as the center of gravity in terms of the energy of the RIR.

The previously described measures were calculated for various RIRs that were (i) generated artificially by the so-called image method [7], (ii) measured [8], (iii) taken from the MARDY database [9], and (iv) modeled by an exponentially damped Gaussian noise (compare equation (6)). A total number of 270 RIRs was used with room reverberation times ranging from $\tau_{60} = 50\mathrm{ms}$ to $\tau_{60} = 1200\mathrm{ms}$.

# Estimation of Optimum Equalizer Delay

We now examine the correlations of the optimum system delay $k_{0,\mathrm{opt}}$ with the previously described measures characterizing the impulse responses to see if one of these measures can be used for an estimation of $k_{0,\mathrm{opt}}$.

For that purpose all equalizers are evaluated by means of the Bark spectral distortion (BSD) measure [10] that was developed for evaluation of speech quality and is widely used for evaluation of dereverberation algorithms and the signal-to-reverberation-ratio-enhancement (SRRE) [4, 11] which is the enhancement of the signal-to-reverberation-ratio achieved by the dereverberation algorithm. The equalizer performance and, thus, both measures depend on the chosen equalizer delay and were calculated for varying $k_0$.

Thus, $k_{0,\mathrm{opt,BSD}} = \mathrm{argmin}_{k_0}\{\mathrm{BSD}\}$ is the equalizer delay for the minimum achievable BSD for a given RIR if the parameter $k_0$ in (3) is varied and $k_{0,\mathrm{opt,SRRE}} = \mathrm{argmax}_{k_0}\{\mathrm{SRRE}\}$ is the corresponding equalizer delay at the maximum SRRE if the parameter $k_0$ in (3) is varied. Please note, that a small BSD indicates a good performance while for the SRRE a high value indicates good performance. Both measures (BSD and SRRE) lead to similar optimum delays for all RIRs tested ($k_{0,\mathrm{opt,BSD}} \approx k_{0,\mathrm{opt,SRRE}}\ \forall\,\mathbf{h}$).

The optimum delays defined by the maximum SRRE and the minimum BSD were calculated for each RIR and for different equalizer orders $L_{c,\mathrm{EQ}} = \{256, 512, 1024\}$ as illustrated in Figure 2 exemplarily for the SRRE and for two different RIRs $h_1[k]$ and $h_2[k]$ that are depicted in sub-figures (a) and (b) having room reverberation times of $\tau_{60} = 500\mathrm{ms}$ and $\tau_{60} = 1\mathrm{s}$, respectively. The corresponding equalizer performances in terms of SRRE in dependence of the system delay $k_0$ is shown in sub-figure (c) for the different equalizer lengths $L_{c,\mathrm{EQ}} = \{256, 512, 1024\}$ with solid lines, dashed lines and dash-dotted lines, respectively. Thick blue lines are used for the curves showing the equalizer performance if the RIR $h_1[k]$ is equalized and thin red lines are used for equalization of $h_2[k]$.

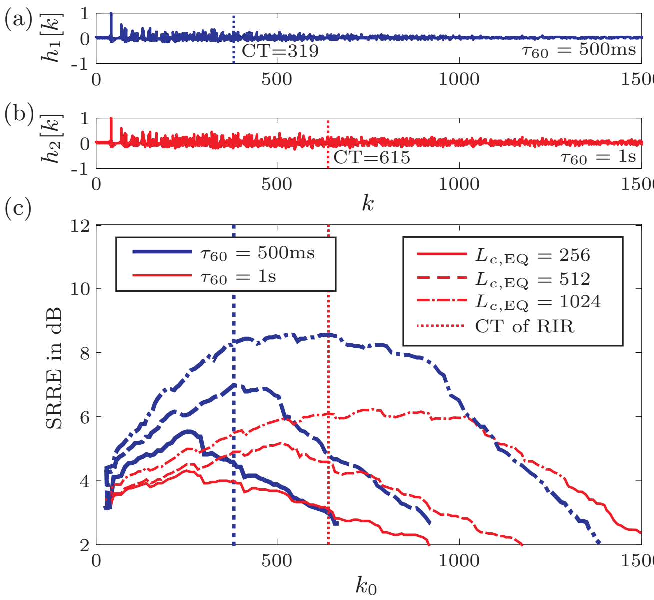


**Figure 2:** (a) RIR with reverberation times of $\tau_{60} = 500\mathrm{ms}$ and its CT in samples. (b) RIR with $\tau_{60} = 1\mathrm{s}$ and its CT. (c) equalizer performance in dependence of delay $k_0$ of the desired system for different equalizer filter lengths $L_{c,\mathrm{EQ}}$ and RIRs (a) (thicker blue lines) and (b) (thinner red lines).

It can be clearly seen from Figure 2 (c) that the equalizer performance depends on the equalizer delay $k_0$. Thus, we calculate the correlation between the measures describing a specific RIR and the optimum $k_{0,\mathrm{opt}}$ in the following.

Table 1 shows the correlations

$$r = \frac{\sum_i (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_i (A_i - \bar{A})^2 \sum_i (B_i - \bar{B})^2}} \tag{4}$$

between the different measures characterizing the RIRs and $k_{0,\text{opt,SRRE}}$ and Table 2 shows the correlations between the different measures characterizing the RIRs $k_{0,\text{opt,BSD}}$. In (4) $A_i$ and $B_i$ denote the specific calculated values of $k_{0,\text{opt}}$ and CT, respectively, and $\bar{A}$ and $\bar{B}$ their mean values.

| | $k_0$ for SRRE correlated with | | | | | |
|---|---|---|---|---|---|---|
| $L_{c,\text{EQ}}$ | $\tau_{60}$ | $k_{h_{max}}$ | DRR | $D_{50}$ | CI | CT |
| 256 | 0.28 | **0.89** | 0.86 | 0.47 | 0.49 | 0.80 |
| 512 | 0.39 | 0.78 | **0.85** | 0.30 | 0.57 | **0.85** |
| 1024 | 0.36 | 0.74 | 0.83 | 0.27 | 0.50 | **0.84** |

**Table 1:** Correlation coefficients between optimum equalizer delay according to SRRE and RIR properties for varying equalizer length.

| | $k_0$ for BSD correlated with | | | | | |
|---|---|---|---|---|---|---|
| $L_{c,\text{EQ}}$ | $\tau_{60}$ | $k_{h_{max}}$ | DRR | $D_{50}$ | CI | CT |
| 256 | 0.37 | **0.84** | 0.84 | 0.37 | 0.56 | 0.82 |
| 512 | 0.39 | 0.70 | 0.74 | 0.23 | 0.58 | **0.75** |
| 1024 | 0.53 | 0.66 | 0.80 | 0.11 | 0.63 | **0.89** |

**Table 2:** Correlation coefficients between optimum equalizer delay according to BSD and RIR properties for varying equalizer length.

Figure 3 exemplarily shows the CT for all 270 RIRs over the optimum equalizer delays $k_{0,\text{opt,BSD}}$ (left) and $k_{0,\text{opt,SRRE}}$ (right).

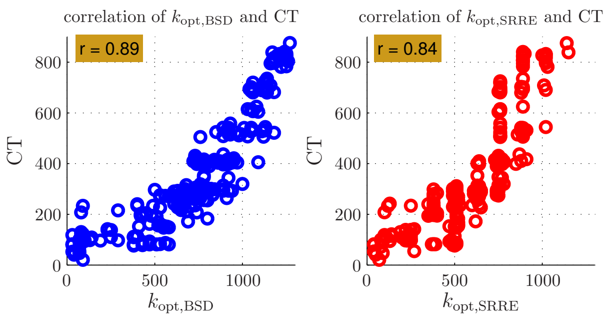


**Figure 3:** Correlations between central time (CT) and optimum equalizer delay given by the minimum of the BSD (left) and maximum of the SRRE (right) for an equalizer length of $L_{c,\text{EQ}} = 1024$.

The highest correlations are indicated by bold letters in Tables 1 and 2 and it can be seen that the central time (CT) seems to be a good indicator for the optimum equalizer delay $k_{0,\text{opt}}$ for both, BSD and SRRE. The somewhat lower correlation for short equalizer lengths in Tables 1 and 2 can be explained by taking a closer look at Figure 3. If the CT is greater than the equalizer length the equalizer may not be capable to introduce the desired delay. Hence, we propose to chose the equalizer delay as follows:

$$\hat{k}_{0,\text{opt}} = \min\{\text{CT}, L_{c,\text{EQ}}\} \tag{5}$$

Using the criterion (5) to determine the optimum equalizer delay achieves 94.4% of the performance in our tests for all 270 RIRs compared to the case that the optimum delay is known a priori. (90.5% is achieved if the CT is used as a direct criterion for determining $k_{0,\text{opt}}$).

# Estimation of the Central Time

In practical systems the delay $k_0$ has to be chosen without a priori information about the RIR which shall be equalized. Thus, we propose to estimate the central time of the RIR by applying a very short acoustic echo canceller (AEC) [4] to identify the initial delay and the first few samples of the RIR and an estimator of the room reverberation time $\tau_{60}$. Different methods exist for the estimation of the reverberation time [12, 13, 14, 15] directly from the reverberant signal or by modeling the RIR as an exponentially damped Gaussian process

$$h_M[k] = b[k] \exp\left(-\frac{(k - k_{\text{init}})}{\beta}\right) u[k - k_{\text{init}}] \tag{6}$$

with $k_{\text{init}}$ being the initial delay of the room impulse response model, $b[k]$ a white Gaussian random process, $u[k - k_{\text{init}}]$ the time-shifted Heaviside step function, $f_s$ the sampling frequency and

$$\beta = \frac{2\tau_{60} f_s}{\ln(10^{-6})} \tag{7}$$

a damping constant that depends on the room reverberation time $\tau_{60}$ as depicted in Figure 4.

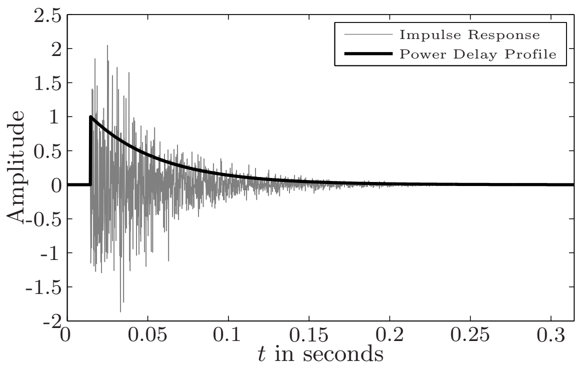


**Figure 4:** RIR and power delay profile (PDP).

Please note that an estimate of the room reverberation time has to be done only once for a specific room, since it does not vary too much for different spatial positions. The length of the AEC can be restricted to a few taps since only the position of the initial RIR coefficients is needed to fit the power delay profile by a least-squares approach [13]. Thus, the AEC will converge extremely fast and has a very low computational complexity.

To avoid inversion of the RIR matrix $\mathbf{H}$ in (1) which has a size of $L_h + L_{c,\text{EQ}} - 1 \times L_{c,\text{EQ}}$ and to allow for tracking of RIR changes we use gradient algorithms working in the block frequency domain as described in [16, 17] for the equalizer as well as for the acoustic echo canceler which identifies the room impulse response. The AEC length was chosen to $L_{c,\text{AEC}} = 256$ at a sampling frequency of $f_s = 8000\text{Hz}$ to identify the initial part of the RIR and to estimate $\beta$ and $k_{\text{init}}$ according to (6) and (7) by least-squares fitting. Afterwards, the central time was calculated from (6) and used as the delay $k_0$ in (3).

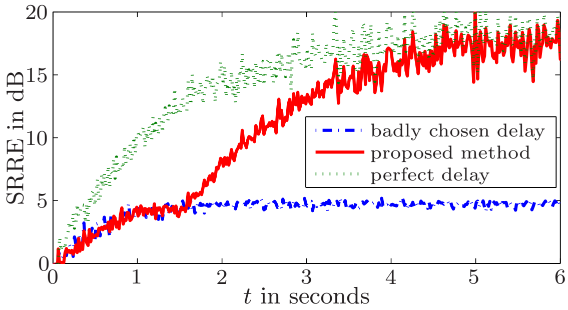


**Figure 5:** Performance comparison of equalizers using different delays $k_0$ in terms of SRRE.

Figure 5 shows the convergence of the equalizer of length $L_{c,\mathrm{EQ}} = 1024$ updated by the so-called decoupled filtered-X least mean square (dFxLMS) algorithm described in [17] for the case of perfect knowledge of the best possible delay $k_0$ (upper curve) and if a *bad guess* was made for the delay (lower curve). The solid curve in the middle shows the convergence behavior if the equalizer delay is switched at about 1.5 seconds from the *bad guess* to the proposed estimate according to (5).

It can be seen that the equalizer delay can be switched without performance loss and that the proposed equalizer reaches nearly the same performance as if the equalizer delay $k_0$ would be known a priori.

## Conclusions

This contribution analyzes the influence of the delay that has to be introduced by an equalizer for speech dereverberation by inverse filtering. A high correlation was found between the central time of a room impulse response that has to be equalized and an optimum equalizer delay w.r.t. maximum dereverberation performance. An estimator for the central time by identification of the first samples of the RIR by an adaptive filter was proposed which allows for an identification of the room reverberation time and, by this, of the central time of the RIR. It was shown that the proposed scheme is capable to enhance the equalizer performance in the case that an improper delay was chosen.

## References

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

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

[3] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse Filtering for Speech Dereverberation Less Sensitive to Noise and Room Transfer Function Fluctuations," *EURASIP J. on Advances in Signal Processing*, vol. Volume 2007, Article ID 34013, 2007,

[4] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "System Identification for Multi-Channel Listening-Room Compensation using an Acoustic Echo Canceller," in *Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, Trento, Italy, pp. 224-227, May 2008.

[5] H. Kuttruff, *Room Acoustics*, Spoon Press, London, 4. edition, 2000.

[6] M. Triki and D.T.M. Slock, "Iterated Delay and Predict Equalization for Blind Speech Dereverberation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, Sept. 2006.

[7] 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.

[8] Massarani P. Müller, S., "Transfer-Function Measurement with Sweeps," *J. of the Audio Engineering Society*, vol. 49, no. 6, pp. 443-471, June 2001.

[9] J.Y.C. Wen, N.D. Gaubitch, E.A.P. Habets, T. Myatt, and P.A. Naylor, "Evaluation of Speech Dereverberation Algorithms using the MARDY Database," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, Sept. 2006.

[10] S. Wang, A. Sekey, and A. Gersho, "An Objective Measure for Predicting Subjective Quality of Speech Coders," *IEEE J. Selected Areas of Communications*, vol. 10, no. 5, pp. 819-829, June 1992.

[11] P.A. Naylor and N.D. Gaubitch, "Speech Dereverberation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sept. 2005.

[12] R. Ratnam, D.L. Jones, B.C. Wheeler, J.W.D. O'Brien, C.R. Lansing, and A.S. Feng, "Blind Estimation of Reverberation Time," *J. of the Acoustical Society of America (JASA)*, vol. 114, no. 5, pp. 2877-2892, 2003.

[13] J. Schroeder, T. Rohdenburg, V. Hohmann, and S.D. Ewert, "Classification of Reverberant Acoustic Situations," in *Int. Conf. on Acoustics (NAG/DAGA'09)*, Rotterdam, The Netherlands, March 2009.

[14] H. Löllmann and P. Vary, "Estimation of the Reverberation Time in Noisy Environments," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattla, USA, Sept. 2008.

[15] T.J. Cox, F. Li, and P. Dalington, "Extracting Room Reverberation Time from Speech Using Artificial Neural Networks," *J. of the Acoustical Society of America (JASA)*, vol. 94, no. 4, pp. 219-230, 2001.

[16] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "A Decoupled Filtered-X LMS Algorithm for Listening-Room Compensation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, Sept. 2008.

[17] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Multi-Channel Listening-Room Compensation unsing a Decoupled Filtered-X LMS Algorithm," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, Oct. 2008.