Non-intrusive Regularization for Least-Squares Multichannel Equalization for Speech Dereverberation

Ina Kodrasi [1]   Stefan Goetze [2]   Simon Doclo [1,2]

[1]Signal Processing Group
University of Oldenburg
Oldenburg, Germany

[2]Project Group Hearing, Speech and Audio Technology
Fraunhofer IDMT
Oldenburg, Germany

15.11.2012

Overview

- **Speech Dereverberation**
- **Acoustic Multichannel Equalization**
  - Regularized partial multichannel equalization based on MINT (P-MINT)
- **Regularization Parameter for Regularized Least-Squares Techniques**
  - Intrusive selection
  - Non-intrusive selection
- **Experimental Results and Summary**

Overview

- **Speech Dereverberation**
- **Acoustic Multichannel Equalization**
  - Regularized partial multichannel equalization based on MINT (P-MINT)
- **Regularization Parameter for Regularized Least-Squares Techniques**
  - Intrusive selection
  - Non-intrusive selection
- **Experimental Results and Summary**

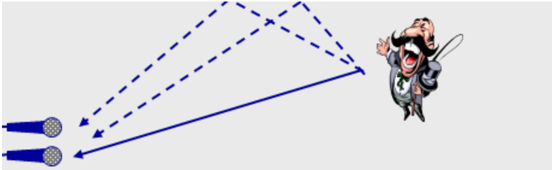
Overview

- **Speech Dereverberation**
- **Acoustic Multichannel Equalization**
  - Regularized partial multichannel equalization based on MINT (P-MINT)
- **Regularization Parameter for Regularized Least-Squares Techniques**
  - Intrusive selection
  - Non-intrusive selection
- **Experimental Results and Summary**

Overview

- **Speech Dereverberation**
- **Acoustic Multichannel Equalization**
  - Regularized partial multichannel equalization based on MINT (P-MINT)
- **Regularization Parameter for Regularized Least-Squares Techniques**
  - Intrusive selection
  - Non-intrusive selection
- **Experimental Results and Summary**

Speech Reverberation and Dereverberation



Reverberation degrades: speech quality, speech intelligibility, performance in ASR systems

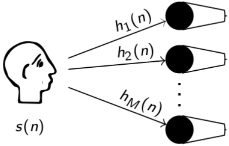
Applications: voice-controlled systems, hearing aids, teleconferencing applications

Techniques: LP processing [Gaubitch, Delcroix], spectral enhancement [Habets, Gannot, Jeub], **acoustic multichannel equalization** [Miyoshi, Naylor, Mertins, Kodrasi]

Acoustic Multichannel Equalization for Speech Dereverberation

Blindly identify RIRs and **design reshaping filters that compensate for their effect**

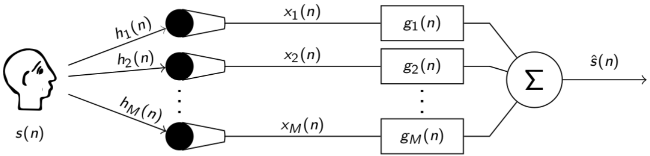
- Single-source multi-microphone system (M microphones)
- Estimate RIRs $\widehat{\mathbf{h}}_m = \left[\hat{h}_m(0) \ \ldots \ \hat{h}_m(L_h - 1)\right]^T$



Acoustic Multichannel Equalization for Speech Dereverberation

Blindly identify RIRs and **design reshaping filters that compensate for their effect**

- Single-source multi-microphone system (M microphones)
- Estimate RIRs $\widehat{\mathbf{h}}_m = \left[\hat{h}_m(0) \; \ldots \; \hat{h}_m(L_h - 1)\right]^T$
- Design reshaping filters $g_m = [g_m(0) \; \ldots \; g_m(L_g - 1)]^T$ based on **different optimality criteria**

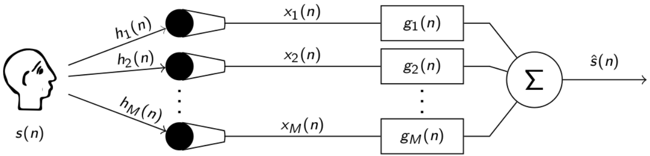


Acoustic Multichannel Equalization for Speech Dereverberation

Blindly identify RIRs and **design reshaping filters that compensate for their effect**

$$\hat{s}(n) = \sum_{m=1}^{M} x_m(n) * g_m(n) = s(n) * \underbrace{\sum_{m=1}^{M} h_m(n) * g_m(n)}_{\text{equalized impulse response} \quad c(n)}$$

$$\boxed{\hat{s}(n) = s(n) * c(n)}$$



Acoustic Multichannel Equalization for Speech Dereverberation

Blindly identify RIRs and **design reshaping filters that compensate for their effect**

$$\hat{s}(n) = \sum_{m=1}^{M} x_m(n) * g_m(n) = s(n) * \underbrace{\sum_{m=1}^{M} h_m(n) * g_m(n)}_{\text{equalized impulse response } c(n)}$$

$$\boxed{\hat{s}(n) = s(n) * c(n)}$$

Optimize:

- ideally, the true EIR $c(n) = \sum_{m=1}^{M} h_m(n) * g_m(n)$
- in practice, the estimated EIR $\hat{c}(n) = \sum_{m=1}^{M} \hat{h}_m(n) * g_m(n)$

Problem Formulation

Using

- the estimated convolution matrices $\hat{\mathbf{H}}_m$

$$\widehat{\mathbf{H}}_m = \begin{bmatrix} \hat{h}_m(0) & 0 & \dots & 0 \\ \hat{h}_m(1) & \hat{h}_m(0) & \ddots & \vdots \\ \vdots & \hat{h}_m(1) & \ddots & 0 \\ \hat{h}_m(L_h-1) & \vdots & \ddots & \hat{h}_m(0) \\ 0 & \hat{h}_m(L_h-1) & \ddots & \hat{h}_m(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \hat{h}_m(L_h-1) \end{bmatrix}_{(L_h+L_g-1)\times L_g}$$

Problem Formulation

Using

- the estimated convolution matrices $\hat{\mathbf{H}}_m$
- the estimated multichannel convolution matrix

$$\widehat{\mathbf{H}} = [\widehat{\mathbf{H}}_1 \ \widehat{\mathbf{H}}_2 \ \ldots \ \widehat{\mathbf{H}}_M]_{(L_h+L_g-1)\times ML_g}$$

- the multichannel reshaping filter vector

$$\mathbf{g} = \left[\mathbf{g}_1^T \ \mathbf{g}_2^T \ \ldots \ \mathbf{g}_M^T\right]^T_{ML_g\times 1}$$

optimize the estimated EIR $\widehat{\mathbf{c}} = \hat{\mathbf{H}}\mathbf{g}$

Problem Formulation

Using

- the estimated convolution matrices $\hat{\mathbf{H}}_m$
- the estimated multichannel convolution matrix

$$\widehat{\mathbf{H}} = [\widehat{\mathbf{H}}_1 \; \widehat{\mathbf{H}}_2 \; \ldots \; \widehat{\mathbf{H}}_M]_{(L_h+L_g-1)\times ML_g}$$

- the multichannel reshaping filter vector

$$\mathbf{g} = \left[\mathbf{g}_1^T \; \mathbf{g}_2^T \; \ldots \; \mathbf{g}_M^T\right]^T_{ML_g\times 1}$$

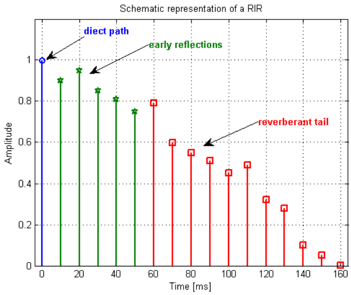
optimize the estimated EIR $\widehat{\mathbf{c}} = \hat{\mathbf{H}}\mathbf{g}$

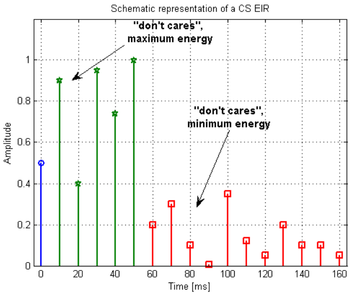
Typical assumptions

- The RIRs do not share any common zeros
- The reshaping filter length is chosen as $L_g \geq \lceil \frac{L_h-1}{M-1} \rceil$
- The estimated convolution matrix $\widehat{\mathbf{H}}$ is full row-rank

Overview of State-of-the-art Techniques

Channel Shortening (CS) [Kallinger 2006] maximizes the energy of the first part of the EIR while minimizing the reverberant energy





Overview of State-of-the-art Techniques

Channel Shortening (CS) [Kallinger 2006] maximizes the energy of the first part of the EIR while minimizing the reverberant energy

Cost function

$$\max_{\mathbf{g}} J_{\mathrm{CS}} = \frac{\|\mathbf{W}_d \hat{\mathbf{H}} \mathbf{g}\|_2^2}{\|\mathbf{W}_u \hat{\mathbf{H}} \mathbf{g}\|_2^2} = \frac{\mathbf{g}^T \hat{\mathbf{B}} \mathbf{g}}{\mathbf{g}^T \hat{\mathbf{A}} \mathbf{g}}$$

$$\hat{\mathbf{B}} = \hat{\mathbf{H}}^T \mathbf{W}_d^T \mathbf{W}_d \hat{\mathbf{H}} \qquad \hat{\mathbf{A}} = \hat{\mathbf{H}}^T \mathbf{W}_u^T \mathbf{W}_u \hat{\mathbf{H}},$$

$$\mathbf{W}_d = \mathrm{diag}\{[\underbrace{1 \ \ldots \ 1}_{L_d} \ 0 \ \ldots \ 0]\} \quad \mathbf{W}_u = \mathrm{diag}\{[\underbrace{0 \ \ldots \ 0}_{L_d} \ 1 \ \ldots \ 1]\}$$

Overview of State-of-the-art Techniques

Channel Shortening (CS) [Kallinger 2006] maximizes the energy of the first part of the EIR while minimizing the reverberant energy

Cost function

$$\max_{\mathbf{g}} J_{\mathrm{CS}} = \frac{\|\mathbf{W}_d \hat{\mathbf{H}} \mathbf{g}\|_2^2}{\|\mathbf{W}_u \hat{\mathbf{H}} \mathbf{g}\|_2^2} = \frac{\mathbf{g}^T \hat{\mathbf{B}} \mathbf{g}}{\mathbf{g}^T \hat{\mathbf{A}} \mathbf{g}}$$

$$\hat{\mathbf{B}} = \hat{\mathbf{H}}^T \mathbf{W}_d^T \mathbf{W}_d \hat{\mathbf{H}} \qquad \hat{\mathbf{A}} = \hat{\mathbf{H}}^T \mathbf{W}_u^T \mathbf{W}_u \hat{\mathbf{H}},$$

$$\mathbf{W}_d = \mathrm{diag}\{[\underbrace{1\ \ldots\ 1}_{L_d}\ 0\ \ldots\ 0]\} \quad \mathbf{W}_u = \mathrm{diag}\{[\underbrace{0\ \ldots\ 0}_{L_d}\ 1\ \ldots\ 1]\}$$

Solution $\hat{\mathbf{B}} \mathbf{g}_{\mathrm{CS}} = \lambda_{\max} \hat{\mathbf{A}} \mathbf{g}_{\mathrm{CS}}$

Overview of State-of-the-art Techniques

Channel Shortening (CS) [Kallinger 2006] maximizes the energy of the first part of the EIR while minimizing the reverberant energy

Cost function

$$\max_{\mathbf{g}} J_{\mathrm{CS}} = \frac{\|\mathbf{W}_d \hat{\mathbf{H}} \mathbf{g}\|_2^2}{\|\mathbf{W}_u \hat{\mathbf{H}} \mathbf{g}\|_2^2} = \frac{\mathbf{g}^T \hat{\mathbf{B}} \mathbf{g}}{\mathbf{g}^T \hat{\mathbf{A}} \mathbf{g}}$$

$$\hat{\mathbf{B}} = \hat{\mathbf{H}}^T \mathbf{W}_d^T \mathbf{W}_d \hat{\mathbf{H}} \qquad \hat{\mathbf{A}} = \hat{\mathbf{H}}^T \mathbf{W}_u^T \mathbf{W}_u \hat{\mathbf{H}},$$

$$\mathbf{W}_d = \mathrm{diag}\{[\underbrace{1 \ \ldots \ 1}_{L_d} \ 0 \ \ldots \ 0]\} \quad \mathbf{W}_u = \mathrm{diag}\{[\underbrace{0 \ \ldots \ 0}_{L_d} \ 1 \ \ldots \ 1]\}$$

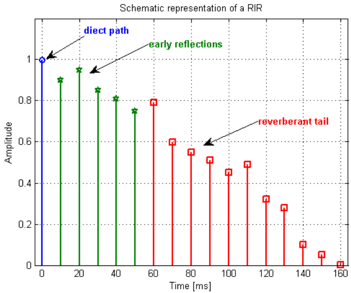
Solution $\hat{\mathbf{B}} \mathbf{g}_{\mathrm{CS}} = \lambda_{\max} \hat{\mathbf{A}} \mathbf{g}_{\mathrm{CS}}$

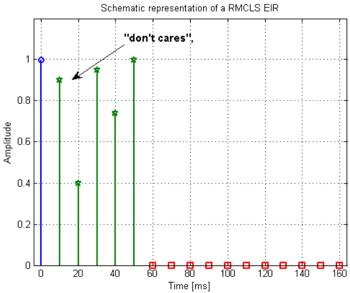
Drawback

No perceptual speech quality preservation

Overview of State-of-the-art Techniques

Relaxed multichannel least-squares (RMCLS) [Zhang 2010] aims to set the reverberant tail of the EIR to **0**





Overview of State-of-the-art Techniques

Relaxed multichannel least-squares (RMCLS) [Zhang 2010] aims to set the reverberant tail of the EIR to **0**

Cost function

$$\min_{\mathbf{g}} J_{\mathrm{RMCLS}} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{g} - \mathbf{d})\|_2^2$$

$$\mathbf{W} = \mathrm{diag}\{[\underbrace{1\ 0\ \ldots\ 0}_{L_d}\ 1\ \ldots\ 1]\}, \quad \mathbf{d} = [1\ 0\ \ldots\ 0]^T$$

Overview of State-of-the-art Techniques

Relaxed multichannel least-squares (RMCLS) [Zhang 2010] aims to set the reverberant tail of the EIR to **0**

Cost function

$$\min_{\mathbf{g}} J_{\mathrm{RMCLS}} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{g} - \mathbf{d})\|_2^2$$

$$\mathbf{W} = \mathrm{diag}\{[\underbrace{1\ 0\ \ldots\ 0}_{L_d}\ 1\ \ldots\ 1]\}, \quad \mathbf{d} = [1\ 0\ \ldots\ 0]^T$$

Solution $\mathbf{g}_{\mathrm{RMCLS}} = (\mathbf{W}\hat{\mathbf{H}})^{+}\mathbf{W}\mathbf{d}$

Overview of State-of-the-art Techniques

Relaxed multichannel least-squares (RMCLS) [Zhang 2010] aims to set the reverberant tail of the EIR to **0**

Cost function

$$\min_{\mathbf{g}} J_{\text{RMCLS}} = \|\mathbf{W}(\hat{\mathbf{H}}\mathbf{g} - \mathbf{d})\|_2^2$$

$$\mathbf{W} = \text{diag}\{[\underbrace{1\ 0\ \ldots\ 0}_{L_d}\ 1\ \ldots\ 1]\},\quad \mathbf{d} = [1\ 0\ \ldots\ 0]^T$$

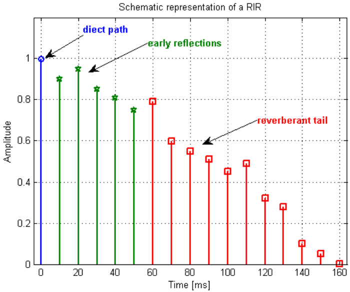
Solution $\mathbf{g}_{\text{RMCLS}} = (\mathbf{W}\hat{\mathbf{H}})^{+}\mathbf{W}\mathbf{d}$

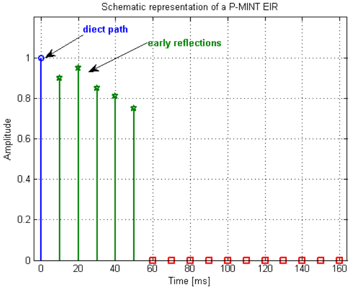
Drawback

No perceptual speech quality preservation

Overview of State-of-the-art Techniques

Regularized P-MINT [Kodrasi 2012] aims to preserve the first part of the EIR and set the reverberant tail to **0**





Overview of State-of-the-art Techniques

Regularized P-MINT [Kodrasi 2012] aims to preserve the first part of the EIR and set the reverberant tail to **0**

Cost function

$$\min_{\mathbf{g}} J_{\mathrm{P-MINT}} = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_p^{\mathrm{d}}\|_2^2$$

$$\hat{\mathbf{h}}_p^{\mathrm{d}} = [\hat{h}_p(0) \; \ldots \; \hat{h}_p(L_d - 1) \; 0 \; \ldots \; 0]^T \quad p \in \{1, \ldots, M\}$$

Overview of State-of-the-art Techniques

Regularized P-MINT [Kodrasi 2012] aims to preserve the first part of the EIR and set the reverberant tail to **0**

Cost function

$$\min_{\mathbf{g}} J^{\mathrm{R}}_{\mathrm{P-MINT}} = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}^{\mathrm{d}}_{p}\|_2^2 + \delta\|\mathbf{g}\|_2^2$$

$$\hat{\mathbf{h}}^{\mathrm{d}}_{p} = [\hat{h}_p(0) \ \ldots \ \hat{h}_p(L_d - 1) \ 0 \ \ldots \ 0]^T \quad p \in \{1, \ldots, M\}$$

Overview of State-of-the-art Techniques

Regularized P-MINT [Kodrasi 2012] aims to preserve the first part of the EIR and set the reverberant tail to **0**

Cost function

$$\min_{\mathbf{g}} J^{\mathrm{R}}_{\mathrm{P-MINT}} = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}^{\mathrm{d}}_p\|_2^2 + \delta\|\mathbf{g}\|_2^2$$

$$\hat{\mathbf{h}}^{\mathrm{d}}_p = [\hat{h}_p(0) \ \ldots \ \hat{h}_p(L_d - 1) \ 0 \ \ldots \ 0]^T \quad p \in \{1, \ldots, M\}$$

Solution $\mathbf{g}^{\mathrm{R}}_{\mathrm{P-MINT}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I})^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{h}}^{\mathrm{d}}_p$

Overview of State-of-the-art Techniques

Regularized P-MINT [Kodrasi 2012] aims to preserve the first part of the EIR and set the reverberant tail to $\mathbf{0}$

Cost function

$$\min_{\mathbf{g}} J^{\mathrm{R}}_{\mathrm{P-MINT}} = \|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}^{\mathrm{d}}_p\|_2^2 + \delta\|\mathbf{g}\|_2^2$$

$$\hat{\mathbf{h}}^{\mathrm{d}}_p = [\hat{h}_p(0)\ \ldots\ \hat{h}_p(L_d-1)\ 0\ \ldots\ 0]^T \quad p \in \{1,\ \ldots,\ M\}$$

Solution $\mathbf{g}^{\mathrm{R}}_{\mathrm{P-MINT}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I})^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{h}}^{\mathrm{d}}_p$

Drawback

The perceptual speech quality is dependent on the regularization parameter δ

Selection of the Regularization Parameter

Intrusive selection procedure

- Compute reshaping filters $\mathbf{g}^k$ for a set of regularization parameters δ^k
- Compute the true EIRs $\mathbf{c}^k = \mathbf{H}\mathbf{g}^k$
- Compute the PESQ scores of $\hat{s}(n) = s(n) * c^k(n)$
- Select δ_{opt} as the parameter yielding the highest PESQ score

Selection of the Regularization Parameter

Intrusive selection procedure

- Compute reshaping filters $\mathbf{g}^k$ for a set of regularization parameters δ^k
- Compute the true EIRs $\mathbf{c}^k = \mathbf{H}\mathbf{g}^k$
- Compute the PESQ scores of $\hat{s}(n) = s(n) * c^k(n)$
- Select δ_{opt} as the parameter yielding the highest PESQ score

✘ Not applicable in practice since knowledge of the true RIRs is required

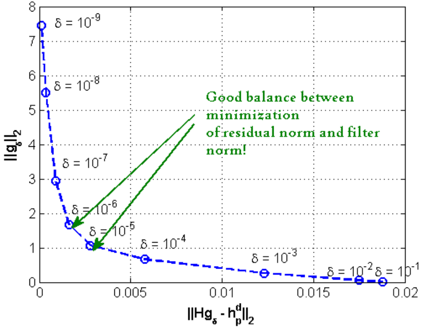
⇨ An automatic non-intrusive procedure is required

Regularized P-MINT Revisited

Cost function $J^R_{\mathrm{P-MINT}} = \underbrace{\|\hat{\mathbf{H}}\mathbf{g}_\delta - \hat{\mathbf{h}}^{\mathrm{d}}_p\|_2^2}_{\text{Residual energy}} + \delta \underbrace{\|\mathbf{g}_\delta\|_2^2}_{\text{Filter energy}}$

Trade-off between residual energy and filter energy

⇒ *L*-shaped curve



Non-intrusive Regularization

Objective Find the regularization parameter which minimizes both the residual and filter energy

Method

- Plot $\eta(\delta)$ vs $\rho(\delta)$ for a set of regularization parameters

$$\eta(\delta) = \|\mathbf{g}_\delta\|_2 \quad \rho(\delta) = \|\hat{\mathbf{H}}\mathbf{g}_\delta - \hat{\mathbf{h}}_p^{\mathrm{d}}\|_2$$

- Locate the maximum curvature

$$\kappa(\delta) = \frac{\rho'\eta'' - \rho''\eta'}{[(\rho')^2 + (\eta')^2]^{\frac{3}{2}}}$$

Non-intrusive Regularization

Objective Find the regularization parameter which minimizes both the residual and filter energy

Method

- Plot $\eta(\delta)$ vs $\rho(\delta)$ for a set of regularization parameters

$$\eta(\delta) = \|\mathbf{g}_\delta\|_2 \quad \rho(\delta) = \|\hat{\mathbf{H}}\mathbf{g}_\delta - \hat{\mathbf{h}}_p^{\mathrm{d}}\|_2$$

- Locate the maximum curvature

$$\kappa(\delta) = \frac{\rho'\eta'' - \rho''\eta'}{[(\rho')^2 + (\eta')^2]^{\frac{3}{2}}}$$

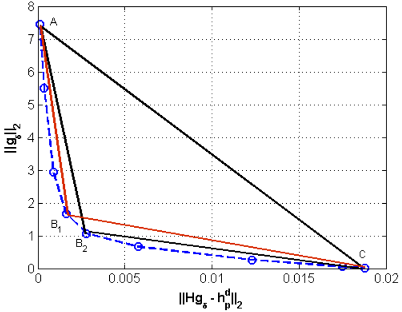
Maximum curvature location method

✘ Analytical maximum curvature location [numerical errors]

⇨ Geometrical maximum curvature location

Geometrical Maximum Curvature Location

Triangle method [Castellanos 2002]



Method summary

- Plot the filter norm vs the residual norm for a set of δ values
- Construct triangles AB_iC
- $\angle B_i \approx 90^\circ \Rightarrow$ maximum curvature

Experimental Results

Simulation parameters

- Measured 2-channel system with $T_{60} \approx 600$ ms as the true system to be equalized
- Simulated estimated RIRs $\hat{h}_m(n) = h_m(n)[1 + e(n)]$
- Normalized channel mismatch $E_m = 10 \log_{10} \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2^2}{\|\mathbf{h}_m\|_2^2}$

$$E_m \in \{-33 \text{ dB},\ -30 \text{ dB},\ -25 \text{ dB},\ -20 \text{ dB},\ -15 \text{ dB}\}$$

- $f_s = 16000$ Hz, $L_h = 2000$, $L_g = 1999$, $p = 1$
- Desired window lengths L_d: 10 ms, 20 ms, 30 ms, 40 ms, 50 ms

Experimental Results

Simulation parameters

- Measured 2-channel system with $T_{60} \approx 600$ ms as the true system to be equalized
- Simulated estimated RIRs $\hat{h}_m(n) = h_m(n)[1 + e(n)]$
- Normalized channel mismatch $E_m = 10 \log_{10} \frac{\|\mathbf{h}_m - \hat{\mathbf{h}}_m\|_2^2}{\|\mathbf{h}_m\|_2^2}$

$$E_m \in \{-33 \text{ dB},\ -30 \text{ dB},\ -25 \text{ dB},\ -20 \text{ dB},\ -15 \text{ dB}\}$$

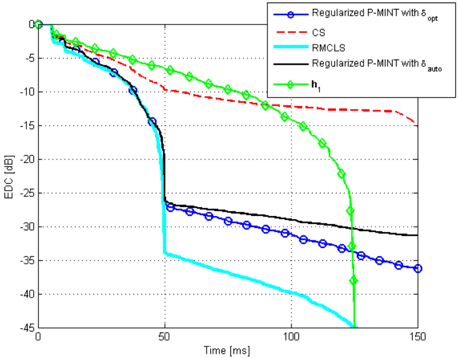
- $f_s = 16000$ Hz, $L_h = 2000$, $L_g = 1999$, $p = 1$
- Desired window lengths L_d: 10 ms, 20 ms, 30 ms, 40 ms, 50 ms

Performance measures

- Reverberant tail suppression: Energy decay curve
- Perceptual speech quality: PESQ [reference $= s(n) * h_1^{\mathrm{d}}(n)$]

Experimental Results

Reverberant tail suppression ($L_d = 50$ ms, $E_m = -33$ dB)

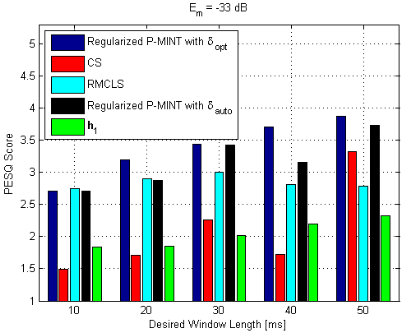


Non-intrusive regularization

yields similar reverberant tail suppression as the intrusive one

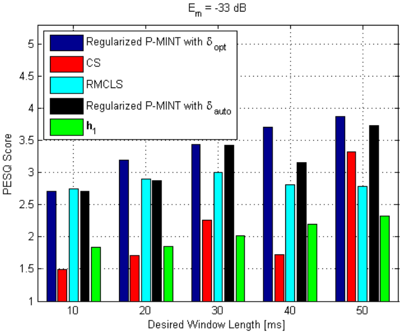
Experimental Results

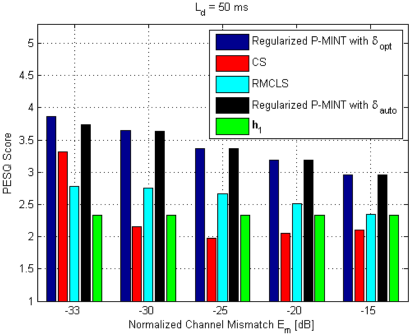
Perceptual speech quality



Experimental Results

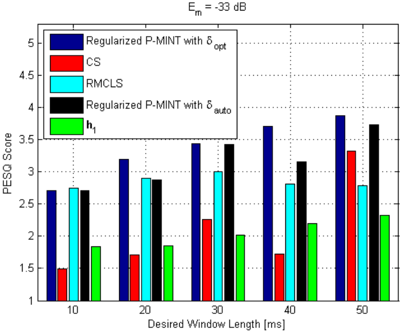
Perceptual speech quality

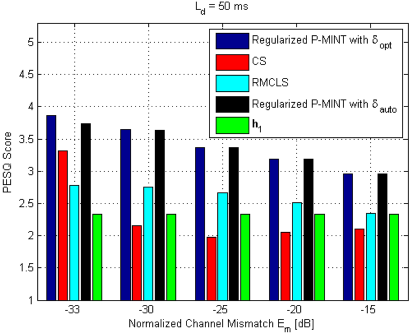




Experimental Results

Perceptual speech quality





Non-intrusively regularized P-MINT

- yields similar perceptual quality as the intrusively regularized P-MINT
- typically outperforms state-of-the-art techniques

Summary

- **Acoustic Multichannel Equalization Techniques**
  - Channel shortening
  - Relaxed multichannel least-squares
  - Regularized P-MINT
- **Regularization Parameter Selection**
  - Intrusive
  - Non-intrusive
- **Non-intrusive Selection Procedure**
  - Achieves similar performance as the intrusive one
  - Can be used for any regularized least-squares technique

Summary

- **Acoustic Multichannel Equalization Techniques**
  - Channel shortening
  - Relaxed multichannel least-squares
  - Regularized P-MINT
- **Regularization Parameter Selection**
  - Intrusive
  - Non-intrusive
- **Non-intrusive Selection Procedure**
  - Achieves similar performance as the intrusive one
  - Can be used for any regularized least-squares technique

Summary

- **Acoustic Multichannel Equalization Techniques**
  - Channel shortening
  - Relaxed multichannel least-squares
  - Regularized P-MINT
- **Regularization Parameter Selection**
  - Intrusive
  - Non-intrusive
- **Non-intrusive Selection Procedure**
  - Achieves similar performance as the intrusive one
  - Can be used for any regularized least-squares technique