Spatial Sensitivity for Listening Room Compensation

S. Goetze, M. Kallinger, K.-D. Kammeyer, and A. Mertins

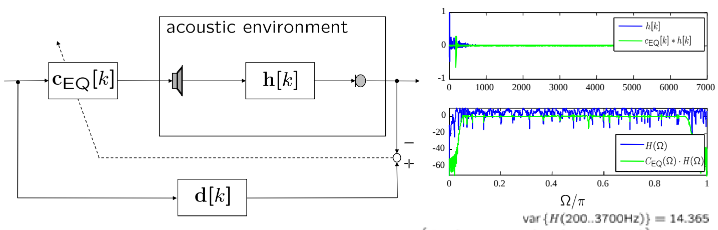
Universität Bremen — Arbeitsbereich Nachrichten Technik
ISIP Institute for Signal Processing, University of Lübeck
Fraunhofer Institut Integrierte Schaltungen IIS

Motivation

- Listening Room Compensation (LRC) / Room Impulse Response Shaping can increase speech intelligibility.
- Since LRC devices are designed for fixed positions of lousdpeaker and microphone spatial mismatch degrades performance.
- This contribution demonstrates the effects of spatial mismatch visually and acoustically.

Listening Room Compensation

- An equalizer precedes the acoustic channel
- Common design method: Least Squares Equalizer $\mathbf{c}_{\text{EQ}} = \mathbf{H}^{+}\mathbf{d}$
- Problem: Channel $\mathbf{h}[k]$ is needed!



$\text{var}\{H(200..3700\text{Hz})\} = 14.365$

$\text{var}\{C_{\text{EQ}}(200..3700\text{Hz}) \cdot H(200..3700\text{Hz})\} = 1.0379$

- The desired system $\mathbf{d}[k]$ is approximated by the overall system of $\mathbf{c}_{\text{EQ}}[k] * h[k]$

Room Impulse Response Shaping

- The goal is not spectral flatness of the overall system but a concentration of the energy at a specified temporal envelope (desired area $\mathbf{d}_d$).

$$\mathbf{d}_d = \text{diag}\{\mathbf{w}_d\}\mathbf{H}\mathbf{c}_{\text{EQ}}$$

$$\mathbf{d}_u = \text{diag}\{\mathbf{1} - \mathbf{w}_d\}\mathbf{H}\mathbf{c}_{\text{EQ}}$$

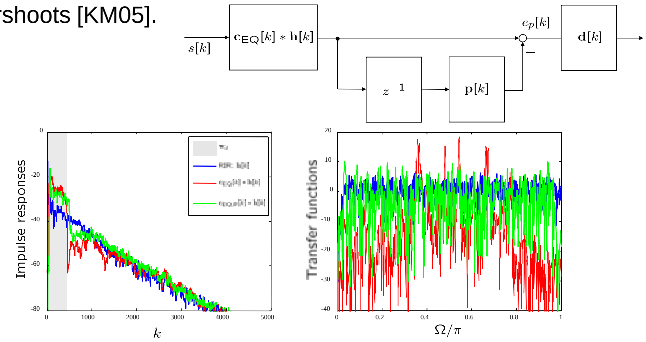
- Maximization of the energy of $\mathbf{d}_d$ while keeping the energy of $\mathbf{d}_u$ constant leads to the impulse response shortener after Melsa [MYR96, MMEJ03].

$$\mathbf{B}_{\text{BP}} \cdot \mathbf{c}_{\text{EQ,opt}} = \mathbf{A} \cdot \mathbf{c}_{\text{EQ,opt}} \cdot \lambda_{\text{max}}$$

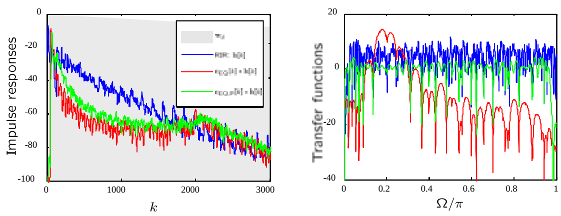
$$\mathbf{A} = \mathbf{H}^{H}\text{diag}\left\{\mathbf{w}_{\text{BP,d}}\right\}^{2}\mathbf{H}$$

$$\mathbf{B}_{\text{BP}} = \mathbf{H}_{\text{BP}}^{H}\text{diag}\left\{\mathbf{w}_{\text{BP,d}}\right\}^{2}\mathbf{H}_{\text{BP}}$$

- Problem: spectral peaks occur in overall transfer function!
- Post processing by a linear prediction filter can reduce the spectral overshoots [KM05].



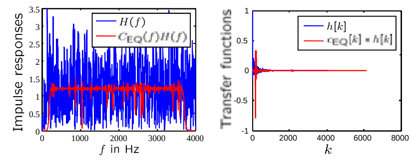
- For reducing the problem of late echoes the impulse response should better be *shaped* than shortened [KM05].
- This can be done by a exponential decreasing window.



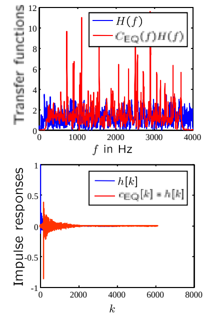
Spatial Sensitivity

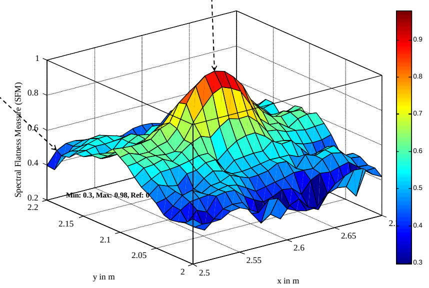
- LRC devices are usually designed for fixed and a-priori known positions of source and microphone.
- If the assumptions of the spatial configuration are violated severe distortions may occur.

A known RIR leads to a good equalization.



For spatial mismatch severe distortions may occur!



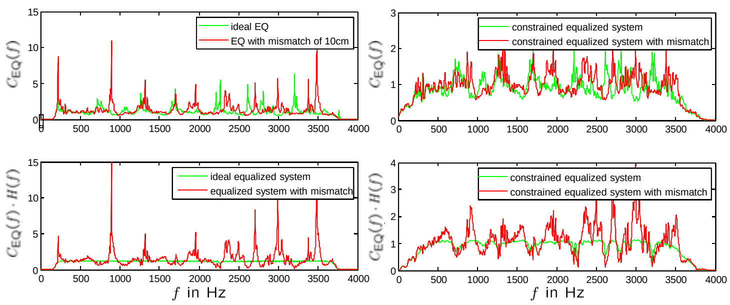


- Spectral Flatness Measure (SFM) for a LS-EQ, order 2048, τ_{60} = 200 ms, position 3

Constrained LS-EQ design

- If an estimation error for the transfer function exist $\mathbf{H} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}$ the least squares EQ can be modified considering these errors.

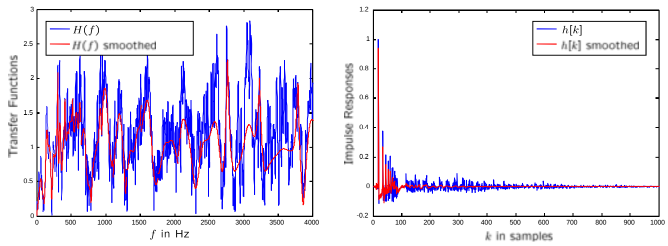
$$\mathbf{c}_{\text{EQ}} = (\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \underbrace{\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}}_{\approx\delta \cdot ||\hat{\mathbf{h}}||^2\mathbf{I}})^{-1}\hat{\mathbf{H}}\mathbf{d}$$



- Constraining the least squares design enhances spatial robustness.
- Spectral peaks due to spatial mismatch can be strongly reduced.

Impulse Response Smoothing

- Complex Fractional Octave Smoothing as proposed in [HM00] is also capable to increase spatial robustness.



- RIR shaping and constrained LS-EQ design lead to spatially more robust LRC designs than Complex Smoothing.

Universität Bremen
Arbeitsbereich Nachrichten Technik

Spatial Sensitivity for Listening Room Compensation

S. Goetze, M. Kallinger, K.-D. Kammeyer, and A. Mertins

ISIP Institute for Signal Processing University of Lübeck

Fraunhofer Institut Integrierte Schaltungen

Quality Assessment

- Subjective quality assessment is time consuming and expensive.
- Reliable and commonly accepted quality measurement (of enhancement) for dereverberation algorithms has yet to be found.

Objective Quality Measures for LRC

- Several measures can be found in the literature to evaluate dereverberation algorithms:

Measures based on the transfer function:

- Variance of logarithmic transfer function [Mou94]:

$$\sigma^2_{\mathcal{H}} = \frac{1}{n_{max} - n_{min} + 1} \sum_{n=n_{min}}^{n_{max}} \left(20 \cdot \log_{10} |\mathcal{H}[n]| - \bar{\mathcal{H}}_{\text{dB}}\right)^2$$

$$\bar{\mathcal{H}}_{\text{dB}} = \frac{1}{n_{max} - n_{min} + 1} \sum_{n=n_{min}}^{n_{max}} 20 \cdot \log_{10} |\mathcal{H}[n]|$$

$\mathcal{H}[n] = H[n] C_{\text{EQ}}[n]$: concatenated system of EQ and RTF
n : discrete frequency index
n_{min} : frequency index corresponding to $f = 200\text{Hz}$
n_{min} : frequency index corresponding to $f = 3600\text{Hz}$

- Spectral Flatness Measure (SFM), [Joh88]

$$\text{SFM} = \frac{G_n}{A_n} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} |\mathcal{H}[n]|^2}}{\frac{1}{N}\sum_{n=0}^{N-1} |\mathcal{H}[n]|^2}$$

$G[n]$: Geometric mean
$A[n]$: Arithmetic mean

Measures based on the impulse response:

- “Deutlichkeit“ (D50), [Kut00]

$$D_{50} = \frac{\sum_{k=0}^{k_{50}-1} (c_{\text{EQ}}[k] * h[k])^2}{\sum_{k=0}^{L_h} (c_{\text{EQ}}[k] * h[k])^2}$$

$k_{50} = 50\text{ms}/f_s$: discrete time index corresponding to a time of 50ms

- Clarity Index (CI), [Kut00]

$$\text{CI} = \frac{\sum_{k=0}^{k_{80}} (c_{\text{EQ}}[k] * h[k])^2[k]}{\sum_{k=k_{80}}^{L_h} (c_{\text{EQ}}[k] * h[k])^2[k]}$$

$k_{80} = 80\text{ms}/f_s$: discrete time index corresponding to a time of 80ms

- Central Time (CT), [Kut00]

$$\text{CT} = \frac{\sum_{k=0}^{L_h} k (c_{\text{EQ}}[k] * h[k])^2[k]}{\sum_{k=0}^{L_h} (c_{\text{EQ}}[k] * h[k])^2[k]}$$

- Direct-to-Reverberation-Ratio (DRR), [TS06]

$$\text{DRR}_{\text{dB}} = 10\log_{10} \frac{(c_{\text{EQ}}[k_{h_{max}}] * h[k_{h_{max}}])^2}{\sum_{k \neq k_{h_{max}}} (c_{\text{EQ}}[k] * h[k])^2}$$

$k_{h_{max}}$: index of maximum of $h[k]$

Signal-based measures:

- Segmental Signal-to-Reverberation Ratio (SSRR), [NG05]

$$\text{SSRR}_{\text{dB}} = \frac{1}{K/L} \sum_{\ell=0}^{K/L-1} 10\log_{10} \frac{\sum_{k=0}^{L-1} \tilde{y}[\ell L + k]^2}{\sum_{k=0}^{L-1} (\tilde{y}[\ell L + k] - y[\ell L + k])^2}$$

K :signal length
L :block length
ℓ :block index
$\tilde{y}[k]$:signal after desired system $\mathrm{d}[k]$
$y[k]$:microphone signal

- Bark Spectral Distortion (BSD), [Yan99]

$$\text{BSD} = \frac{\frac{1}{K/L} \sum_{\ell=0}^{K/L-1} \sum_{i=1}^{Q} \left[L_{\tilde{y}}^{(\ell)}(i) - L_{y}^{(\ell)}(i)\right]^2}{\frac{1}{K/L} \sum_{\ell=0}^{K/L-1} \sum_{i=1}^{Q} \left[L_{\tilde{y}}^{(\ell)}(i)\right]^2}$$

Q : No. of critical bands
i : No. of critical band
$L_{\tilde{y}}(\ell)$: Bark spectrum of frame ℓ
$L_{y}(\ell)$: Bark spectrum of frame ℓ

- Cepstral Distance (CD), [Yan99]

$$\text{CD} = \frac{1}{K/L} \sum_{\ell=1}^{K/L-1} \left\{ \frac{10}{\ln 10} \sqrt{\left[C_{\text{LP}}^{(\tilde{y})}(1,\ell) - C_{\text{LP}}^{(y)}(1,\ell)\right]^2 + 2\sum_{i=2}^{m} \left[C_{\text{LP}}^{(\tilde{y})}(i,\ell) - C_{\text{LP}}^{(y)}(i,\ell)\right]^2} \right\}$$

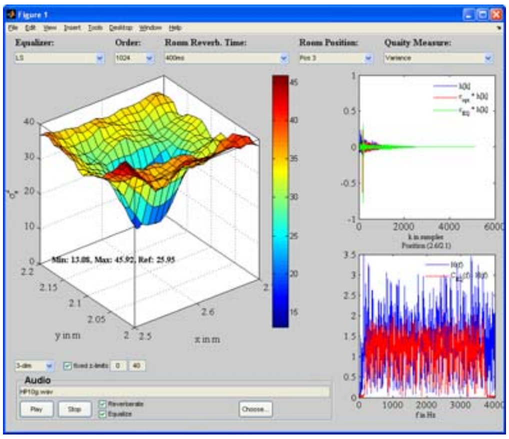
$$C_{\text{LP}}^{(x)}(i) = a_i + \frac{1}{i} \sum_{j=1}^{i-1} (i-j) \cdot C_{\text{LP}}^{(x)}(i-j) \cdot a_j \quad \text{for } i = 1, 2, 3 \ldots$$

$$\mathbf{a} = -\mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{r}_{\mathbf{xx}}$$

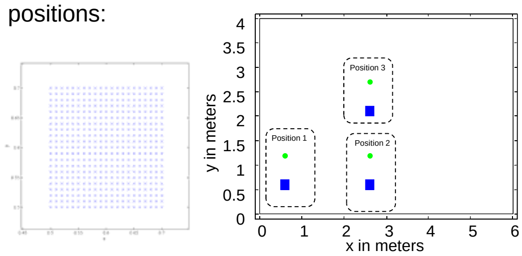
$C_{\text{LP}}^{(y)}$: Cepstral Coefficients for signal y
i : Cepstral Coefficient index
$\mathbf{a}$: Coefficients of Linear Predictor

MATLAB® Demo

- The different LRC schemes can be compared by means of various objective measures (see left column). Furthermore a subjective quality assessment is possible by applying the precomputed equalizers to a given sound-file and by listening to the equalized sound-file acoustically.



- Possible EQ designs: LS-EQ, Constrained LS-EQ, RIR smoothing, RIR shaping
- Possible EQ orders (@8kHz): 128, 256, 512, 1024, 2048
- Possible room reverberation times: 50ms, 100ms, 200ms, 400ms, 800ms, 1200ms
- Possible spatial positions:



References

[HM00] P. Hatziantoniou and J.N. Mourjopoulos: **“Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses”**, J. of the Audio Engineering Society , vol. 48, no. 4, pp. 259-280, April 2000

[Joh88] J. D. Johnston**: “Transform Coding of Audio Signals using Perceptual Noise Criteria”**, IEEE Journal on Selected Areas in Communication, vol. 6, no. 2, pp. 314-232, Feb. 1988

[KM05] M. Kallinger and A. Mertins: **“Room Impulse Response Shortening by Channel Shortening Concepts”**, Proc. of Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, USA, 2005

[Kut00] H. Kuttruff: **“Room Acoustics”**, 4th Edition, 2000

[Mou94] J. N. Mourjopoulos. **“Digital Equalization of Room Acoustics”**, Journal of the Audio Engineering Society, vol. 42, no.11, pp. 884–900, Nov. 1994.

[MYR96] P.J.W. Melsa, R.C. Younce, and C.E. Rohrs: “Impulse Response Shortening for Discrete Multitone Transceivers”, IEEE Trans. On Communications, vol. 44, no. 12, pp. 1662-1672, Dec. 1996

[MMEJ03] R.K. Martin, D. Ming, B.L. Evans, and C.R. Johnson Jr.: “Efficient Channel Shortening Equalizer Design”, J. on Applied Signal Processing, vol. 13, pp 1279-1290, Dec. 2003

[NG05] P.A. Naylor, and N.D. Gaubitch: **“Speech Dereverberation”**, in Proc. of the International Workshop on Acoustic Echo and Noise Control, (IWAENC’05), Eindhoven, The Netherlands, 2005

[TS06] M. Triki and D.T.M. Slock: **“Iterated Delay and Predict Equalization for Blind Speech Dereverberation”**, in Proc. of the International Workshop on Acoustic Echo and Noise Control, (IWAENC’06), Paris, France, Sep. 2006.

[WSG92] S. Wang and A. Sekey and A. Gersho: **“An Objective Measure for Predicting Subjective Quality of Speech Coders”**, IEEE J. on Selected Areas of Communications, vol. 10, no. 5, pp. 819-829, Feb. 1992

[Yan99] W. Yang, “**Enhanced Modified Bark Spectral Distortion (EMBSD): A Objective Speech Quality Measure Based on Audible Distortion and Cognition Model**”, PhD-thesis, Temple University, Philadelphia, USA, May 1990