Quality Assessment for Listening-Room Compensation Algorithms

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

Fraunhofer Institute for Digital Media Technology (IDMT), Project Group Hearing, Speech and Audio Technology (HSA), Oldenburg, Germany

University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg, Germany

University of Lübeck, Institute for Signal Processing, Lübeck, Germany

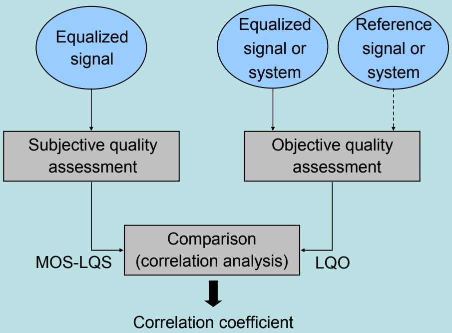
University of Bremen, Dept. Of Communication Engineering, Bremen, Germany

Motivation

- Listening-room compensation (LRC) / room impulse response shaping is capable to increase speech intelligibility
- LRC algorithms may introduce distortions
  - Small in Amplitude but clearly perceivable
- Commonly accepted objective quality measures not available
- This contribution analyses objective quality measures for LRC

Quality Assessment

- Subjects / humans assess quality based on their internal reference
- Objective quality measures (mostly intrusive) need reference signal or system
- Goal: find a quality measure that shows high correlation to subjective rating!

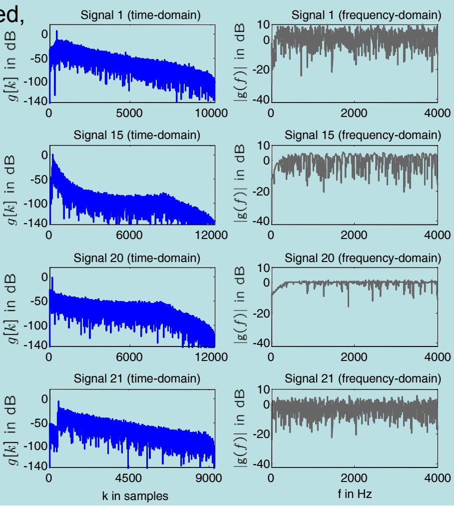


Subjective Listening Tests

- Room reverberation time: {500, 1000} ms
- Room size: 6 m x 4 m x 2.6 m
- Loudspeaker-microphone distance: 0.8 m
- EQ lengths: 1024, 2048, 4096, 8192 at sampling rate of 8 kHz
- 19 audio samples (male and female) of 8 sec
- 24 normal-hearing listeners

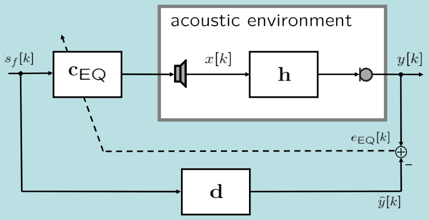
- Assessment of four attributes: reverberant, colored/distorted, distant, overall quality
- Different LRC types:
  - Least-squares equalizer (LS-EQ)
  - Weighted least-squares EQ (WLS-EQ)
  - Impulse response shaper with spectral post-processing (ISwPP)

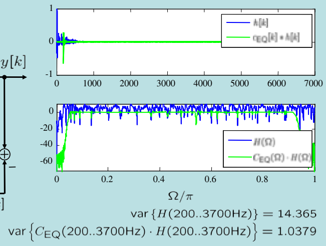
Signal	τ_{60} of RIR	Equalizer	L_{EQ}	Signal
1	800 ms	*WLS*	2048	male
⋮	⋮	⋮	⋮	⋮
15	400 ms	*WLS*	8192	male
⋮	⋮	⋮	⋮	⋮
20	400 ms	*LS*	8192	female
21	800 ms	*ISwPP*	1024	male



Listening-Room Compensation

- An equalizer precedes the acoustic channel
- Common design method: Least-squares equalizer: $\mathbf{c}_{\mathrm{EQ}} = \mathbf{H}^{+}\mathbf{d}$
- Knowledge of channel $\mathbf{h}$ is needed!

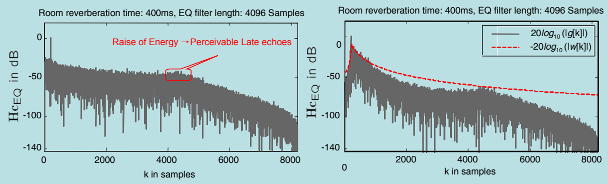




- The desired system $\mathbf{d}$ is approximated by the overall system of $\mathbf{H}\mathbf{c}_{\mathrm{EQ}}$
- For reducing the problem of late echoes the impulse response should better be *shaped* than equalized to flat transfer function
- This can be done by a exponential decreasing window.

$$\mathbf{e}_{\mathrm{EQ}} = \mathbf{W}(\mathbf{H}\mathbf{c}_{\mathrm{EQ}} - \mathbf{d})$$

$$\mathbf{c}_{\mathrm{EQ}} = (\mathbf{W}\mathbf{H})^{+}\mathbf{W}\mathbf{d}$$



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

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

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

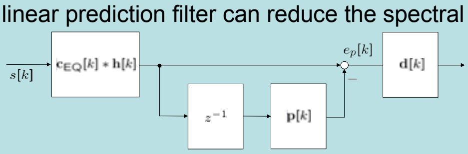
- Maximization of the energy of $\mathbf{d}_d$ while keeping the energy of $\mathbf{d}_u$ constant leads to impulse response shaper.

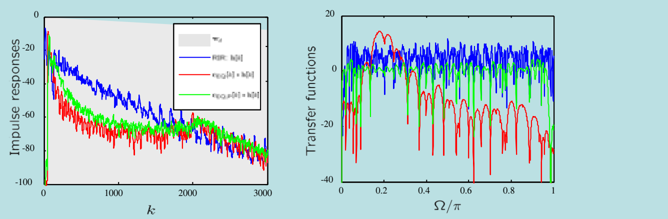
$$\mathbf{B}_{\mathrm{BP}} \cdot \mathbf{c}_{\mathrm{EQ,opt}} = \mathbf{A} \cdot \mathbf{c}_{\mathrm{EQ,opt}} \cdot \lambda_{\max}$$

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

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

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





Fraunhofer IDMT

Quality Assessment for Listening-Room Compensation Algorithms

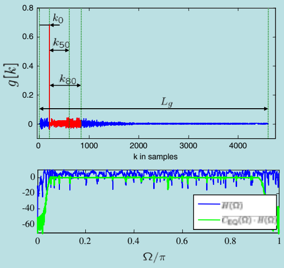
[1]Stefan Goetze, [1]Eugen Albertin, [2]Markus Kallinger, [3]Alfred Mertins, and [4]Karl-Dirk Kammeyer

Fraunhofer Institute for Digital Media Technology (IDMT), Project Group Hearing, Speech and Audio Technology (HSA), Oldenburg, Germany
University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg, Germany
University of Lübeck, Institute for Signal Processing, Lübeck, Germany
University of Bremen, Dept. Of Communication Engineering, Bremen, Germany

Channel-based Objective Quality Measures for LRC

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

Channel-based measures			
Acronym	**Measure**	**Acronym**	**Measure**
D_{50}	Definition (50 ms)	CT	Center Time
D_{80}	Definition(80ms)	DRR	Direct to Reverberation Ratio
C_{80}	Clarity Index (80ms)	SFM	Spectral Flatness Measure
C_{50}	Clarity Index (50ms)	VAR	Spectral Variance



Definition

$$D_{50} = \frac{\sum_{k=k_0}^{k_0+k_{50}-1} g^2[k]}{\sum_{k=0}^{L_g-1} g^2[k]}$$

Direct-to-Reverberation Ratio

$$\mathrm{DRR}_{\mathrm{dB}} = 10\log_{10}\frac{g^2[k_{h_{max}}]}{\sum_{k\neq k_{h_{max}}} g^2[k]}$$

Clarity

$$C_{80} = 10\log_{10}\frac{\sum_{k=k_0}^{k_0+k_{80}-1} g^2[k]}{\sum_{k=0}^{k_0-1} g^2[k] + \sum_{k=k_0+k_{80}}^{L_g-1} g^2[k]}$$

Spectral Flatness Measure

$$\mathrm{SFM} = \frac{\sqrt[N]{\prod_{n=0}^{N-1}|G[n]|^2}}{\frac{1}{N}\sum_{n=0}^{N-1}|G[n]|^2}$$

Central Time

$$CT_g = \frac{\sum_{k=0}^{L_g-1} g^2[k]\cdot k}{\sum_{k=0}^{L_g-1} g^2[k]}$$

Variance of transfer function

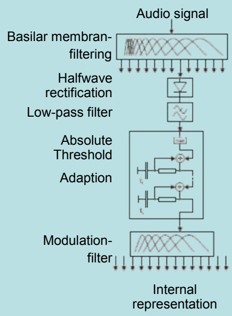
$$\mathrm{VAR} = \frac{1}{n_{max}-n_{min}+1}\sum_{n=n_{min}}^{n_{max}}\left(20\cdot\log_{10}|G[n]| - \bar{G}_{\mathrm{dB}}\right)^2$$

- Channel-based measures evaluate energy ratios of early and late part of impulse response or spectral flatness of transfer function

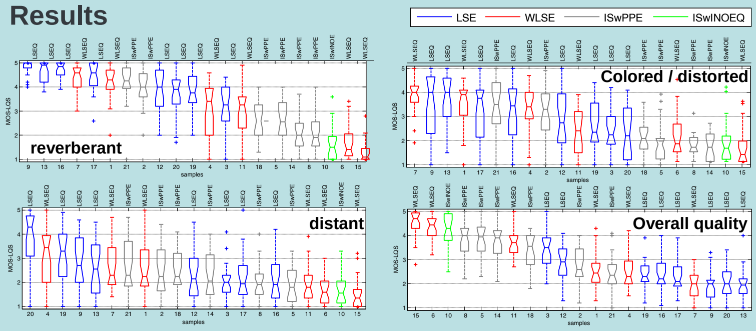
Signal-based Objective Quality Measures for LRC

Signal-based measures	
Acronym	**Measure**
SSRR	Segmental Signal to Reverberation Ratio
SRRE	Signal to Reverberation Ratio Enhancement
FWSSRR	Frequency Weighted SSRR
WSS	Weighted Spectral Slope
OMCR	Objective Measure of Colouration in Reverberation
IS, CEP	Itakura-Saito-Distance, Cepstral Distance
LAR, LLR	Log Area Ratio, Log Likelihood Ratio
LSD	Log Spectral Distortion
BSD	Bark Spectral Distortion
R_{DT}	Reverberation Decay Tail Measure
PSM	Perceptual Similarity Measure
PSM_t	Perceptual Similarity Measure (time)
ΔPSM	PSM enhancement
ΔPSM_t	PSM_t enhancement
PESQ	Perceptual Evaluation of Speech Quality
SRMR	Speech to Reverberation Modulation Energy Ratio

- Signal-based measures partly incorporate models of human auditory systems:

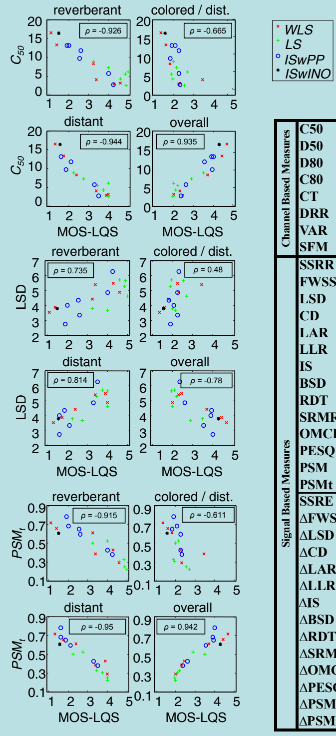


Results



- Good ratings for WLS-EQ and shaping approaches
- LS-EQ shows lower performance than IR shaping even for high filter lengths

Correlation analysis



		Reverberated	Colored/Dist.	Distant	Overall
Channel Based Measures	C50	**0,93**	**0,67**	**0,94**	**0,94**
	D50	0,86	0,63	**0,94**	0,91
	D80	0,90	0,50	0,91	0,90
	C80	**0,93**	0,61	0,89	0,91
	CT	0,85	0,61	0,93	0,91
	DRR	0,24	0,10	0,18	0,13
	VAR	0,03	0,37	0,23	0,16
	SFM	0,13	0,27	0,13	0,05
Signal Based Measures	SSRR	0,33	0,29	0,43	0,40
	FWSSRR	0,44	0,40	0,57	0,55
	LSD	0,74	0,48	0,81	0,78
	CD	0,63	0,41	0,70	0,67
	LAR	0,52	0,38	0,61	0,59
	LLR	0,66	0,43	0,75	0,71
	IS	0,64	0,35	0,69	0,68
	BSD	0,04	0,30	0,24	0,20
	RDT	0,67	0,51	0,79	0,75
	SRMR	0,53	0,24	0,59	0,51
	OMCR	0,05	0,13	0,03	0,05
	PESQ	0,60	0,35	0,69	0,63
	PSM	0,80	**0,63**	0,90	0,87
	PSMt	**0,91**	0,61	**0,95**	**0,94**
	SSRE	0,00	0,14	0,02	0,03
	ΔFWSSRR	0,15	0,04	0,11	0,09
	ΔLSD	0,07	0,06	0,03	0,03
	ΔCD	0,52	0,37	0,47	0,49
	ΔLAR	0,24	0,23	0,25	0,26
	ΔLLR	0,50	0,31	0,46	0,45
	ΔIS	0,46	0,16	0,37	0,42
	ΔBSD	0,66	0,25	0,57	0,60
	ΔRDT	0,67	0,51	0,71	0,72
	ΔSRMR	0,42	0,14	0,45	0,36
	ΔOMCR	0,52	0,24	0,45	0,43
	ΔPESQ	0,41	0,18	0,43	0,37
	ΔPSM	0,44	0,41	0,49	0,47
	ΔPSMt	0,84	0,53	0,85	0,86

- IR-based measures show high correlation ⊕ ‘Winner’: C50
- Measures based on transfer functions show lower correlation ⊖
  - Distortions in time-domain are perceptually prominent
- Low correlation of all measures with dimension coloration/distortion ⊖
  - Measures for coloration assessment assess pure coloration only
- Δ measures show low correlation ⊖
- Simple signal-based measures like SSRR show low correlation ⊖
- Signal-based measures based on human speech perception better ⊕
  - At least speech production models (like for LSD) should be used
- ‘Winner’ PSM incorporates model of human auditory system ⊕

Conclusions

- Performance of various objective quality measures for LRC was analysed:
- LS-EQ shows lower performance than IR shaping even for high filter lengths; distortions in time-domain perceptually disturbing
- Channel-based measures show high correlation with subjective data
  - C50 and D50 show high correlation
- Coloration is difficult to assess due to perceptually relevant distortions
  - Late echoes and pre-echoes
- If channel is not available (e.g. for reverberation suppression) objective measures relying on auditory models should be used
  - Perceptual Similarity Measure (PSM) shows highest correlation to our test data

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