QUALITY ASSESSMENT FOR LISTENING-ROOM COMPENSATION ALGORITHMS

*Stefan Goetze, *Eugen Albertin, †Markus Kallinger, ‡Alfred Mertins, and §Karl-Dirk Kammeyer

ABSTRACT

In this contribution various objective measures that can be used to evaluate speech dereverberation algorithms by means of listening-room compensation (LRC) are compared to subjective listening tests. It is shown that technical measures describing the impulse responses are suitable for evaluation of such algorithms. Most signal-based objective measures fail to judge the specific distortions that may be introduced by LRC algorithms like late reverberation since these artifacts are small in amplitude but perceptually relevant due to the loss of masking of the room impulse response. Only one signal-based measure, the so-called perceptual similarity measure (PSM), showed high correlation with subjective rating for the given test setup.

Index Terms— Quality Assessment, Listening-Room Compensation, Equalization, Dereverberation

1. INTRODUCTION

Reverberation is caused in enclosed spaces by numerous reflections of a signal emitted by a sound source until it is picked up by a microphone or the human ear. Reverberation decreases speech intelligibility and reverberant sound is perceived as sounding distant and echoic. Although speech dereverberation by means of listening-room compensation (LRC) has been research topic for some years now [1, 2], no commonly accepted objective measure for evaluating such algorithms has established itself [3, 4, 5, 6]. Often the performance of dereverberation algorithms is judged by simple technical measures which are based on energy ratios as it is common in the field of noise reduction. However, lots of more sophisticated measures have been developed recently incorporating psychoacoustical knowledge [7] or even complex models of the human auditory system [8, 9]. Since the perception of the sound quality of dereverberated signals is multidimensional, at least the dimensions coloration [10] and what is called the reverberation tail effect (reverberation without the influence of a spectral modification) have to be considered [11]. However, for the LRC algorithms that are analyzed in this paper these two dimensions (influences in spectral domain and in time domain) may not be sufficient. For example, late reverberation introduced by least-squares approaches may be mathematically negligible since it is of small amplitude but it is perceptually relevant due to the loss of temporal masking of the impulse response. Furthermore, measures that were developed to judge coloration [2, 10] may fail if spectral distortion is perceptually more relevant to subjective listeners.

In this contribution, different objective measures that are expected to be able to judge algorithms for LRC are compared to results from subjective listening tests and evaluated w.r.t. their capability to predict important artifacts or the overall quality of the dereverberated speech sample.

The remainder of this paper is organized as follows: Section 2 briefly summarizes the methods for LRC that were used for speech dereverberation. Section 3 introduces objective measures that can principally be used for the evaluation of dereverberated signals. Section 4 describes the experimental setup for the subjective listening tests. Results and correlations between subjective and objective data are presented in Section 5, and Section 6 concludes the paper. A list of all objective measures that were evaluated and the corresponding references are given in Section 7 as a help to deal with the acronyms.

2. LISTENING-ROOM COMPENSATION ALGORITHMS

Fig. 1 shows a general setup for LRC with an equalization filter $\mathbf{c}_{\mathrm{EQ}}$ preceding the acoustic channel $\mathbf{h}$. Here, $\mathbf{c}_{\mathrm{EQ}}$ and $\mathbf{h}$ denote the coefficient vectors of the equalization filter of length L_{EQ} and the coefficient vector of the room impulse response (RIR) of length L_h, respectively [12]. Since usual RIRs are mixed-phase systems of high order direct inversion by a stable causal filter is not possible in general [1]. Thus, common least-squares approaches [12, 1] try to minimize the Euclidean distance between the concatenated system of $\mathbf{c}_{\mathrm{EQ}}$ and $\mathbf{h}$ and a given target system $\mathbf{d}$ which usually is chosen as a delay, band pass or high pass [13]. An appropriate weighting of the error vector $\mathbf{e}_{\mathrm{EQ}}$ as introduced in [14] for cross-talk cancelation can avoid perceptually disturbing late echoes by accentuating on late samples of $\mathbf{e}_{\mathrm{EQ}}$. Minimizing the weighted mean squared error signal $\mathrm{E}\{||\mathbf{e}_{\mathrm{EQ}}||^2\} = \mathrm{E}\{||\mathbf{W}(\mathbf{H}\mathbf{c}_{\mathrm{EQ}} - \mathbf{d})||^2\}$ leads to a weighted least-squares equalizer

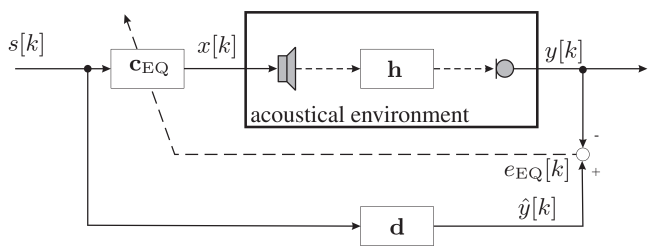

Fig. 1. General setup for listening-room compensation.

$$\mathbf{c}_{\mathrm{EQ}} = (\mathbf{W}\mathbf{H})^{+}\mathbf{W}\mathbf{d} \quad (1)$$

with $\mathbf{H}$ being the channel convolution matrix built up by the RIR coefficients and $(\cdot)^{+}$ being the Moore-Penrose pseudo-inverse. The weighting matrix $\mathbf{W}$ is a diagonal matrix containing a weighting

*Stefan Goetze (s.goetze@idmt.fraunhofer.de) and Eugen Albertin are with the Fraunhofer Institute for Digital Media Technology (IDMT), Project group Hearing Speech and Audio Technology (HSA), 26129 Oldenburg, Germany, †Markus Kallinger is with the Fraunhofer Institute for Integrated Circuits (IIS), 91058 Erlangen, Germany, he contributed to this work while he was with the University of Oldenburg, Signal Processing Group, 26111 Oldenburg, Germany, ‡Alfred Mertins is with the University of Lübeck, Institute for Signal Processing, 23538 Lübeck, Germany, §Karl-Dirk Kammeyer is with University of Bremen, Dept. of Communications Engineering, 28334 Bremen, Germany. This work has been supported in parts by the German Research Foundation DFG under Grants No. Ka841-17 and Me1170/3.

window on its main diagonal. By a proper choice of $\mathbf{W}$, RIR shortening or RIR shaping instead of a straightforward equalization can be achieved, which leads to perceptually better results.

Another approach for RIR shaping was discussed in [15] and is based on the solution of a generalized eigenvalue problem

$$\mathbf{A}\mathbf{c}'_{\mathrm{EQ}} = \lambda_{\max}\mathbf{B}\mathbf{c}'_{\mathrm{EQ}} \tag{2}$$
$$\mathbf{A} = \mathbf{H}^T\mathbf{W}_u^T\mathbf{W}_u\mathbf{H} \tag{3}$$
$$\mathbf{B} = \mathbf{H}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{H}. \tag{4}$$

Similar to (1) $\mathbf{W}_u$ and $\mathbf{W}_d$ are diagonal matrices with window functions defining a desired part of the RIR and an undesired part of the RIR. The greatest eigenvalue is denoted by $\lambda_{\max}$ in (2). To avoid spectral distortion a post processor based on liner prediction [15] is used after applying (2). For a more detailed discussion the reader is referred to [15, 16].

3. OBJECTIVE MEASURES

We classify objective measures that may be capable to evaluate speech dereverberation algorithms in two different classes: measures that are based on the (i) impulse response or the transfer function of a system (system-based measures) and (ii) measures that are based on signals. Generally, for LRC algorithms both the filter impulse response $c_{\mathrm{EQ}}[k]$ and the RIR $h[k]$ are available during simulations. However, if gradient algorithms [17] are used to avoid computational complex matrix inversions (e.g. as in (1)) or to track time-varying environments or if blind dereverberation approaches, e.g. [5], shall be applied, the necessary impulse responses of the room or the filter may not be accessible or not appropriate to apply those measures. Such situations restrict the number of applicable measures to those based on signals as described in Section 3.2.

3.1. System-Based Measures

Room impulse responses can be characterized by several objective measures, see e.g. [18]. Most of them are based on a ratio between early and late part of the impulse response. Since impulse responses of equalized systems $v[k] = h[k]*c_{\mathrm{EQ}}[k]$ may look slightly different from common RIRs, Fig. 2 shows an equalized impulse response $v[k]$. We calculated six different measures that are widely used for

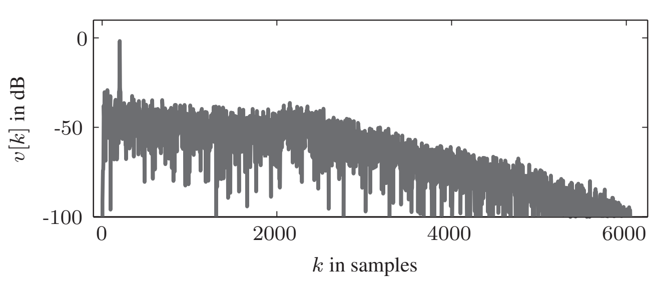

Fig. 2. Impulse response of an equalized system $v[k] = h[k] * c_{\mathrm{EQ}}[k]$ in dB (f_s=8kHz).

characterizing impulse responses. The ratio between the first 50ms or the first 80ms after the main peak to the overall energy of the RIR is called *Definition* and is denoted by D50 or D80, respectively [18]. The so-called *Clarity* [18], denoted here by C50 or C80, is the logarithmic ratio of 50ms (80ms) after the main peak to the rest of the impulse response. The *Direct-to-Reverberation-Ratio* DRR [19] is defined as the logarithmic ratio between the main peak and all others. The so-called *Central Time* CT [18] is no direct ratio but the center of gravity in terms of the energy of the RIR.

Since equalization often aims at a flat spectrum using the *variance* (VAR) of the logarithmic overall transfer function $V[n] = H[n]C_{\mathrm{EQ}}[n]$ was proposed in [2] to evaluate LRC algorithms. A second measure that judges a flat overall transfer function is the so-called *Spectral Flatness Measure* (SFM) [20] that calculates the ratio of geometric mean and the arithmetic mean of $V[n]$.

3.2. Measures Based on Signals

Whenever impulse responses or transfer functions are not obtainable for objective testing, e.g. for blind dereverberation [5], algorithms have to be evaluated based on the signals only. The most simple measures are the *Segmental Signal-to-Reverberation Ratio* (SSRR) [3] and the *SRR Enhancement* (SRRE) [12] that are defined similarly to SNR-based measures known from noise reduction quality assessment. The *Frequency-Weighted SSRR* (FWSSRR) [21] represents a first step towards consideration of the human auditory system by analyzing the SSRR in critical bands. Apart from the SRRE, all improvements between processed and unprocessed signals are indicated by a Δ, e.g. as for ΔFWSSRR. To account for logarithmic loudness perception of the human auditory system the *Log-Spectral Distortion* (LSD) compares logarithmical weighted spectra. Since dereverberation of speech is the aim in most scenarios, we also tested measures based on the LPC models as the *Log-Area Ratio* (LAR) [22], the *Log-Likelihood Ratio* (LLR) [21], the Itakura-Saito Distance (IS) [21], and the *Cepstral Distance* (CD) [21]. As a further extension towards modeling of the human auditory system the *Bark Spectral Distortion* measure (BSD) [23] compares perceived loudness incorporating spectral masking effects.
Recently, objective measures have been proposed especially designed for assessment of dereverberation algorithms. For this contribution we tested the so-called *Reverberation Decay Tail* (RDT) measure [11], the *Speech-to-Reverberation Modulation Energy Ratio* (SRMR) [24] and the *Objective Measure for Coloration in Reverberation* (OMCR) [10].
From quality assessment in the field of audio coding or noise reduction it is known that measures that are based on more exact models of the human auditory system show high correlation with subjective data. Thus, we also tested the *Perceptual Evaluation of Speech Quality* (PESQ) measure [25, 21] and the *Perceptual Similarity Measure* (PSM, PSMt) from PEMO-Q [8] that compares internal representations according to the auditory model of [9].

4. SUBJECTIVE QUALITY ASSESSMENT

Speech samples for the subjective tests were generated by convolving male and female utterances with room impulse responses generated by the image method [26] for a room having a size of $(6 \times 4 \times 2.6)\mathrm{m}^3$. The distance between sound source and microphone was approximately 0.8m. These RIRs had room reverberation times of approximately $\tau_{60} = \{500, 1000\}$ms corresponding to normal and somewhat larger office environments. To generate the dereverberated speech samples that later were presented to the subjects the reverberant speech samples were convolved with the equalization filters $c_{\mathrm{EQ}}[k]$ stemming from the different algorithms described in Section 2. Filter lengths of these equalizers were $L_{\mathrm{EQ}} = \{1024, 2048, 4096, 8196\}$ at a sampling rate of 8000Hz.

From all speech samples 19 audio samples were chosen which represented the full scale of reverberation and possible distortions.

These audio samples had a length of 8 sec and were scaled to have the same root-mean-square value. An audiovisual presentation of the samples and the corresponding systems can be found at [27]. They were presented diotically to 24 normal-hearing listeners via headphones (Sennheiser HD650) in quiet after an appropriate training period by example audio samples. Training and listening could be repeated as often as desired. A graphical user interface was programmed for the listening test based on the suggestions of [28] (with slight differences) asking to judge the attributes *reverberant, colored (distorted), distant and overall quality* on a continuous 5-point *Mean Opinion Score* (MOS) scale. It was expected that attributes "reverberant" and "distant" would lead to the same result. Since for LRC algorithms frequency distortion is perceptually much more prominent than what usually is understood as coloration, we asked to judge coloration/distortion as one spectral attribute. This leads to the fact that common measures that were designed to judge coloration may not correlate well to the subjective data. However, these distortions dominate the spectral perception of subjective quality.

5. RESULTS

Table 1 shows the correlations of subjective data with all measures described in Section 3 using the methodology described in Section 4. It is obvious that measures that are based on the impulse response of the equalized system show high correlation with the subjective data for all four attributes (with the exception of the DRR measure and somewhat lower correlations obtained for the attribute colored/distorted). The frequency-domain measures VAR and SFM are less correlated to all attributes. The signal-based measures show lower correlation to subjective data than the system-based measures. Furthermore, the Δ-measures show lowest correlation. Comparing the signal-based measures it is obvious that LPC-based measures outperform purely signal-based measures like the SSRR. By far, the highest correlations are obtained by measures relying on auditory models like PESQ, PSM and PSMt. PSMt, in addition to PSM, evaluates short-time behavior of the correlations of internal signal representations and focuses on low correlations as it is done by human listeners [8]. The auditory-model based measures show even higher correlation than RDT, SRMR and OMCR although the latter were designed to explicitly judge reverberation. The RDT and OMCR measures rely on internal parameters that can be adjusted and, by this, higher correlation to the specific set of samples can be obtained. However, we used standard values for these parameters given in [11, 10]. Furthermore, it has to be emphasized that the attribute coloration/distortion is most difficult to assess by objective measures at least for the discussed LRC algorithms, since distortions are perceptually relevant and measures like OMCR try to judge coloration effects only (the same holds for the variance measure). They succeed in doing so, but coloration alone is not well correlated to our subjective data. One reason for that surely is that the source-receiver distance for our experiment is larger than the critical distance. Fig. 3 gives an example for this. Although the spectral characteristics are clearly enhanced (sub-figure (b)) and in time domain (sub-figure (a)) much energy of the impulse response is suppressed, a high amount of late reverberation occurs after sample 5000 that is small in amplitude but perceptually relevant since the temporal masking effect of the main peak is less distinct for those late taps [16]. Fig. 3 (c) which depicts the subjective rating for the given system in Fig. 3 (a) and (b), clearly shows that a high amount of reverberation is perceived by the subjects as well as relatively high spectral coloration/distortions given that the transfer function is clearly enhanced compared to the unprocessed room transfer function. Furthermore, depending on the delay that is introduced by the equalizer, perceptually disturbing pre-echoes occur as observable in Fig. 3 (a). None of the tested measure

		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

Table 1. Correlations between subjective quality assessment and objective measures. Length of bars in each cell correspond to correlation coefficients r. High correlations between $0.8 \leq r \leq 1$ are indicated by green bars, medium correlations between $0.6 \leq r < 0.8$ by yellow bars and low correlations $r < 0.6$ by red bars.

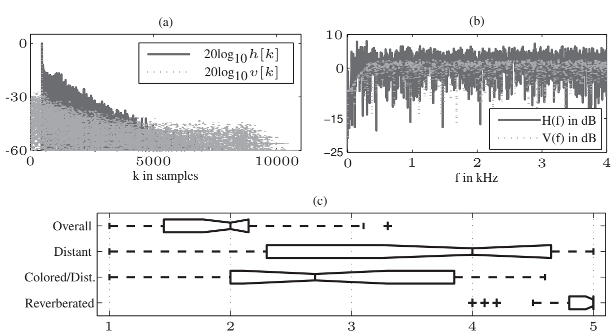

Fig. 3. (a) IR or room ($\tau_{60} = 1$s) and equalized system in time domain (EQ filter length was 8192 samples, delay has been compensated to match main peaks of IRs), (b) corresponding transfer functions, (c) subjective rating for system in (a), (b).

is capable to explicitly judge those influences. Thus, the authors believe that developing such a measure would be valuable future work.

It should be noted that even those measures showing low correlation to the subjective data can be used for assessment within a specific test setup (e.g., for the enhancement of one specific algorithm). For example, the SSRR measure that is used widely to judge dereverberation algorithms is capable of clearly indicating a decrease of reverberation energy. However, a thoughless use of this measure, e.g., to compare different algorithms, may be imprudent.

As it was expected, the attributes reverberation and distance show similar results (correlation between reverberation and distance was 94%). Furthermore, for the given set of speech data, the correlation between the attributes *overall quality* and the attributes *distant* as well as *reverberated* show high correlation as summarized in Table 2. Thus, the perceived audio quality is strongly influenced by reverberation (including late reverberation).

	Colored/distorted	Distant	Overall
Reverberated	0.56	0.94	0.95
Colored/distorted		0.65	0.70
Distant			0.97

Table 2. Inter-attribute correlations.

6. CONCLUSION AND OUTLOOK

In this contribution it has been analyzed to what extent various objective quality measures are capable to assess quality of speech samples that have been dereverberated by means of listening-room compensation algorithms. Measures that are based on the impulse response (like the common C50 measure) showed much higher correlation between objective and subjective data than most of the tested measures that are based on the signals only. However, if impulse responses are not properly accessible, e.g. as for blind dereverberation algorithms, measures that incorporate sophisticated auditory models should be used for quality assessment. The so-called Perceptual Similarity Measure showed highest correlations to subjective data.

7. LIST OF OBJECTIVE MEASURES

BSD: *Bark Spectral Distortion* [23], **C50, C80:** *Clarity* [18], **CD:** *Cepstral Distance* [21], **CT:** *Central Time* [18], **D50,D80:** *Definition* [18], **DRR:** *Direct-to-Reverberation-Ratio* [19], **FWSSRR:** *Frequency-Weighted SSRR* [21], **IS:** *Itakura-Saito Distance* [21], **LAR:** *Log-Area Ratio* [22], **LLR:** *Log-Likelihood Ratio* [21], **LSD:** *Log-Spectral Distortion* [21], **OMCR:** *Objective Measure for Coloration in Reverberation* [10], **PESQ:** *Perceptual Evaluation of Speech Quality* [21], **PSM, PSMt:** *Perceptual Similarity Measure* [8] **RDT:** *Reverberation Decay Tail* [11], **SFM:** *Spectral Flatness Measure* [20], **SRMR:** *Speech-to-Reverberation Modulation Energy Ratio* [24], **SSRR:** *Segmental Signal-to-Reverberation Ratio* [3], **VAR:** *Variance of logarithmic transfer function* [2].

8. REFERENCES

[1] 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.

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

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

[4] J. Wen, N. Gaubitch, E. Habets, T. Myatt, and P. 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.

[5] E. Habets, "Single and Multi-Microphone Speech Dereverberation using Spectral Enhancement," Ph.D. dissertation, University of Eindhoven, Eindhoven, The Netherlands, June 2007.

[6] E. Albertin, J. Rennies and S. Goetze, "Objective Quality Measures for Dereverberation Methods based on Room Impulse Response Equalization," in *Proc. German Annual Conference on Acoustics (DAGA)*, Berlin, Germany, Mar. 2010.

[7] 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, June 1992.

[8] R. Huber and B. Kollmeier, "PEMO-Q - A New Method for Objective Audio Quality Assessment using a Model of Auditory Perception," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14, no. 6, Special Issue on Objectiv Quality Assessment of Speech and Audio, 2006.

[9] T. Dau, D. Püschel, and A. Kohlrausch, "A Quantitative Model of the Effective Signal Processing in the Auditory System: I. Model Structure," *Journal of the Acoustical Society of America (JASA)*, vol. 99, no. 6, June 1996.

[10] J. Wen and P. Naylor, "Objective Measurement of Colouration in Reverberation," in *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Poznan, Poland, Sept. 2007.

[11] J. Wen and P. Naylor, "An Evaluation Measure for Reverberant Speech using Decay Tail Modeling," in *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sept. 2006.

[12] 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, May 2008.

[13] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Estimation of the Optimum System Delay for Speech Dereverberation by Inverse Filtering," in *Int. Conf. on Acoustics (NAG/DAGA 2009)*, Rotterdam, The Netherlands, Mar. 2009.

[14] M. Kallinger and A. Mertins, "A Spatially Robust Least Squares Crosstalk Canceller," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, 15-20 April 2007.

[15] M. Kallinger and A. Mertins, "Room Impulse Response Shaping – A Study," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2006.

[16] A. Mertins, T. Mei and M. Kallinger, "Room Impulse Response Shortening/Reshaping with Infinity- and p-Norm Optimization," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 249–259, Feb. 2010.

[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.

[18] H. Kuttruff, *Room Acoustics*, 4th ed. Spoon Press, London, 2000.

[19] M. Triki and D. 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.

[20] J. D. Johnston, "Transform Coding of Audio Signals using Perceptual Noise Criteria," *IEEE Journal on Selected Areas in Communication*, vol. 6, no. 2, Feb. 1988.

[21] P. Loizou, *Speech Enhancement: Theory and Practice*. CRC Press Inc., Boca Raton, USA, 2007.

[22] J. Hansen and B. Pellom, "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms," in *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, vol. 7, Sydney, Australia, Dec. 1998.

[23] W. Yang, "Enhanced Modified Bark Sectral Distortion (EMBSD): A Objective Speech Quality Measure Based on Audible Distortion and Cognition Model," Ph.D. thesis, Temple University, Philadelphia, USA, May 1999.

[24] T. Falk and W.-Y. Chan, "A Non-Intrusive Quality Measure of Dereverberated Speech," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, Sept. 2008.

[25] ITU-T P.862, "Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, ITU-T Recommendation P.862," Feb. 2001.

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

[27] Sound samples available online at http://www.ant.uni-bremen.de/∼goetze/icassp2010/index.html.

[28] ITU-T P.835, "Subjective Test Methology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm, ITU-T Recommendation P.835," Nov. 2003.