Speech Quality Assessment for Listening-Room Compensation

Stefan Goetze[1], Eugen Albertin[1], Jan Rennies[1], Emanuël A.P. Habets[2], and Karl-Dirk Kammeyer[3]

[1]*Fraunhofer Institue for Digital Media Technology (IDMT), Hearing, Speech and Audio Technology, 26129 Oldenburg, Germany*

[2]*Imperial College London, Dept. of Electrical and Electronic Engineering, London SW7 2AZ, UK*

[3]*University of Bremen, Dept. of Communications Engineering, 28359 Bremen, Germany*

Correspondence should be addressed to Stefan Goetze (s.goetze@idmt.fraunhofer.de)

ABSTRACT

In this contribution objective measures for quality assessment of speech signals are evaluated for listening-room compensation algorithms. Dereverberation of speech signals by means of equalization of the room impulse response and reverberation suppression has been an active research topic within the last years. However, no commonly accepted objective quality measures exist for assessment of the enhancement achieved by those algorithms. This paper discusses several objective quality measures and their applicability for dereverberation of speech signals focusing on algorithms for listening-room compensation.

1. INTRODUCTION

State-of-the-art hands-free communication devices as they are used e.g. in offices or car environments use algorithms to reduce ambient noise, acoustic echoes and reverberation. Reverberation is caused by numerous reflections of the signal on room boundaries (walls, floor and ceiling) in enclosed spaces. Reverberant speech sounds distant and echoic [1]. Large amounts of reverberation decrease speech intelligibility and perceived quality at the position of the near-end speaker of a communication system [2–4]. In general, two distinct reverberation reduction classes exist, viz. reverberation suppression and reverberation cancellation. Reverberation suppression approaches focus on removing the reverberant part of the speech signal by calculating a spectral weighting rule for each time-frequency coefficient in a way similar to well-known approaches for noise reduction (cf. e.g. [5] and the references therein). Reverberation cancellation approaches remove the influence of the acoustic channel between the sound source and the listener by equalizing the room impulse response (RIR) of the channel. This knowledge can be obtained by means of blind [6] or non-blind [7] system identification. The equalizer can be applied to the loudspeaker signal or the microphone signal. Listening-room compensation is achieved in the former case, i.e. when the equalizer is applied to the signal that is emitted by the loudspeaker such that the influence of reverberation on the perceived signal is reduced at the position the listener is assumed to be located. In order to compute the equalizer one requires knowledge of the RIR, which in the context of listening-room compensation (LRC) is often obtained using non-blind system identification [7–9]. This contribution focuses on such non-blind approaches for LRC. While the aim of LRC algorithms is to improve the sound quality of the dereverberated signal, they may also decrease the sound quality if they are not designed properly [7, 10]. Thus, especially during algorithm design periods a reliable objective quality measure is required to evaluate and compare different algorithms and their parameters. It should be noted that at least the signal-based objective quality measures that will be described in Section 4.2 are applicable for quality assessment of all kinds of dereverberation algorithms.

In general, whenever signal processing strategies change a signal, e.g. to enhance speech quality, speech intelligibility, listening effort, etc., the question arises how to assess the achieved enhancement. Since subjective listening tests that involve humans are not applicable in every case because they are time consuming and costly, objective quality measures that assess the performance of the dereverberation algorithm based on impulse responses, transfer functions or signals are needed [11]. While several commonly accepted quality measures exist to assess

the performance of noise reduction algorithms or acoustic echo cancellers, the assessment of dereverberation algorithms is still an open issue [1, 10, 12].

This work discusses several measures that can be used for evaluating dereverberation algorithms. An evaluation of the sound quality of the dereverberated signals was conducted by subjective listening tests and compared to the results of the objective measures. As previously shown by the authors [10], most signal-based measures have difficulties to assess the performance of dereverberation algorithms properly, especially if distortions are introduced that are small in amplitude but clearly perceivable by the human listener. However, especially these measures are of particular interest since, e.g. for nonlinear dereverberation suppression approaches, channel-based measures may not be applicable since the impulse response of such an algorithm may be neither linear nor time-invariant. Thus, artifacts that may be introduced by the dereverberation algorithms such as late echoes or spectral distortions, and their effect on the quality measures are analyzed and discussed. The algorithms were analyzed regarding their capability to assess the properties *reverberation*, *coloration*, *spectral distortion*, perceived *distance*, and *overall quality* of the signals. Since objective measures that rely on purely technical energy ratios, as it is common for quality assessment for noise reduction algorithms, do not show good correlation to subjective tests [13], we will particularly focus on quality measures that incorporate knowledge about the human auditory system for quality assessment.

The remainder of this paper is organized as follows. Methods for LRC that were used for generating the test signals are briefly summarized in Section 2 and some general remarks on quality assessment for LRC algorithms are given in Section 3. Section 4 gives an overview of objective quality measures that principally can be used for quality assessment of LRC algorithms and Section 5 describes the experimental setup for the subjective listening tests. Results of the correlation analysis are presented in Section 6 and Section 7 concludes the paper.

Notation: The following notation is used throughout the paper. Vectors and matrices are printed in boldface while scalars are printed in italic. The discrete time and frequency indices are denoted by n and k, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^+$ denote the transposition and the Moore-Penrose pseudoinverse, respectively. The operator $*$ denotes the convolution of two sequences, $\mathrm{E}\{\cdot\}$ is the expectation operator, and the operator $\mathrm{convmtx}\{\mathbf{h}, L_{\mathrm{EQ}}\}$ generates a convolution matrix of size $(L_{\mathrm{EQ}} + L_h - 1) \times L_{\mathrm{EQ}}$. The operator $\mathrm{diag}\{\cdot\}$ yields a matrix of size $L \times L$ from a vector of size $L \times 1$ that has the vector's elements on its main diagonal and zeros elsewhere.

2. LISTENING-ROOM COMPENSATION

For LRC the equalization filter

$$\mathbf{c}_{\mathrm{EQ}} = \left[c_{\mathrm{EQ},0}, c_{\mathrm{EQ},1}, \ldots, c_{\mathrm{EQ},L_{\mathrm{EQ}}-1}\right]^T \tag{1}$$

of length L_{EQ} precedes the acoustic channel characterized by the RIR

$$\mathbf{h} = \left[h_0, h_1, \ldots, h_{L_h}\right]^T \tag{2}$$

of length L_h. The aim of the equalizer is to remove the influence of the RIR at the position of the reference microphone [8, 14] and, by this, to remove reverberation from the signal. A general setup for listening-room compensation is shown in Fig. 1. Four different LRC approaches were used to generate sound samples with the goal of covering a large amount of distortions that may occur while using such algorithms. These four approaches are briefly introduced in the following. For a deeper discussion of the LRC algorithms please refer to the respective references.

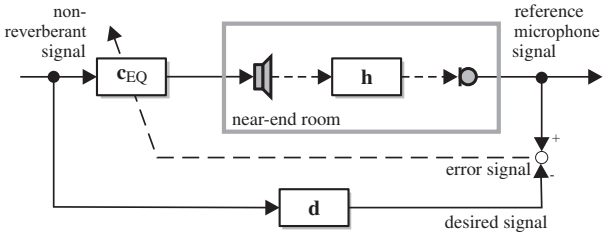

Fig. 1: General setup for listening-room compensation (LRC) using an equalizer filter $\mathbf{c}_{\mathrm{EQ}}$.

Since an RIR is a mixed-phase system having thousands of zeros close to or even outside the unit-circle in z-domain, a direct inversion by a causal stable filter is not possible in general [15]. Therefore, least-squares approaches focus on minimizing the error vector

$$\mathbf{e}_{\mathrm{EQ}} = \mathbf{H}\mathbf{c}_{\mathrm{EQ}} - \mathbf{d}, \tag{3}$$

where $\mathbf{H} = \text{convmtx}\{\mathbf{h}, L_{\text{EQ}}\}$ is the channel convolution matrix built up by the RIR coefficients and

$$\mathbf{d} = [\underbrace{0,\ldots,0}_{n_0}, d_0, d_1, \ldots, d_{L_d-1}, \underbrace{0,\ldots,0}_{L_h+L_{\text{EQ}}-1-L_d-n_0}]^T \quad (4)$$

is the desired response of length $L_h + L_{\text{EQ}} - 1$ with n_0 being the delay introduced by the equalizer (cf. [16] for a discussion of n_0). Rather than minimizing the norm of the error vector, one can minimize the norm of a weighted error vector. By a proper choice of a weighting vector, RIR shortening or RIR shaping can be achieved. Preferably, the weighting is based on the psychoacoustic property of masking observed in the human auditory system in order to alleviate perceptually disturbing late echoes [17, 18]. Hence, minimizing the ℓ_2-norm of the weighted error vector $\mathbf{We}_{\text{EQ}}$ leads to a weighted least-squares equalizer

$$\mathbf{c}_{\text{EQ}} = (\mathbf{WH})^{+}\mathbf{Wd} \quad (5)$$

with

$$\mathbf{W} = \text{diag}\{\mathbf{w}\} \quad (6)$$

$$\mathbf{w} = [\underbrace{1,1,...,1}_{N_1}, \underbrace{w_0, w_1, ..., w_{N_2-1}}_{N_2}]^T \quad (7)$$

$$w_i = 10^{\frac{3\alpha}{\log_{10}(N_0/N_1)}\log_{10}(i/N_1)+0.5}. \quad (8)$$

Here, $\mathbf{W}$ is a diagonal matrix containing a window weighting vector $\mathbf{w}$ on its main diagonal adopted from [18] with $N_0 = (t_0 + 0.2)f_s$, $N_1 = (t_0 + 0.004)f_s$ and $N_2 = L_h + L_{\text{EQ}} - 1 - N_1$. The time of the direct sound is denoted by t_0 and $\alpha \leq 1$ is a factor that influences the steepness of the window. For $\alpha = 1$ the window corresponds to the masking found in human subjects [17, 18].

Another approach for RIR shaping was discussed in [19] and is based on the solution of a generalized eigenvalue problem

$$\mathbf{Ac}'_{\text{EQ}} = \lambda_{\max}\mathbf{Bc}'_{\text{EQ}} \quad (9)$$

$$\mathbf{A} = \mathbf{H}^T\mathbf{W}_{\text{u}}^T\mathbf{W}_{\text{u}}\mathbf{H} \quad (10)$$

$$\mathbf{B} = \mathbf{H}^T\mathbf{W}_{\text{d}}^T\mathbf{W}_{\text{d}}\mathbf{H}. \quad (11)$$

Similar to (5), $\mathbf{W}_{\text{u}}$ and $\mathbf{W}_{\text{d}}$ 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 (9). To avoid spectral distortion a post-processor based on linear prediction [19] is used after applying (9). For a more detailed discussion the reader is referred to [18, 19].

An approach that jointly shapes the impulse response (IR) of the equalized system and minimizes spectral distortions is described in [18]. Additionally, the psychoacoustic property of masking is exploited during the filter design in [18].

Table 1 summarizes the four approaches and the respective acronyms used for LRC and for generating dereverberated signals that were used for the subjective tests described in Section 5.

	Acronym	Description of method
1.	LS-EQ	Least-squares equalizer according to (5) without weighting of error signal ($\mathbf{w} = \mathbf{1}$)
2.	WLS-EQ	Least-squares equalizer according to (5) with window function according to (7)
3.	ISwPP	Impulse response shaping (IS) according to (9) with post-processing (PP) [19]
4.	ISwINO	Impulse response shaping (IS) with infinity-norm optimization (INO) according to [18]

Table 1: LRC approaches.

An RIR $\mathbf{h}$ that is processed by the equalizers described above is exemplarily shown in Fig. 2 in time-domain (left panels) and frequency-domain (right panels). The room reverberation time of the RIR is $\tau_{60} = 0.5$ s and the respective filter length of the equalizer is $L_{\text{EQ}} = 4096$ at a sampling rate of $f_s = 8$ kHz. The LS-EQ approach seems to show good results in the time-domain (a) as well as in the frequency-domain (b). However, although the desired system $\mathbf{d}$ which was chosen as a delayed high-pass is closely approximated a large amount of late reverberation can be observed e.g. around sample $n = 4000$. Although small in amplitude this late reverberation is clearly perceivable and disturbing since it is no longer masked by the natural decay of common RIRs [17, 18]. Furthermore, pre-echoes that occur before the main peak of the equalized channel's impulse response further disturb a natural sound perception. By applying the window as defined in (7) both pre-echoes and late echoes are reduced at the cost of a less flat transfer function in the frequency-domain (cf. sub-figures (c) and (d)). Figures 2 (e) and (f) show the impulse response shaping approach according to [19] and sub-figures (g) and (h) show the approach according to [18] that explicitly focuses on hiding all echoes below the masking curve.

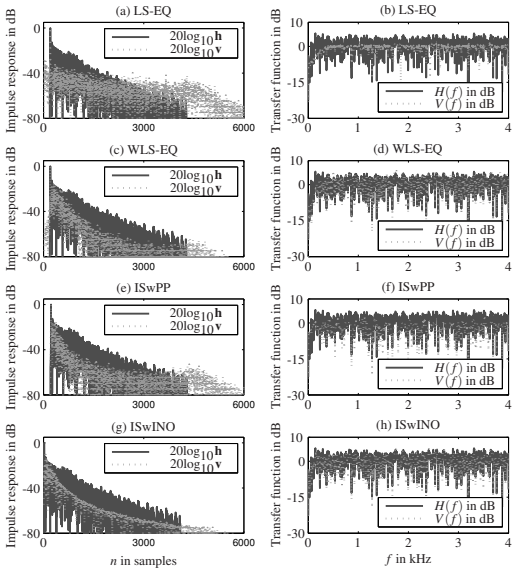

Fig. 2: Performance of the LRC algorithms. RIR $\mathbf{h}$ and equalized IR $\mathbf{v} = \mathbf{H}\mathbf{c}_{\mathrm{EQ}}$ are shown in time-domain in dB in sub-figures (a), (c), (e), (g) and the corresponding squared-magnitude spectra in dB in sub-figures (b), (d), (f), (h).

3. QUALITY ASSESSMENT FOR LRC ALGORITHMS

Within this contribution, quality assessment involving human *subjects* is called *subjective* quality assessment while quality assessment based on technical measures is denoted by the term *objective*. If humans are asked for their opinion about the quality of a specific sound sample they are able to assess the quality based on an internal reference. This reference is created throughout their life while listening to various sounds and allows the subject to distinguish between *good quality* and *bad quality*. Unfortunately, subjective quality assessment is time consuming and costly. Thus, especially during algorithm design and test periods reliable objective quality measures are needed that show high correlation with subjective ratings. Since no commonly accepted measure for LRC quality assessment has been identified yet, we analyzed the correlation between subjective quality ratings and various objective measures that are assumed to be applicable for LRC quality assessment as depicted in Fig. 3.

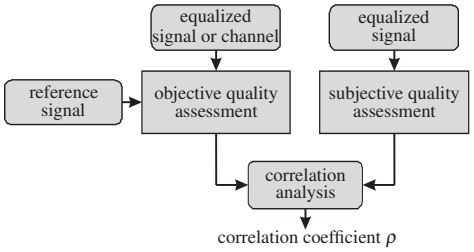

Fig. 3: Quality assessment by means of subjective and objective testing.

Here, the reverberant signal is processed by the LRC algorithm under test that produces a processed signal and a corresponding equalized impulse response. This signal is assessed by human subjects. The objective measures described in Section 4 either take the equalized impulse response (channel-based measures) or the processed signal (signal-based measures) as an input. The correlation between the subjective and objective ratings was determined by

$$\rho = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2 \sum_i (b_i - \bar{b})^2}}, \qquad (12)$$

where a_i and b_i are the subjective and objective ratings on a specific sound sample and $\bar{a}$ and $\bar{b}$ the respective mean values.

It should be noted that besides the *Speech-to-Reverberation Modulation Energy Ratio* measure all objective measures used in this contribution belong to the class of *intrusive* measures, which means that they explicitly need a reference signal or system while human subjects rely on their internal reference.

4. OBJECTIVE QUALITY ASSESSMENT

This section focuses on the description of several objective quality measures that are assumed to be capable to assess quality of signals processed by LRC algorithms. Two classes of objective quality measures for LRC can be defined: (i) measures that are based on the impulse response or the transfer function of a system (channel-based measures) and (ii) measures that are based on signals only. For LRC algorithms, both the filter impulse response $\mathbf{c}_{\mathrm{EQ}}$ and the RIR $\mathbf{h}$ are available during simulations. However, if gradient algorithms [14] are used to avoid computational complex matrix inversions, e.g. as

in (5), or to track time-varying environments or if the effect of the dereverberation algorithm cannot be characterized in terms of an linear time invariant (LTI) impulse response, e.g. as in [5, 20, 21], the necessary impulse responses of the room or the filter may not be accessible or it may be inappropriate to apply those measures [22]. Such situations restrict the number of applicable measures to those based on signals as described in Section 4.2.

4.1. Channel-Based Measures

Objective measures to characterize room impulse responses are mostly based on the ratio of early and late part of the RIR, see e.g. [23]. Since the IR of an equalized system **v** may look slightly different compared to a normal RIR some objective measures were adapted from their original definitions to account for this. Fig. 4 shows such an equalized system and illustrates the definitions of the lags n_0, which is the position of the main peak of the impulse response, $N_{50} = \lfloor 0.05\,\mathrm{s}\cdot f_s \rfloor$ and $N_{80} = \lfloor 0.08\,\mathrm{s}\cdot f_s \rfloor$ which are the samples 50 ms and 80 ms later than the main peak, respectively. The definitions of six measures that are widely used to characterize RIRs are given in the following for the equalized system **v** and are also applicable for an RIR **h**. The ratio between the first 50 or 80 ms after the main

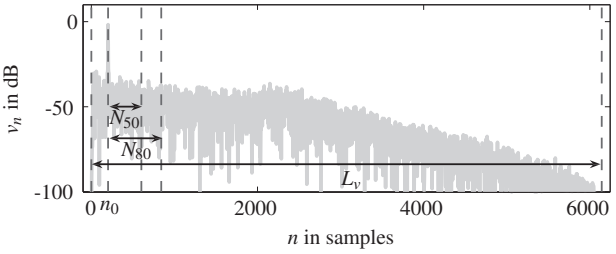

Fig. 4: Impulse response of an equalized system $\mathbf{v} = \mathbf{H}\mathbf{c}_{\mathrm{EQ}}$ in dB and the corresponding definitions of the position of the main peak n_0, and the discrete samples following 50 ms and 80 ms after this main peak N_{50} and N_{80}. Sampling frequency is $f_s = 8$ kHz.

peak to the overall energy of the RIR is called *Definition* and is denoted by D_{50} or D_{80}, respectively [23]: $\mathrm{D}_{\{50,80\}} = \frac{\sum_{n=n_0}^{n_0+N_{\{50,80\}}-1} v_n^2}{\sum_{n=0}^{L_v-1} v_n^2}$. The *Clarity* [23], denoted here by C_{50} or C_{80}, is the logarithmic ratio of 50 or 80 ms after the main peak to the rest of the impulse response: $\mathrm{C}_{\{50,80\}} = 10\log_{10}\frac{\sum_{n=n_0}^{n_0+N_{\{50,80\}}-1} v_n^2}{\sum_{n=0}^{n_0-1} v_n^2 + \sum_{n=n_0+N_{\{50,80\}}}^{L_v-1} v_n^2}$. The *Direct-to-Reverberation-Ratio* DRR [24] is defined as the logarithmic ratio between the energy of the direct path of the impulse response and the energy of all reflections. However, since the direct path, in general, does not match the sampling grid, a small range around the main peak is considered as the direct path energy [5, 24]: $\mathrm{DRR} = 10\log_{10}\frac{\sum_{n=n_0-n_\Delta}^{n_0+n_\Delta-1} v_n^2}{\sum_{n=0}^{n_0-n_\Delta-1} v_n^2 + \sum_{n=n_0+n_\Delta}^{L_v} v_n^2}$. Here, we chose $n_\Delta = 4\,\mathrm{ms}\cdot f_s$. The *Central Time* CT [23] is no direct ratio but the center of gravity in terms of the energy of the RIR: $\mathrm{CT} = \frac{\sum_{n=0}^{L_v} n\cdot v_n^2}{\sum_{n=0}^{L_v} v_n^2}$. Additionally to the time-domain measures described above, we evaluated two common spectral channel-based measures to account for the coloration effect [2,12]. Since equalization often aims at a flat spectrum using the *variance* (VAR) of the logarithmic overall transfer function $V_k = H_k C_{\mathrm{EQ},k}$ as an objective measure to evaluate LRC algorithms was proposed in [9,25]: $\mathrm{VAR} = \frac{1}{K_{\max}-K_{\min}+1}\sum_{k=K_{\min}}^{K_{\max}} \left(20\log_{10}|V_k| - \bar{V}_{\mathrm{dB}}\right)^2$. Here, $\bar{V}_{\mathrm{dB}} = \frac{1}{K_{\max}-K_{\min}+1}\sum_{k=K_{\min}}^{K_{\max}} 20\log_{10}|V_k|$ is the mean logarithmic spectrum and $K_{\min}$ and $K_{\max}$ are the frequency indices that limit the considered frequency range in which the equalized transfer function is desired to be flat. We chose $K_{\min}$ and $K_{\max}$ corresponding to 200 Hz and 3700 Hz to account for a high-pass or band-pass characteristic of the desired system vector in (4). A second measure for the quality of equalization in frequency-domain is the *spectral flatness measure* (SFM) that is the ratio of geometric mean and the arithmetic mean of V_k [26]: $\mathrm{SFM} = \frac{\sqrt[K]{\prod_{k=0}^{K-1}|V_k|^2}}{\frac{1}{K}\sum_{k=0}^{K-1}|V_k|^2}$, where K denotes the number of frequency bins.

4.2. Signal-Based Measures

For non-linear dereverberation suppression approaches as in [5], impulse responses or transfer functions are not obtainable or applicable for objective testing. Thus, such algorithms have to be evaluated based on the signals only. Several signal-based measures that exist for assessment of LRC approaches and dereverberation suppression approaches are briefly summarized in the following. Due to space limitation the interested reader is referred to the respective references. Simple measures like the *Segmental Signal-to-Reverberation Ratio* (SSRR) [1] are defined similarly to SNR-based measures known from noise reduction quality assessment. As al-

ready known from speech quality assessment for noise reduction, quality measures incorporating models of the human auditory system show higher correlation with subjective rating [13]. The *Frequency-Weighted SSRR* (FWSSRR) [27] and the *Weighted Spectral Slope* (WSS) [27] represent a first step towards consideration of the human auditory system by analyzing the SSRR in critical bands. To account for logarithmic loudness perception within the human auditory system the *Log-Spectral Distortion* (LSD) compares logarithmically weighted spectra. Since dereverberation of speech is the aim in most scenarios, we also tested measures based on the LPC models such as the *Log-Area Ratio* (LAR) [28], the *Log-Likelihood Ratio* (LLR) [27], the Itakura-Saito Distance (ISD) [27], and the *Cepstral Distance* (CD) [27]. As a further extension towards modeling of the human auditory system the *Bark Spectral Distortion* measure (BSD) [29] 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 *Reverberation Decay Tail* (RDT) measure [30], the *Speech-to-Reverberation Modulation Energy Ratio* (SRMR) [31] and the *Objective Measure for Coloration in Reverberation* (OMCR) [32].

From quality assessment in the fields of audio coding and 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 [13]. Thus, we also tested the *Perceptual Evaluation of Speech Quality* (PESQ) measure [27, 33] and the *Perceptual Similarity Measure* (PSM, PSM$_t$) from PEMO-Q [34] that compares internal representations according to the auditory model of [35].

5. SUBJECTIVE QUALITY ASSESSMENT

For the subjective listening tests, reverberant speech samples were calculated by first convolving room impulse responses generated by the image method [36] for a room having a size of 6 m $\times$ 4 m $\times$ 2.6 m (length $\times$ width $\times$ height) with male and female utterances. The distance between sound source and microphone was approximately 0.8 m. Room reverberation times were approximately $\tau_{60} = \{500, 1000\}$ ms corresponding to normal and somewhat larger office environments. These reverberant speech samples were then processed by the four LRC approaches discussed in Section 2 and presented to the subjects. Filter lengths of these equalizers were $L_{\mathrm{EQ}} = \{1024, 2048, 4096, 8196\}$ at a sampling rate of 8 kHz. The parameters α in (7) was chosen to $\alpha = 0.8$.

From all generated speech samples 21 audio samples were chosen which represented a wide variety of acoustic conditions and possible distortions. These audio samples had a length of 8 s and were scaled to have the same root-mean-square value. An audiovisual presentation of the samples and the corresponding systems can be found in [37]. They were presented diotically to 24 normal-hearing listeners via headphones (Sennheiser HD650) in quiet after a 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 [38] (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. An overview of the training and listening test as well as the GUI can also be obtained at [37]. For the algorithms under test, it was expected that attributes *reverberant* and *distant* would lead to similar results. 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.

6. RESULTS

6.1. Rating of the Sound Samples

The subjective ratings of the sound samples [37] for the four attributes *reverberant*, *colored/distorted*, *distant*, and *overall quality* are shown in Fig. 5 by means of boxplots.

The sound samples are ordered according to their median value for the respective attribute. Consequently, the order is different for the different sub-figures. In general, the results of the shaping approaches are better than the LS approaches. Increasing the filter length of the LS-EQ does not improve the results due to the fact that despite a 'good equalization' perceptually relevant late echoes and pre-echoes are clearly perceived as disturbing by the listeners. Good ratings are achieved by the WLS-EQ and

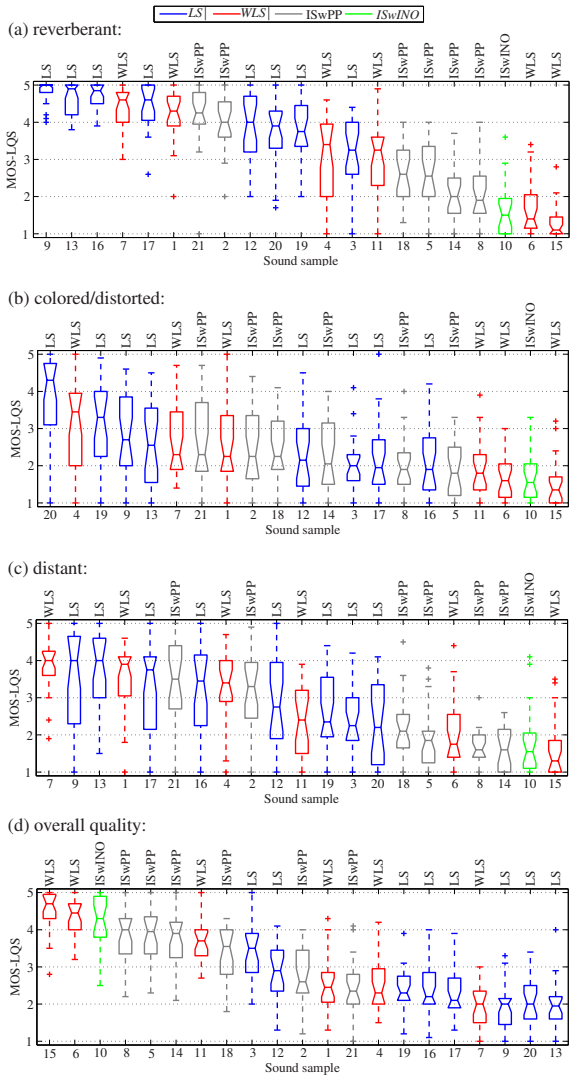

Fig. 5: Subjective rating of sound samples for attribute (a) *reverberant*, (b) *colored/distorted*, (c) *distant*, and (d) *overall quality*

the impulse response shaping based on infinity-norm optimization that considers human masking (see e.g. sound samples no. 15 and 10).

Table 2 shows the inter-attribute correlations for the given set of speech samples. As expected, the attributes *reverberant* and *distant* show high inter-attribute correlation (0.94) although the attribute *distant* leads to a higher interquartile range (IQR) as it can be seen comparing sub-figures (a) and (c) in Fig. 5. Furthermore, the correlation between the attributes *overall quality* and the attributes *distant* as well as *reverberant* is high. Thus, the perceived audio quality is strongly influenced by reverberation (including late reverberation).

Attribute	Colored/distorted	Distant	Overall
Reverberant	0.44	**0.91**	**0.94**
Colored/distorted	-	0.29	0.66
Distant	-	-	0.86

Table 2: Inter-attribute correlations.

6.2. Correlation analyses

The correlations of subjective rating for the four attributes and the channel-based objective measures are shown in Table 3 while correlations with signal-based objective measures are shown in Table 4. For each objective measure correlations with the subjective ratings are given for the case that all LRC approaches of Section 2 are considered (Method: All EQs) and for the case that only one LRC approach is used. For the latter case no correlation was calculated for the impulse-response shaping approach based on infinity-norm optimization because the number of sound samples was too low for a reliable correlation analysis. The highest correlation for each attribute and approach is highlighted in boldface in the tables. The reason for additionally calculating correlations for each LRC approach separately is exemplarily illustrated in Fig. 6 for the SFM.

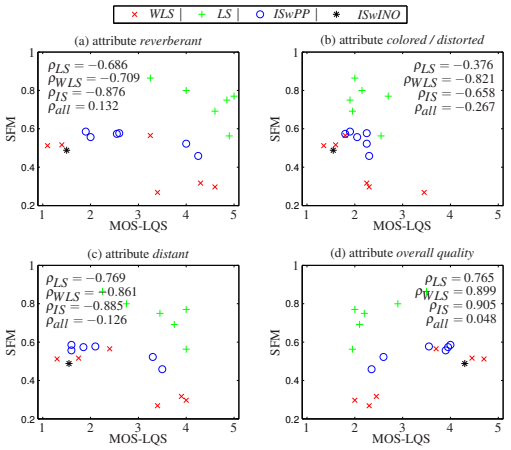

Fig. 6: Correlations of subjective ratings and SFM measure for all four attributes.

Here the SFM shows much higher correlation when a sin-

Measure	Method	Reverberant	Col./dist.	Distant	Overall
D_{50}	All EQs	0.860	0.629	0.937	0.910
	LS-EQ	0.711	0.329	0.795	0.794
	WLS-EQ	0.942	0.735	**0.993**	**0.982**
	ISwPP	0.943	0.611	0.940	0.934
D_{80}	All EQs	0.905	0.504	0.911	0.904
	LS-EQ	0.733	0.311	0.815	0.817
	WLS-EQ	0.941	0.585	0.976	0.931
	ISwPP	0.850	0.546	0.844	0.844
C_{80}	All EQs	**0.930**	0.607	0.888	0.907
	LS-EQ	0.804	0.305	0.865	0.877
	WLS-EQ	0.982	0.690	0.987	0.963
	ISwPP	0.916	0.543	0.899	0.882
C_{50}	All EQs	0.926	**0.665**	**0.944**	**0.935**
	LS-EQ	0.783	0.320	0.846	0.857
	WLS-EQ	**0.965**	0.755	0.981	0.971
	ISwPP	**0.976**	0.580	0.958	0.933
CT	All EQs	0.845	0.607	0.927	0.911
	LS-EQ	**0.909**	0.288	**0.938**	**0.949**
	WLS-EQ	0.857	0.785	0.958	0.966
	ISwPP	0.973	0.667	**0.979**	**0.974**
DRR	All EQs	0.238	0.101	0.179	0.131
	LS-EQ	0.769	0.335	0.835	0.843
	WLS-EQ	0.399	**0.858**	0.597	0.696
	ISwPP	0.249	**0.692**	0.273	0.360
VAR	All EQs	0.028	0.374	0.231	0.156
	LS-EQ	0.618	**0.416**	0.708	0.694
	WLS-EQ	0.687	0.809	0.841	0.883
	ISwPP	0.599	0.462	0.608	0.647
SFM	All EQs	0.132	0.267	0.126	0.048
	LS-EQ	0.686	0.376	0.769	0.765
	WLS-EQ	0.709	0.821	0.861	0.899
	ISwPP	0.876	0.658	0.885	0.905

Table 3: Correlations $|\rho|$ of MOS values of subjective ratings and channel-based objective measures (maxima are indicated in boldface).

gle rather than all LRC approaches are considered. However, the time-domain channel-based measures show consistent correlations for all LRC approaches. The interested reader is referred to [37] for an overview of all correlation patterns. It can be seen from Table 3 that the time-domain channel-based objective measures show high correlation with the subjective data for the attributes *reverberation, distance* and *overall quality* (with the exception of the DRR measure). The frequency-domain channel-based measures VAR and SFM show much lower correlation. However, as stated before, they may show somewhat higher correlation for single LRC approaches such as SFM for the WLS-EQ. In general, and this is also true for the signal-based measures (cf. Table 4), only low correlation was obtained with the attribute *colored/distorted* for all measures. This can be attributed to the fact that the source-receiver distance for our experiment (0.8 m) is larger than the critical distance.

As shown in Fig. 7 and in consilience with the findings

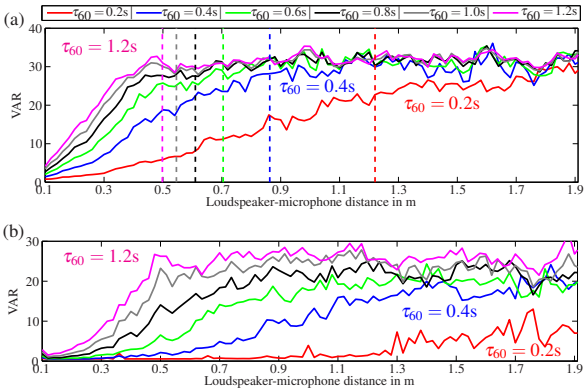

Fig. 7: VAR measure of (a) RIR H_k and (b) equalized channel V_k over loudspeaker-microphone distance for different room reverberation times (critical distances are indicated as dashed vertical lines). Sub-figure (b) shows the VAR measure for an equalized system using an LS-EQ with $L_{\mathrm{EQ}} = 2048$ at $f_s = 8$ kHz.

in [5, 25], the variance does not increase once it reaches its maximum value that was calculated to be at about 31 dB in [25] for RIRs. This point is approximately reached at the critical distance as it is shown in Fig. 7. However, another reason for lower correlations for the spectral measure VAR and SFM may be that they equally judge spectral peaks which are perceived as being very annoying [19] and spectral dips that do not decrease the perceived quality to a great extend.

Table 4 shows the correlations of subjective ratings with signal-based objective measures. It can be seen that the signal-based measures show lower correlation to subjective data than the system-based measures. The LPC-based measures outperform purely signal-based measures like the SSRR. By far, the highest correlations are obtained by the measures PSM and PSMt that rely on auditory models. PSMt, in addition to PSM, evaluates short-time behaviour of the correlations of internal signal representations and focuses on low correlations as it is done by human listeners [34]. The auditory-model based measures show even higher correlation than RDT, SRMR and OMCR although the latter were designed to explicitly judge reverberation. The performance of RDT and OMCR measures can be adjusted by changing internal parameters. By this, higher correlation to the specific set of samples can be obtained. However, we used standard values for these parameters given in [30, 32]. 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 due to distortions like late echoes and pre-echoes which are much more prominent than the coloration effect [37]. As tested measures are incapable to explicitly judge those influences further development of objective measures is required.

7. CONCLUSION

Objective quality measures were compared to data from subjective listening tests to identify objective measures that can be used to evaluate the performance of listening-room compensation algorithms. Channel-based measures showed higher correlations between objective and subjective data than most of the tested signal-based measures. However, especially if impulse responses are not properly accessible, e.g. as for dereverberation suppression algorithms, measures that incorporate sophisticated auditory models should be used for quality assessment. The Perceptual Similarity Measure (PSM) showed highest correlations to subjective data. A detailed assessment of coloration effects and distortions that may be introduced by LRC algorithms is a topic for future research.

8. REFERENCES

[1] 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.

[2] J.B. Allen, "Effects of Small Room Reverberation on Subjective Preference," *Journal of the Acoustical Society of America (JASA)*, vol. 71, no. 1, p. S5, 1982

[3] D.A. Berkley, "Normal Listeners in Typical Rooms - Reverberation Perception, Simulation, and Reduction," in *Acoustical factors affecting hearing aid performance*, pp. 3–24. University Park Press, Baltimore, 1980.

[4] IEC 1998, "Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index," 1998.

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

[6] J. Benesty Y. Huang and J. Chen, "A Blind Channel Identification-Based Two-Stage Approach to Separation and Dereverberation of Speech Signals in a Reverberant Environment," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 5, pp. 882–895, Sept. 2005.

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

[8] S. J. Elliott and P. A. Nelson, "Multiple-Point Equalization in a Room Using Adaptive Digital Filters," *Journal of the Audio Engineering Society*, vol. 37, no. 11, pp. 899–907, Nov. 1989.

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

[10] S. Goetze, E. Albertin, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Quality Assessment for Listening- Room Compensation Algorithms," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, USA, Mar. 2010.

[11] R. Huber, *Objective Assessment of Audio Quality Using an Auditory Processing Model*, Ph.D. thesis, University of Oldenburg, Germany, 2003.

[12] 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.

[13] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Objective Measures for the Evaluation of Noise Reduction Schemes," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2005.

[14] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer, "Multi-Channel Listening-Room Compensation using a Decoupled Filtered-X LMS Algorithm," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, USA, pp. 811–815, Oct. 2008.

[15] 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.

[16] 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, pp. 976–979, Mar. 2009.

[17] L. D. Fielder, "Practical Limits for Room Equalization," in *AES Convention (Audio Engineering Society)*, New York, NY, USA, vol. 111, pp. 1 – 20, Sept. 2001.

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

[19] M. Kallinger and A. Mertins, "Room Impulse Response Shaping – A Study," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. V101–V104, 2006.

[20] S. M. Griebel and M. S. Brandstein, "Wavelet Transform Extrema Clustering for Multi-Channel Speech Dereverberation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Pocono Manor, PA, USA, Sept. 1999.

[21] B. Yegnanarayana and P.S. Murthy, "Enhancement of Reverberant Speech Using LP Residual Signal," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 3, pp. 267–280, May 2000.

[22] P.A. Naylor, N.D. Gaubitch, and E.A.P. Habets, "Signal-Based Performance Evaluation of Dereverberation Algorithms," *Journal of Electrical and Computer Engineering, Article ID 127513*, 2010.

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

[24] 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.

[25] J.J. Jetzt, "Critical Distance Measurement of Rooms from the Sound Energy Spectral Response," *Journal of the Acoustical Society of America (JASA)*, vol. 65, no. 5, pp. 1204–1211, May 1979.

[26] 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.

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

[28] J.H.L. Hansen and B. Pellom, "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms," in *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, Sydney, Australia, vol. 7, pp. 2819–2822, Dec. 1998.

[29] 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.

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

[31] T.H. 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.

[32] J.Y.C. Wen and P.A. Naylor, "Objective Measurement of Colouration in Reverberation," in *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Poznan, Poland, Sept. 2007, pp. 1615–1619.

[33] 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.

[34] 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 - Special Issue on Objectiv Quality Assessment of Speech and Audio*, vol. 14, no. 6, 2006.

[35] 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, pp. 3615–3622, June 1996.

[36] 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.

[37] Sound samples, correlation patterns, and MATLAB code for quality assessment available online at http://www.ant.uni-bremen.de/∼goetze/aes2010/.

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

Measure	Method	Reverberant	Col./dist.	Distant	Overall
SSRR	All EQs	0.332	0.290	0.432	0.403
	LS-EQ	0.596	0.152	0.648	0.673
	WLS-EQ	0.802	0.737	0.827	0.798
	ISwPP	0.703	0.338	0.652	0.641
FWSSRR	All EQs	0.440	0.404	0.568	0.551
	LS-EQ	0.792	0.037	0.821	0.852
	WLS-EQ	0.943	0.778	0.989	**0.984**
	ISwPP	0.807	0.458	0.763	0.752
WSS	All EQs	0.603	0.580	0.762	0.713
	LS-EQ	0.788	0.441	0.866	0.847
	WLS-EQ	0.892	0.760	0.959	0.981
	ISwPP	0.909	0.580	0.874	0.860
ISD	All EQs	0.639	0.347	0.693	0.684
	LS-EQ	0.352	0.444	0.364	0.408
	WLS-EQ	**0.964**	0.709	**0.999**	0.980
	ISwPP	0.701	0.374	0.672	0.677
CD	All EQs	0.627	0.414	0.702	0.674
	LS-EQ	0.445	0.371	0.478	0.523
	WLS-EQ	0.893	0.811	0.942	0.933
	ISwPP	0.797	0.416	0.749	0.731
LAR	All EQs	0.517	0.384	0.612	0.588
	LS-EQ	0.332	0.504	0.356	0.419
	WLS-EQ	0.934	0.779	0.985	0.976
	ISwPP	0.749	0.386	0.700	0.686
LLR	All EQs	0.663	0.432	0.753	0.713
	LS-EQ	0.469	0.365	0.495	0.544
	WLS-EQ	0.893	0.845	0.956	0.962
	ISwPP	0.836	0.450	0.795	0.778
LSD	All EQs	0.735	0.480	0.814	0.780
	LS-EQ	0.753	0.065	0.809	0.832
	WLS-EQ	0.867	0.834	0.923	0.921
	ISwPP	0.865	0.500	0.833	0.823
BSD	All EQs	0.043	0.303	0.237	0.195
	LS-EQ	0.526	0.470	0.634	0.602
	WLS-EQ	0.848	0.644	0.938	0.937
	ISwPP	0.907	0.635	0.926	0.937
OMCR	All EQs	0.051	0.134	0.028	0.052
	LS-EQ	0.519	**0.827**	0.620	0.538
	WLS-EQ	0.631	0.233	0.640	0.649
	ISwPP	0.163	0.453	0.239	0.257
RDT	All EQs	0.670	0.505	0.790	0.746
	LS-EQ	0.690	0.430	0.776	0.767
	WLS-EQ	0.810	0.745	0.883	0.933
	ISwPP	0.943	0.574	0.922	0.901
SRMR	All EQs	0.526	0.242	0.593	0.511
	LS-EQ	0.437	0.154	0.509	0.538
	WLS-EQ	0.747	**0.885**	0.734	0.803
	ISwPP	0.785	0.451	0.722	0.695
PSM	All EQs	0.803	**0.627**	0.902	0.866
	LS-EQ	0.844	0.642	0.905	0.877
	WLS-EQ	0.843	0.832	0.922	0.971
	ISwPP	**0.982**	0.653	0.963	0.945
PSM_t	All EQs	**0.915**	0.611	**0.950**	**0.942**
	LS-EQ	**0.895**	0.558	**0.958**	**0.920**
	WLS-EQ	0.896	0.761	0.960	**0.984**
	ISwPP	0.979	**0.787**	**0.970**	**0.964**
PESQ	All EQs	0.596	0.349	0.691	0.628
	LS-EQ	0.465	0.354	0.503	0.552
	WLS-EQ	0.842	0.772	0.898	0.874
	ISwPP	0.893	0.458	0.847	0.816

Table 4: Correlations $|\rho|$ of MOS values of subjective ratings and signal-based objective measures (maxima are indicated in boldface).