A STUDY ON SPEECH QUALITY AND SPEECH INTELLIGIBILITY MEASURES FOR QUALITY ASSESSMENT OF SINGLE-CHANNEL DEREVERBERATION ALGORITHMS

Stefan Goetze[1,5], *Anna Warzybok*[2,5], *Ina Kodrasi*[2,5], *Jan Ole Jungmann*[3], *Benjamin Cauchi*[1,5], *Jan Rennies*[1,5], *Emanuël A.P. Habets*[4], *Alfred Mertins*[3], *Timo Gerkmann*[2,5], *Simon Doclo*[1,2,5], *Birger Kollmeier*[1,2,5]

[1] Fraunhofer Institute for Digital Media Technology IDMT, Oldenburg, Germany
[2] University of Oldenburg, Department of Medical Physics and Acoustics, Oldenburg, Germany
[3] University of Lübeck, Institute for Signal Processing, Lübeck, Germany
[4] International Audio Laboratories, Erlangen, Germany
[5] Cluster of Excellence Hearing4all

ABSTRACT

This paper reports on the evaluation of several objective quality measures for predicting the quality of the dereverberated speech signals. The correlations between subjective quality assessment for single-channel dereverberation techniques and objective speech quality as well as speech intelligibility measures are analyzed and discussed. Six different single-channel dereverberation algorithms were included in the evaluation to account for different types of distortions. The subjective quality was assessed along the four attributes *reverberant*, *colored*, *distorted* and *overall quality* following the recommendations of ITU-T P.835. The objective measures included system-based, i.e. channel-based, as well as signal-based measures.

Index Terms— Objective quality measures, subjective listening test, speech dereverberation

1. INTRODUCTION

Generally, the signal quality of an audio signal can be assessed in two ways: subjectively and objectively. Subjective quality measurements are based on the subjective opinion of the listeners, measured by e.g. ranking the signal quality on a predetermined scale. To obtain results with a relatively low variation, a reasonable number of listeners is needed which is time-consuming as well as costly. To overcome this, a number of objective measures have been developed to predict speech quality. For this, high correlation to the subjective rating in the respective task is essential [1, 2]. However, still no commonly accepted quality measure for assessing of dereverberation algorithms has been proposed. In this contribution, the applicability of several objective measures applied to speech signals processed by single-channel dereverberation algorithms is analyzed and discussed. Different classes of dereverberation algorithms (cf. Section 2) are included in the evaluation to account for different types of distortions that may be introduced by dereverberation algorithms, e.g. pre-, late-, and ringing echoes, distortions of the remaining speech signal or residual reverberation [3]. The subjective quality is assessed for the dimensions *reverberant*, *colored*, *distorted* and *overall quality* (cf. also Section 3.1) and compared to the results of the objective measures. While this paper focuses on the correlations between subjective data and objective quality measures, the detailed analysis of the raw subjective data can be found in [4]. The objective measures encompass several system- and signal-based measures that are summarized in Section 3.2. Section 4 discusses the correlation analysis and Section 5 concludes the paper.

This work was partially supported by the project Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS, project no. 316969) funded by the European Commission (EC), as well as by the DFG-Cluster of Excellence EXC 1077/1 "Hearing4all".

The International Audio Laboratories Erlangen (AudioLabs) is a joint institution of the University of Erlangen-Nürnberg and Fraunhofer Institute for Integrated Circuits IIS.

2. ALGORITHMS UNDER TEST

The following algorithms have been included for the listening tests: least-squares equalization [5], impulse-response reshaping by weighting of the error used for least-squares minimization [6] or by aiming at *hiding* the equalized impulse response (IR) under the temporal masking threshold [7]. Furthermore, two spectral suppression methods have been assessed: one based on a statistical reverberation model [8, 9], and a second one based on an estimate of the room impulse response (RIR) [10].

The most simple impulse response equalization technique is known as least-squares (LS) equalization [5] which is defined by

$$\mathbf{c}_{\mathrm{EQ}}^{\mathrm{LS}} = \mathbf{H}^{+}\mathbf{d}, \tag{1}$$

with $\mathbf{H}^{+}$ and $\mathbf{d}$ denoting the Moore-Penrose pseudo inverse of the channel convolution matrix and the desired system response, respectively. A weighting of the error signal with an appropriate window function $\mathbf{w}$, e.g.

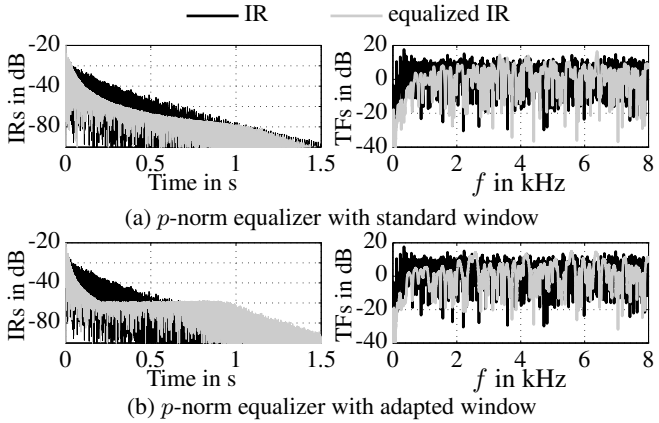
$$\mathbf{W} = \mathrm{diag}\left\{\mathbf{w}\right\}, \tag{2}$$

$$\mathbf{w} = [\underbrace{1, 1, ..., 1}_{N_1}, \underbrace{w_0, w_1, ..., w_{N_2-1}}_{N_2}]^T, \tag{3}$$

$$w_i = 10^{\frac{3\alpha}{\log_{10}(N_0/N_1)}\log_{10}(i/N_1)+0.5}, \tag{4}$$

leads to the weighted least-squares (WLS) equalizer

$$\mathbf{c}_{\mathrm{EQ}}^{\mathrm{WLS}} = (\mathbf{WH})^{+}\mathbf{Wd}. \tag{5}$$



(a) p-norm equalizer with standard window

(b) p-norm equalizer with adapted window

Fig. 1: RIR $\mathbf{h}$ (RT60=1.1 s) and IRs of equalized systems $\mathbf{v} = \mathbf{H}\mathbf{c}_{\mathrm{EQ}}^{p-\mathrm{norm}}$ in dB (left) as well as respective transfer functions (right) for the p-norm equalizer with (a) standard window and (b) adapted window.

Contrary to complete equalization as in (1), RIR shaping as e.g. in (5) emphasizes the suppression of late parts of the equalized IR to prevent perceptually disturbing late echoes [2, 3]. In (3) and (4), the constants N_0, N_1 and N_2 are defined as follows [7]:

$N_0 = (t_0 + 0.2)f_s$, $N_1 = (t_0 + 0.004)f_s$ and $N_2 = L_h + L_{\mathrm{EQ}} - 1 - N_1$ with t_0, f_s, L_h and L_{EQ} being the time of the direct path, the sampling rate, the length of the RIR and of the equalizer, respectively. The constant α is a factor that influences the steepness of the window. For $\alpha = 1$ the window corresponds to the masking found in human subjects [11].

The third algorithm under test is the p-norm-based IR shaping approach according to [7] implemented here in two variants, i.e. (i) using the window function defined in (4) with $\alpha = 1$ (p-norm standard, PNormS) and (ii) using the same approach with a window function limited to -60 dB (p-norm adapted, PNormA). The latter is motivated by the assumption that reverberation cannot be perceived more than -60 dB below the main peak of the RIR. The resulting equalized IRs and respective transfer functions are shown in Fig. 1.

Furthermore, two methods for dereverberation in the spectral domain are assessed: For the reverberation suppression rule according to [8, 9], the clean speech was estimated using the log-spectral amplitude estimator [12] and the late reverberant spectral variance was estimated using [9] assuming that the frequency-independent reverberation time RT60 and direct-to-reverberation ratio (DRR) were known. The last method assessed in the subjective listening tests is a frequency-domain technique proposed in [10], called F-Inv in this paper, that designs an approximate regularized inverse filter

$$G_\delta[k] = \frac{H^*[k]}{\|H[k]\|^2 + \delta}, \tag{6}$$

where $k = 0, \ldots, K-1$, denotes the frequency index with $K \geq L_h$, $H[k]$ and $H^*[k]$ denote the acoustic transfer function and it's conjugate, respectively, and δ is a regularization parameter [10]. Since the filter in (6) is acausal and causes pre-echoes in the processed signal, a single channel speech enhancement scheme is incorporated afterwards, for which the spectral analysis is done using an FFT size $K' \ll K$. The parameters K = 262144, δ = 10^{-4} and $K' = 512$ at an overlap of 50 % [10] were used for processing the signals under test.

Table 1 summarizes the algorithms under test and their respective acronyms.

Table 1: Different dereverberation approaches and their respective acronyms.

Acronym	Description of method
LS-EQ	Least-squares equalizer $\mathbf{c}_{\mathrm{EQ}}^{\mathrm{LS}}$ according to (1) without weighting of error signal, i.e. $\mathbf{w}_{\mathrm{I}} = \mathbf{1}$ if using (5)
WLS-EQ	Weighted least-squares equalizer $\mathbf{c}_{\mathrm{EQ}}^{\mathrm{WLS}}$ according to (5) with window function according to (4)
PNormS	Standard p-norm RIR shaping according to [7] using the window function according to (4) with $\alpha = 1$
PNormA	Adapted p-norm RIR shaping according to [7] using the window function according to (4) with $\alpha = 1$ limited to a minimum of -60 dB
Spec Sup	Spectral reverberation suppression according to [8, 9]
F-Inv	Regularized spectral inverse with pre-echo removal according to [10]

3. QUALITY ASSESSMENT

3.1. Subjective Quality Assessment

21 normal-hearing listeners were asked to assess the quality of speech signals regarding four attributes: *reverberant*, *colored*, *distorted*, and *overall quality*. For each algorithm, speech quality was assessed for 5 reverberation times (RT60): 0.7 s, 1.0 s, 1.1 s, 1.6 s, and 3.8 s. The RIRs for RT60s of 0.7 s, 1.1 s, 1.6 s. and 3.8 s were simulated using the image method [13] for a room size of 6 x 4 x 2.6 m^3. The RIR with RT60 of 1 s was measured in a real room having a size of 3.9 x 3.1 x 2.3 m^3. For all RIRs, the source-receiver distance was fixed at 0.54 m. Each sound sample (sampled at $f_s = 16$ kHz), consisting of 2 sentences taken from the Oldenburg sentence corpus [14], was convolved with the respective RIRs. The reverberated speech signals were then processed by the algorithms described in Section 2. The filter lengths for LS and WLS equalizers were $L_{\mathrm{EQ}} = 8192$ and for the p-norm approaches $L_{\mathrm{EQ}} = 16384$. The Spec Sup algorithm processed the reverberant speech signals in the short-term spectral domain based on an estimate of the RT60 and the DRR [8, 9]. Altogether, 35 speech samples (5 RT60s x 6 algorithms and 5 reverberated samples as a reference condition) were included in the subjective quality assessment. The root mean square (RMS) values of the processed speech samples were set to the RMS of the original (clean) signals to allow for a comparison across the different algorithms. Prior to the measurements, a training session was conducted to familiarize the listeners with the stimuli under test and the task. All speech samples were presented diotically via headphones (Sennheiser HDA200) in quiet at a comfortable level which could be adjusted individually during the training session. The listeners' task was to assess the speech quality at the 5-point mean opinion score (MOS) scale according to the ITU-T P.835 recommendations [15] (with slight modifications, cf. [2]) ranging from 1 (corresponding to bad *overall quality*, very *reverberant*, *distorted* or *colored* signals) to 5 (corresponding to excellent *overall quality*, not *reverberant*, *colored* or *distorted* signals) with steps of 0.1. The order of listening conditions (RT60s and algorithms) was randomized across listeners.

3.2. Objective Quality Measures

In general, various objective quality measures exist that can be applied for quality assessment of dereverberated speech signals.

Following [2, 16, 3], we separated them into (i) measures that are based on the IR or the transfer function of a system (channel-based, i.e. system-based measures, cf. Section 3.2.1) and (ii) measures that are based on signals (cf. Section 3.2.2). The set of the objective quality measures used in this study is similar to that used in [16, 3]. The detailed description of the implementation of the objective quality measures for the dereverberation algorithms can be found in [3]. Generally, for listening-room compensation (LRC) algorithms (e.g. LS, WLS, p-norm), both the filter impulse response $\mathbf{c}_{\mathrm{EQ}}$ and the RIR $\mathbf{h}$ are available during simulations, thus system-based measures can be used. However, e.g. algorithms working in short-term spectral domain (Spec Sup and F-Inv) can only be assessed based on processed signal and reference signal (by means of signal-based measures).

3.2.1. System-Based Measures and Speech Intelligibility Measures

Acoustic impulse responses can be characterized by several objective measures, see e.g. [1, 3, 17], often based on a ratio between early and late part of the respective IR.

We will analyse different measures assessing IRs in time- as well as frequency domain. The ratio between the energy of the first 50 ms (or the first 80 ms) after the main peak to the overall energy of the RIR is called *Definition*. It is denoted by D50 or D80, respectively [17]. The *Clarity* [17] is the logarithmic ratio of the energy within 50 ms (80 ms) after the main peak to the rest of the IR, denoted here by C50 and C80, respectively. The *Direct-to-Reverberation-Ratio* (DRR) [18] is defined as the logarithmic ratio between the main peak and all others. The *Central Time* (CT) [17] is the center of gravity of the energy of an impulse response (IR). In addition to the previously introduced six measures commonly used to describe IRs, two more quality measures are analyzed in this study developed for assessing reverberation explicitly, i.e. the *Reverberation Quantization Measure* (RQ) [19] and the *perceivable Reverberation Quantization Measure* (pRQ) [20] that assess the energy of the equalized IR exceeding the temporal masking limit on the logarithmic scale, i.e. the amount of reverberation that is perceivable.

Dereverberation by mean of channel equalization often aims at archiving a flat spectrum of an equalized transfer function. Thus, the *variance* (VAR) of the logarithmic equalized transfer function was proposed in [21] to evaluate LRC algorithms. A second measure that assesses the flatness of the equalized transfer function is the so-called *Spectral Flatness Measure* (SFM) [22].

A further class of objective measures used in this study are speech intelligibility (SI) measures. We evaluated the *Speech Transmission Index* (STI) [23], the *Rapid STI* (RASTI) [24], and the *STI for Telecommunication Systems* (STITEL) [25]. Although these algorithms have been developed to assess speech intelligibility rather than speech quality, they may, in general, be used for both purposes. We chose the implementations of the SI measures based on the knowledge of the used IRs and therefore the SI algorithms are considered as system-based measures although signal-based implementations exist as well.

3.2.2. Signal-Based Measures

For spectral-domain reverberation suppression algorithms (such as Spec Sup or F-Inv in this study), equalized linear time-invariant (LTI) IRs or transfer functions are not accessible or appropriate for objective testing. Thus, these algorithms have to be assessed based on the processed signals. Several signal-based measures exist that can, in general, be used for assessment of dereverberation approaches. Due to the large extent of this topic, the chosen measures are just briefly summarized in the following and the interested reader is referred to the respective references for further reading. A more detailed summary can be found in [3].

Simple measures like the *Signal-to-Noise-Ratio* (SNR) or the *Segmental Signal-to-Reverberation Ratio* (SSRR) [26] have been adopted from SNR-based measures for noise-reduction quality assessment [27]. The *Frequency-Weighted SSRR* (FWSSRR) [1] and the *Weighted Spectral Slope* (WSS) [1] represent a first step towards exploiting findings in 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 linear predictive coding (LPC) models such as the *Log-Area Ratio* (LAR) [28], the *Log-Likelihood Ratio* (LLR) [1], the *Itakura-Saito Distance* (ISD) [1], and the *Cepstral Distance* (CD) [1]. As a further extension towards modeling of the human auditory system the *Bark Spectral Distortion* (BSD) [29] compares perceived loudness based in spectral masking effects.

More recent objective measures like the *Reverberation Decay Tail* (RDT) [30], the *Speech-to-Reverberation Modulation Energy Ratio* (SRMR) [31] and the *Objective Measure for Coloration in Reverberation* (OMCR) [32] have been specifically designed for the assessment of dereverberation algorithms.

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 [27]. Thus, we also incorporated the *Perceptual Evaluation of Speech Quality* (PESQ) measure [33] and the *Perceptual Similarity Measure* (PSM, PSM$_t$) from PEMO-Q [34] that compares internal representations according to the auditory model described in [35]. PSM$_t$ calculates the 5th percentile of the PSM output vector and showed high correlation with subjective ratings for quality assessment of audio codes [34].

4. RESULTS AND DISCUSSION

Table 2 shows the correlations between subjective data and system-based quality measures and Table 3 the correlations between subjective data and signal-based quality measures. Correlations $|r|$ of 0.75 or greater are highlighted using bold-face letters. Stars indicate statistically significant correlations ($p < 0.05$). For each quality measure the correlations are shown (i) for the case that all algorithms under test are considered (see 'All algos' in Tables 2 and 3) and (ii) the mean and standard deviation for the correlations for single algorithms ('Mean (Std)'). It can be seen, that the correlations are generally higher, if they are applied to single algorithms than if they are used for comparison over all algorithms.

To illustrate this, Fig. 2 exemplarily shows the correlation plot for the quality measure PESQ and the respective correlations for each single algorithm are given in Table 4. It can be seen e.g. for the attribute distorted (lower left panel) that the subjective and objective ratings mostly correlate well for the single algorithms (e.g. between $r_{\mathrm{SpecSup}} = 0.67$ and $r_{\mathrm{PNormA}} = 0.93$ with a mean of $r_{\mathrm{Mean}} = 0.84$), however, the correlation if all algorithms are considered is considerably lower ($r_{\mathrm{All}} = 0.43$).

The correlations in Tables 2 and 3 indicate, that none of the quality measures correlates well with the attribute *colored* which is in consilience with the findings in [16, 3]. This reflects the difficulties that listeners reported assessing coloration in general and, furthermore, distinguishing between coloration and distortion in the signals

Table 2: Pearson correlations coefficient $|r|$ between MOS values of subjective ratings and system-based objective measures (values above 0.75 are indicated in boldface).

Measure	Method	Reverberant	Colored	Distorted	Overall
D50	All algos	.30	.45*	.52*	.55*
	Mean (Std)	**.85** (.06)	.38 (.34)	**.81** (.15)	**.75** (.23)
D80	All algos	.27	.41*	.48*	.51*
	Mean (Std)	**.85** (.06)	.38 (.34)	**.77** (.19)	**.75** (.17)
C50	All algos	.30	.44*	.54*	.54*
	Mean (Std)	**.84** (.08)	.39 (.34)	**.82** (.13)	.71 (.26)
C80	All algos	.25	.41*	.48*	.48*
	Mean (Std)	**.81** (.09)	.41 (.32)	**.77** (.20)	.70 (.22)
CT	All algos	-.48*	-.14	-.20	-.50*
	Mean (Std)	**.91** (.05)	.39 (.29)	**.77** (.10)	**.81** (.19)
DRR	All algos	.42*	.23	.48*	.56*
	Mean (Std)	**.84** (.10)	.39 (.30)	**.80** (.15)	.63 (.29)
RQ	All algos	-.48*	-.23	-.37	-.47*
	Mean (Std)	**.81** (.18)	.40 (.38)	**.79** (.06)	.71 (.29)
pRQ	All algos	-.54*	.06	-.04	-.46*
	Mean (Std)	**.90** (.03)	.43 (.23)	**.78** (.06)	**.81** (.20)
VAR	All algos	.03	-.55*	-.35	-.35
	Mean (Std)	.44 (.29)	.64 (.25)	.60 (.26)	.43 (.35)
SFM	All algos	.13	.52*	.39	.47*
	Mean (Std)	.50 (.33)	.69 (.34)	.71 (.21)	.52 (.31)
STI	All algos	.28	.37	.46*	.45*
	Mean (Std)	**.90** (.02)	.37 (.33)	**.79** (.12)	**.79** (.22)
RASTI	All algos	.27	.35	.40*	.42*
	Mean (Std)	**.88** (.07)	.33 (.34)	**.78** (.10)	**.76** (.22)
STITEL	All algos	.27	.38	.44*	.44*
	Mean (Std)	**.89** (.04)	.35 (.34)	**.81** (.12)	**.79** (.23)

Table 3: Pearson correlations coefficient $|r|$ between MOS values of subjective ratings and signal-based objective measures (values above 0.75 are indicated in boldface).

Measure	Method	Reverberant	Colored	Distorted	Overall
BSD	All algos	.35*	.04	.03	.23
	Mean (Std)	.59 (.37)	.56 (.23)	.70 (.24)	.58 (.31)
CD	All algos	-.81*	-.56*	-.45*	**-.81***
	Mean (Std)	**.89** (.12)	.46 (.30)	**.86** (0.09)	**.77** (0.26)
FWSSRR	All algos	.74*	.65*	.49*	**.82***
	Mean (Std)	**.86** (.14)	.37 (.20)	.67 (.15)	**.75** (.23)
ISD	All algos	-.60*	-.31	-.36*	-.60*
	Mean (Std)	**.81** (.26)	.40 (.25)	**.80** (.11)	**.75** (.25)
LAR	All algos	**-.79***	-.47*	-.38*	**-.77***
	Mean (Std)	**.90** (.07)	.38 (.21)	**.77** (.15)	.69 (.36)
LLR	All algos	**-.80***	-.64*	-.48*	**-.82***
	Mean (Std)	**.88** (.09)	.43 (.22)	**.78** (.18)	.73 (.26)
LSD	All algos	-.29	-.40*	-.27	-.34*
	Mean (Std)	.73 (.30)	.40 (.31)	.62 (.24)	.56 (.23)
OMCR	All algos	.23	.28	.29	.35*
	Mean (Std)	.42 (.33)	.46 (.32)	.52 (.32)	.44 (.34)
PESQ	All algos	.70*	.66*	.43*	**.77***
	Mean (Std)	**.75** (.15)	.53 (.25)	**.84** (.11)	.72 (.27)
PSM	All algos	.69*	.52*	.37*	.72*
	Mean (Std)	**.88** (.09)	.52 (.32)	**.89** (.06)	**.77** (.34)
PSMt	All algos	**.80***	.27	.24	.68*
	Mean (Std)	**.86** (.12)	.46 (.29)	.73 (.28)	.74 (.31)
RDT	All algos	-.38*	-.23	-.11	-.38*
	Mean (Std)	**.87** (.13)	.47 (.32)	**.82** (.09)	**.78** (.29)
SSRR	All algos	.52*	.22	.23	.50*
	Mean (Std)	.65 (.29)	.36 (.16)	.53 (.33)	.53 (.33)
SNR	All algos	.06	.12	-.12	.06
	Mean (Std)	.23 (.11)	.51 (.22)	.37 (.22)	.44 (.35)
SRMR	All algos	.49*	.29	.03	.43*
	Mean (Std)	.54 (.33)	.37 (.24)	.52 (.38)	.61 (.36)
WSS	All algos	-0.66*	-0.65*	-0.43*	**-0.76***
	Mean (Std)	**.85** (.11)	.54 (.26)	**.87** (.16)	**.79** (.30)

under test. E.g. for the LS equalizer, typical time-domain artefacts (late echoes) sometimes sound like frequency-domain distortions.

Correlations for the system-based measures shown in Table 2 show that most time-domain measures correlate well with the attribute *reverberant* and *distorted* at least on the basis of single algorithms. The correlations for the frequency-domain measures is much lower (cf. also [16, 3]). Although the speech intelligibility measures

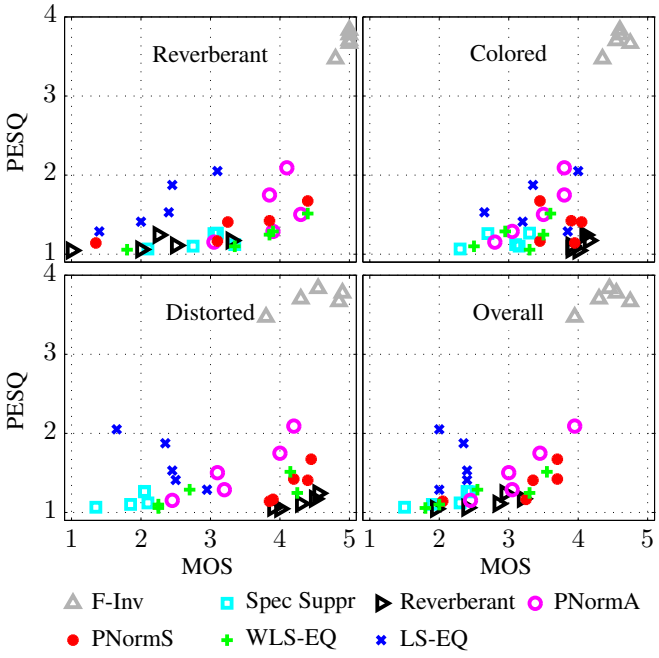


Fig. 2: PESQ score as function of MOS for the four attributes.

Table 4: Pearson correlation coefficient r, between PESQ scores and MOS for the four attributes and the algorithms under test.

	Reverberant	Colored	Distorted	Overall
Reverberant	0.59	0.72	0.93	0.80
LS-EQ	0.92	0.30	-0.90	-0.16
WLS-EQ	0.83	0.55	0.76	0.88
PNormS	0.84	-0.22	0.92	0.74
PNormA	0.59	0.92	0.93	0.96
Spec Sup	0.57	0.37	0.67	0.84
F-Inv	0.88	0.64	0.74	0.66
Mean (Std)	0.75 (0.15)	0.53 (0.25)	0.84 (0.11)	0.72 (0.27)
All	0.70	0.66	0.43	0.77

(STI and its variants) have not been designed to assess speech quality, they show very high correlations for all attributes besides *colored* for the data under test in this study. A thorough study regarding speech intelligibility measurements will be subject to future work.

The correlations for the signal-based measures in Table 3 show that high correlations can also be achieved for the attributes *reverberant* and *distorted*, again mostly for single algorithms. Regarding comparison between different algorithms ('All algos'), FWSSRR and LLR show highest correlation (0.82) for the attribute *overall quality* and LLR and PSMt show highest correlation (0.8) for the attribute reverberant.

5. CONCLUSION

This paper presented a correlation analysis between data from subjective listening test for dereverberated sound samples and different objective quality measures. While several objective quality measures showed high intra-class correlations, i.e. for single algorithms (e.g. for comparison of different parameters), much lower correlation was found if several algorithms are compared with each other. Surprisingly, speech intelligibility measures like the STI correlate well with subjective rating for quality although they aim at assessing speech intelligibility rather than quality.

6. REFERENCES

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

[2] S. Goetze, E. Albertin, J. Rennies, E.A.P. Habets, and K.-D. Kammeyer, "Speech Quality Assessment for Listening-Room Compensation," in *38th AES Conference*, Pitea, Sweden, July 2010, pp. 11–20.

[3] S. Goetze, *On the Combination of Systems for Listening-Room Compensation and Acoustic Echo Cancellation in Hands-Free Telecommunication Systems*, Ph.D. thesis, Dept. of Telecommunications, University of Bremen (FB-1), Bremen, Germany, 2013.

[4] A. Warzybok, I. Kodrasi, J. O. Jungmann, E.A.P. Habets, T. Gerkmann, A. Mertins, S. Doclo, B. Kollmeier, and S. Goetze, "Subjective Speech Quality and Speech Intelligibility Evaluation of Single-Channel Dereverberation Algorithms," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC 2014)*, Antibes, France, Sep. 2014.

[5] 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.

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

[7] A. Mertins, T. Mei, and M. 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, DOI:10.1109/TASL.2009.2025789.

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

[9] E.A.P. Habets, S. Gannot, and I. Cohen, "Late Reverberant Spectral Variance Estimation based on a Statistical Model," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–773, Sep. 2009.

[10] I. Kodrasi, T. Gerkmann, and S. Doclo, "Frequency-Domain Single-Channel Inverse Filtering for Speech Dereverberation: Therory and Practice," in *Proc. 2014 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 5214–5218.

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

[12] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean–square error log-spectral amplitude estimator," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.

[13] 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.

[14] K. Wagener, V. Kühnel, and B. Kollmeier, "Entwicklung und Evaluation einesSatztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (In German language)," *Zeitschrift für Audiologie / Audiological Acoustics*, vol. 38, pp. 86–95, 1999.

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

[16] 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, March 2010.

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

[18] 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.

[19] T. Mei and A. Mertins, "On the Robustness of Room Impulse Reshaping," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

[20] J.O. Jungmann, T. Mei, S. Goetze, and A. Mertins, "Room Impulse Response Reshaping by Joint Optimization of Multiple P-Norm Based Criteria," in *Proc. 19th European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, Aug. 2011, pp. 1658–1662.

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

[22] 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.

[23] H. Steeneken and T. Houtgast, "Basics of the STI-Measuring Method," *Past, Present, and Future of the Speech Transmission Index, International Symposium on STI, The Netherlands*, pp. 13–44, Oct. 2002.

[24] P. Larm and V. Hongisto, "Experimental Comparison between Speech Transmission Index, Rapid Speech Transmission Index, and Speech Intelligibility Index," *Journal of the Acoustical Society of America (JASA)*, vol. 119, no. 2, pp. 1106–1117, Feb. 2006.

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

[26] 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.

[27] 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)*, Eindhoven, The Netherlands, 2005.

[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, Dec. 1998, vol. 7, pp. 2819–2822.

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

[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.