Objective Quality Measures for Dereverberation Methods based on Room Impulse Response Equalization

Eugen Albertin, Jan Rennies, and Stefan Goetze

Fraunhofer Institute for Digital Media Technology, Project group Hearing, Speech, and Audio Technology, 26129 Oldenburg

Abstract

Hands-free communication devices are commonly used e.g. in offices or car environments. These devices have to cope with several problems such as ambient noise, acoustic echoes and room reverberation. Reverberation is caused by reflections at room boundaries in enclosed spaces and increases with the spatial distance between the loudspeaker and the near-end user. A high amount of reverberation decreases speech intelligibility at the position of the near-end listener. Approaches for listening-room compensation (LRC) can be used to reduce the influence of reverberation on acoustic signals by equalization of the acoustic channel. However, such algorithms may also have negative impact on the perception of the sound quality of the dereverberated signal if they are not designed properly. In this contribution, objective measures that are expected to be able to assess quality of LRC algorithms are identified since a commonly accepted objective measure is not yet available.

Listening Room Compensation

The equalization of acoustic channels is topic of research for some decades now [1]. Figure 1 shows a common setup for LRC with the equalization filter $\mathbf{c}_{\mathrm{EQ}}$ preceding the room impulse respose (RIR) $\mathbf{h}$.

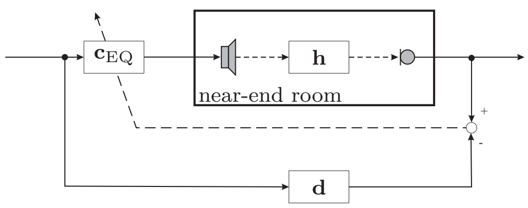

Figure 1: General setup for listening-room compensation.

To reduce the reverberation that is introduced by the RIR $\mathbf{h}$ the equalizer $\mathbf{c}_{\mathrm{EQ}}$ tries to minimize the Euclidean distance between the concatenated system of $\mathbf{c}_{\mathrm{EQ}}$ convolved with $\mathbf{h}$ and the desired target system $\mathbf{d}$ [2]. The minimization of the mean squared error signal $\mathrm{E}\{||\mathbf{e}_{\mathrm{EQ}}||^2\} = \mathrm{E}\{||\mathbf{H}\mathbf{c}_{\mathrm{EQ}} - \mathbf{d}||^2\}$ leads to the least-squares equalizer [2]

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

Here, $\mathbf{H}^{+}$ is the Moore-Penrose pseudoinverse of the convolution matrix $\mathbf{H}$ built up by the RIR coefficiens. A weighting of the error vector $\mathbf{e}_{\mathrm{EQ}}$ as used in [3] leads the weighed least squares equalizer that was also used in this contribution. Another concept used here for LRC is RIR shaping as introduced in [4] which is based on solving a generalized eigenvalue problem.

Quality Assessment

The objective quality assessment for the dereverberated signals generated by the LRC approaches described above was done by using different objective measures that were classified in (i) measures that are based on the impulse response of the equalized systems $v[k] = h[k] * c_{\mathrm{EQ}}[k]$ or of the transfer function of the system $V[n] = H[n]C_{\mathrm{EQ}}[n]$ (system-based measures) and (ii) measures that are based on signals only. Six different measures belong to the class of system-based measures. The *Definition* calculates a ratio between the first 50ms (80ms) after the main peak and the overall energy of the equalized system. The *Clarity* [5] is the logarithmic ratio of 50ms (80ms) after the main peak to the rest of the impulse response. The *Direct-to-Reverberation-Ratio* (DRR) describes the logarithmic ratio between the main peak and the rest of the impulse response [5]. The *variance* (VAR) [1] of the logarithmic transfer function $V[n]$ and the so-called *Spectral Flatness Measure* (SFM) [6] evaluate the equalization performance in frequency domain.

For some dereverberation approaches, e.g. blind reverberation suppression [7], impulse responses are not accessible, thus, only signal based measures can be used for quality assessment. The simplest signal-based measures are the *Segmental Signal-to-Reverberation Ratio* (SSRR) [8] that compares the reference signal and signal under test block by block in time-domain and the *Frequency-Weighted SSRR* (FWSSRR) [9] that is similarly computed in the frequency-domain. Furthermore, measures based on the LPC models such as the *Log-Area Ratio* (LAR) [10], the *Log-Likelihood Ratio* (LLR) [9], the Itakura-Saito Distance (IS) [9], and the *Cepstral Distance* (CD) [9] can be used. Recently, quality measures have been proposed based on the human auditory system. We tested the *Bark Spectral Distortion* measure (BSD) [11], the *Reverberation Decay Tail* (RDT) measure [12] and the *Objective Measure for Coloration in Reverberation* (OMCR) [13]. Additionally, we analyzed the *Speech-to-Reverberation Modulation Energy Ratio* (SRMR) [14] that, by the way, was the only non-intrusive measure. Moreover, we used the *Perceptual Evaluation of Speech Quality* (PESQ) measure [9, 15] and the *Perceptual Similarity Measure* (PSM) from PEMO-Q [16] that were developed for objective quality assessment in the field of audio coding.

A more elaborate description and a correlation analysis with the subjective data that was carried out can be found in [17]. The tested signals, impulse responses and

Work was supported in parts by German Research Foundation DFG under Grant Ka841-17.

transfer functions of the systems used for this correlation analysis are available at [18].

Results

Figure 2 shows the equalized impulse responses and transfer functions obtained by the three LRC approaches.

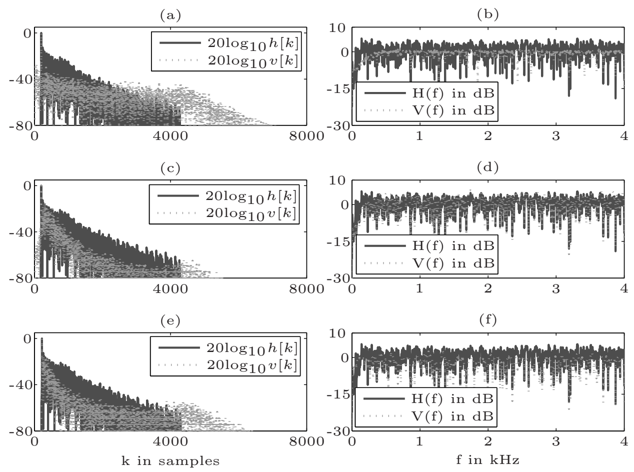

Figure 2: Different LRC approaches: (a) least-squares equalizer in time domain; (b) transfer function of (a); (c) weighted least-squares equalizer in time domain; (d) transfer function of (c); (e) room shaping [4] in time domain; (f) transfer function of (e); Reverberation time of RIR was $\tau_{60} = 0.5$s, EQ filter length was 4096 samples, v[k] denotes equalized impulse response.

The least-squares algorithm (1) achieves a flat spectrum but leads to mathematically small but perceptually annoying late echoes even for large filter lengths that can be seen around sample $k = 4000$ in Figure 2 a). The analysis of the subjective data yields to the conclusion that the distortions in time-domain caused by the LRC algorithms are perceptually much more annoying than the distortions in frequency-domain [17]. Thus, the least-squares algorithm shows a poor overall performance in comparison to the shaping approaches in terms of perceived quality.

Furthermore, as shown in [17] in more detail, quality measures based on impulse responses (with the exception of the DRR) measure were highly correlated with the subjective data for the tested attributes *reverberant*, *distant* and *overall quality*. No objective measure showed a high correlation with the subjective ratings for the attribute *colored/distorted*. The frequency-domain system-based measures VAR and SFM showed low correlation since they failed in judgement of the specific distortions introduced by the algorithms [17].

Only the *Perceptual Similarity Measure* (PSM and PSMt) showed a high correlation with the subjective data from the class of signal-based measures. PSM and PSMt use an auditory model and compare internal representations that are assumed to be found in the human auditory system. These internal representations are correlated in a certain scheme considering the weighting of the loudness levels. The authors believe that the application of an auditory model is necessary to obtain objective quality measures that show high correlation to subjective quality assessment.

Conclusion

Various objective quality measures were briefly described that can be used for the evaluation of LRC algorithms. The correlations between objective and subjective data show that the measures based on impulse responses (like *Definition* or *Central Time*) showed high correlation to subjective quality assessment. Most of the signal-based measures failed to assess the quality of the dereverberated signals. Only the *Perceptual Similarity Measure* (PSM) showed high correlation to subjective data.

Literatur

[1] J. N. Mourjopoulos. Digital Equalization of Room Acoustics. *Journal of the Audio Engineering Society*, 42(11):884–900, November 1994.

[2] 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)*, pages 976–979, Rotterdam, The Netherlands, March 2009.

[3] M. Kallinger and A. Mertins. A Spatially Robust Least Squares Crosstalk Canceller. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages I–177 – I–180, 15-20 April 2007.

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

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

[6] J. D. Johnston. Transform Coding of Audio Signals using Perceptual Noise Criteria. *IEEE Journal on Selected Areas in Communication*, 6(2):314–232, February 1988.

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

[8] P.A. Naylor and N.D. Gaubitch. Speech Dereverberation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, September 2005.

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

[10] 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)*, volume 7, pages 2819–2822, Sydney, Australia, December 1998.

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

[12] 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, September 2006.

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

[14] 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, September 2008.

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

[16] 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*, 14(6), 2006. Special Issue on Objectiv Quality Assessment of Speech and Audio.

[17] 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, März 2010.

[18] Sound samples available online at. URL `http://www.ant.uni-bremen.de/~goetze/icassp2010/index.html`.