Room Impulse Response Shaping based on Estimates of Room Impulse Responses

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

[1] *University of Bremen, Dept. of Communications Engineering, D-28334 Bremen, Email: goetze@uni-bremen.de*

[2] *University of Oldenburg, Signal Processing Group, D-26111 Oldenburg, Email: markus.kallinger@uni-oldenburg.de*

[3] *University of Lübeck, Institute for Signal Processing, D-23538 Lübeck, Email: alfred.mertins@isip.uni-luebeck.de*

Abstract

Modern hands-free telecommunication systems have to reduce different acoustic disturbances of the desired speech signal, amongst them echoes of the far-end speaker due to the acoustic coupling between loudspeaker and microphone and room reverberation for the near-end listener caused by reflections at the room boundaries. Common systems for listening-room compensation (LRC) try to design an equalization filter which is the inverse of the room impulse response to achieve spectral flatness of the concatenated overall system of equalizer and room transfer function. Such designs need reliable knowledge of the room impulse response (RIR) which is not available in real systems. Furthermore, it could be shown that shaping approaches which preserve the masking effects of room impulse responses lead to perceptually better results especially in case of estimation errors. In this contribution a system for room impulse response shaping is analyzed depending on the influences of realistic RIR estimates, which are obtained by an acoustic echo canceller (AEC).

Listening Room Compensation

In an acoustic scenario for listening-room compensation (LRC) the equalizer $c_{EQ}[k]$ precedes the acoustic channel - the room impulse response (RIR) $h[k]$ - as depicted in Figure 1. Its goal is to reduce the influence of the RIR at the position of the reference microphone where the near-end user of the telecommunication system is assumed to be located. A straightforward inversion of a RIR by a stable, causal infinite impulse response (IIR) filter is not possible, in general, since RIRs have hundreds of zeros inside and outside close to the unit circle in the z-domain [4]. Thus a common approach is the least-squares equalizer (LS-EQ) [3] which minimizes the error signal $e_{EQ}[k]$ and by this the distance between the overall system of $c_{EQ}[k]$ convolved with $h[k]$ and a desired target system $d[k]$. Least-squares equalizers need very accurate estimates of the RIR which may not always be available, since a common RIR is time variant, e.g. due to possible speaker movements, and may have a length of thousands of taps. Thus, the RIR has to be identified by an appropriate adaptive filter and, especially in periods of initial convergence or after RIR changes, the system identification may be insufficient for a good design of the equalizer [1].

The goal of RIR shaping approaches [2] is to shorten the RIR to maximize the energy in its first 50ms since the early reflections in the first 50ms are known to enhance speech intelligibility while the reverberant tail of the RIR decreases speech intelligibility. To avoid late echoes occurring in the equalized system RIR shaping which aims in reducing the room reverberation time τ_{60} is a promising approach [2].

All methods for LRC need knowledge about the RIR which has to be measured or estimated adaptively and thus may by deficient. In hands-free systems acoustic echo cancelers (AECs) are common systems to reduce echoes for the far-end listener. This is done by identifying the RIRs.

System Identification

Since the AEC is an adaptive filter which identifies the room impulse response (RIR) its estimate can be used for the EQ design.
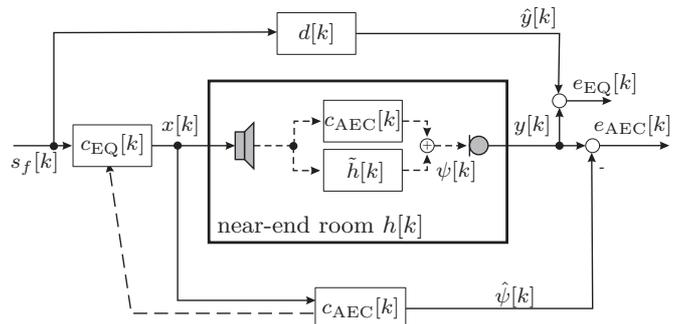

Figure 1: Combined system with LRC filter $c_{EQ}[k]$ and acoustic echo canceller $c_{AEC}[k]$. The RIR can be split into a part modeled by the AEC $c_{AEC}[k]$ and an estimation error $\tilde{h}[k]$.

As depicted in Figure 1 the RIR $h[k]$ can be split up into one part $\hat{h}[k] = c_{AEC}[k]$ which is correctly identified by the AEC and an estimation error $\tilde{h}[k]$:

$$h[k] = \hat{h}[k] + \tilde{h}[k] = c_{AEC}[k] + \tilde{h}[k] \tag{1}$$

Estimation errors $\tilde{h}[k]$ mainly have two reasons: 1) the system identification is incomplete due to unfinished filter convergence and/or a too short filter to identify the RIR on its full length, 2) the equalizer is designed for the spatial position of the reference microphone and the user is not located at this position.

Simulation Results

The influences of the two possible errors described before on the performance of the equalizer will be analyzed in the following.

Insufficient AEC convergence: Figure 2 compares the conventional least-squares EQ with the RIR shaping approach according to [2] by means of the Bark spectral distortion (BSD) measure [5] which is common to evaluate dereverberation algorithms. The BSD is shown for different states of convergence of the AEC filter which delivers the RIR estimate.
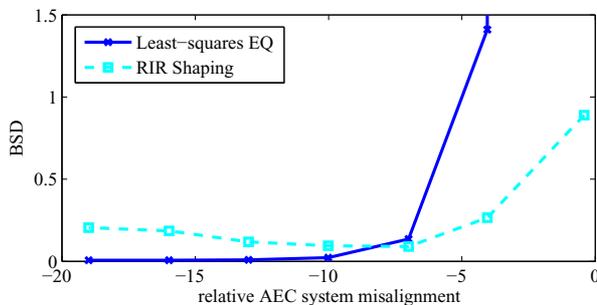

Figure 2: Bark spectral distortion (BSD) depending on the normalized AEC convergence state $D = ||\tilde{h}[k]||^2/||h[k]||^2$.

It can be seen from Figure 2 that the EQ performance is slightly better for the LS design for a good RIR estimate ($D < -7$dB). However, in practical situations the AEC has to follow the time variant RIR and thus its convergence state may be worse than -7dB for most of the time. In this region the LS-EQ may lead to severe signal distortions as it can be seen from the steep rise of the BSD while the RIR shaping approach shows considerably better performance.
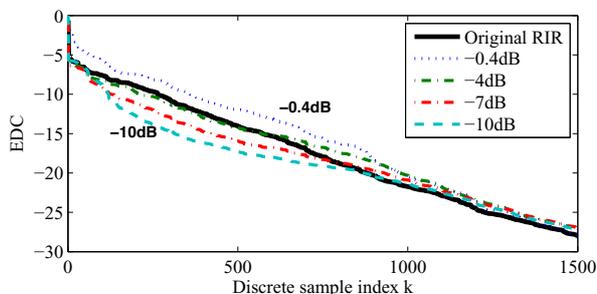

Figure 3: Energy decay curve (EDC) depending on the AEC convergence state $D = ||\tilde{h}[k]||^2/||h[k]||^2$.

Figure 3 shows the performance of the RIR shaping approach for different states of AEC convergence. The thick solid black line shows the energy decay curve (EDC) of the original RIR and the other lines the EDCs of the equalized systems for different AEC convergence states. It can be seen that a bad AEC convergence state decreases the performance of the EQ but that it is robust up to $D = -4$dB and in combination with Figure 2 it can be seen that even for a very bad AEC convergence the distortions introduced to the desired signal are by far less than those for the LS-EQ. Thus, it can be stated that the RIR shaping approach according to [2] is more robust with regard to RIR estimation errors.

Spatial mismatch: Figure 4 shows the robustness against spatial mismatch with respect to the BSD measure. An EQ is designed for a fixed position (x=2.6 meters and y=2.1 meters from the room corner). This point is located in the centers of the left and right subplot. Then the EQ is applied to spatially differing positions (10 cm in every direction). It is clear that the LS-EQ performs best at the exact position it is designed for (center of left plot) but the performance heavily degrades for a spatial mismatch. In contrast to that the RIR shaping approach according to [2] (right subplot) is spatially robust which is a very important property since the user of the hands-free system will be located at some spatial distance from the reference microphone.
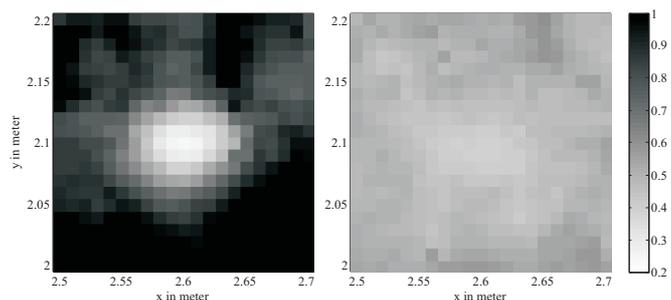

Figure 4: BSD measure for spatial mismatch for LS-EQ (left) and room impulse response shaping (right).

Conclusion

In this contribution listening-room compensation approaches are evaluated for the case of imperfect system identification due to incomplete RIR estimation or due to spatial mismatch. From the presented results it can be stated that room impulse-response shaping approaches are very promising for real-world hands-free systems because they perform well even for imperfect knowledge about the room impulse response, which often is assumed to be perfectly known in literature but, in general, has to be estimated.

Literatur

[1] 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.

[2] M. Kallinger and A. Mertins. Room Impulse Response Shortening by Channel Shortening Concepts. In *Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA , USA*, pages 209–212, Oct. 30 - Nov. 2 2005.

[3] J. N. Mourjopoulos, P. M. Clarkson, and J.K. Hammond. A Comparative Study of Least-Quares and Homomorphic Techniques for the Inversion of Mixed Phase Signals. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1858–1861, 1982.

[4] S. T. Neely and J. B. Allen. Invertibility of a Room Impulse Response. *Journal of the Acoustical Society of America (JASA)*, 66:165–169, July 1979.

[5] S. Wang, A. Sekey, and A. Gersho. An Objective Measure for Predicting Subjective Quality of Speech Coders. *IEEE J. Selected Areas of Communications*, 10(5):819–829, June 1992.