ABOUT ROOM RESPONSE EQUALIZATION AND DEREVERBERATION

Matti Karjalainen
Tuomas Paatero

Lab. of Acoustics and Audio Signal Processing
Helsinki University of Technology
P.O.Box 3000, FI-02015 TKK, Finland
matti.karjalainen@hut.fi

John N. Mourjopoulos
Panagiotis D. Hatziantoniou

Audio Group, Wire Communications Lab.
Electrical and Computer Engineering Department
University of Patras, Patras, 265 00 Greece
mourjop@wcl.ee.upatras.gr

ABSTRACT

Improvement of sound reproduction in rooms by signal processing techniques has been studied seriously for more than twenty years. Such methods are called for example dereverberation, room (response) equalization, and room (response) correction. Understanding the limitations and quality criteria for such methods is gradually achieved through studies of this highly multidimensional problem. In the present paper we discuss the terminology and goals, compare some recent approaches of these techniques in practical cases, and finally point out future challenges.

1. INTRODUCTION

Improvement of sound reproduction through loudspeakers in rooms using DSP is a challenging problem due to the multidimensionality of sound quality perception and the inherent complexities of the physical systems involved. In a carefully designed listening room with high-quality loudspeakers there is need only for minor adjustments of the reproduction chain. Even in such cases flaws may appear that are expensive to be corrected by acoustic redesign. In spaces such as living rooms or car cabins the quality of audio reproduction may gain much more from proper equalization.

In this paper we discuss some of the developments in room equalization and dereverberation and propose directions and ideas towards a more full picture of the related techniques (for background literature see for example [1–10]. It should be pointed out that such methods rely on some measurements of the Room Transfer Function (RTF), and are implemented via the existing audio reproduction chain, i.e., no additional audio channels are employed for reverberation control.

2. CONCEPTS AND GOALS

The terminology used in related literature is not very systematic. This is partly due to the many aspects and goals of room response improvement. Here we first define our concepts and terms.

Room (response) equalization is used here as a general term to denote methods that improve objective or subjective quality of reproduction in a room. *Room response correction* (rather than just 'room correction') can be used instead for the same purpose. *Room response inversion* (rather than *room inversion*) is the process to find an inverse filter that works as a room response equalizer to counteract undesired properties in the loudspeaker-room response. *Loudspeaker equalization* means improving the (anechoic) response of the loudspeaker only or improving the acoustic coupling to the room.

Dereverberation denotes methods that reduce objective or subjective reverberation at least at some frequencies. This may mean lowering the ratio of late (reverberant) vs. early sound level and/or shortening the decay rate of reverberation. *Enhanced reverberation* is sometimes needed in a case where the objective or subjective level of reverberation to early sound needs to be increased.

Modal equalization means methods in which the behavior of resonances is controlled so that the decay time of the modes can be changed, in most cases shortened [4]. This is in practice possible only at low frequencies (up to about 200 Hz).

Requirements of room response equalization and/or dereveberation vary widely according to specific reproduction tasks and room acoustics conditions. In carefully designed listening and sound control rooms the conditions are controlled primarily acoustically and by high-quality loudspeakers, while in home audio the conditions can vary in many ways, often having highly reverberant spaces, strong reflections, and boomy bass due to undamped low-frequency modes or wrong positioning of loudspeakers. Cars are another common listening environment with their own acoustic properties. PA systems for speech require high intelligibility, whereby counteracting reverberation is a common problem.

Listeners' preferences and *program material* can also remarkably affect the preferences of good reproduction and equalization needed. Often the desired properties of 'pleasant listening' deviate from 'HiFi-listening', for example by requiring extra sound level at low frequencies, enhanced presence or intelligibility at mid frequencies, boosted/tilted high frequencies, or special spatial effects.

For the robust application of room response equalization, there are many limitations and constraints from the points of view of *room acoustics*, *psychoacoustics*, and *signal processing techniques*. The lack of knowledge in perceptual aspects is found the most limiting factor, which calls for increased research on perception and sound quality modeling. However, presently such methods must rely on existing knowledge of psychoacoustics.

Among other constraints are variation of listener position, variation of room response due to different acoustic factors, limitations in on-site response measurements for equalizer design, dynamic range limitations when boosting too weak frequency ranges, compromising between pre-echo and reverberation created by equalizers in dereverberation, allowable processing latency from millisecond range to seconds depending on application, requirements for real-time processing power (especially for multi-channel reproduction) and challenges in joint optimization in reproduction of multiple channels, up to tens or hundreds of them for wave field synthesis.

A map illustrating this multidimensionality of alternatives for room inverse filter design is shown in Fig. 1. It can be observed that each alternative choice per line can be combined with any option appearing in the other lines.

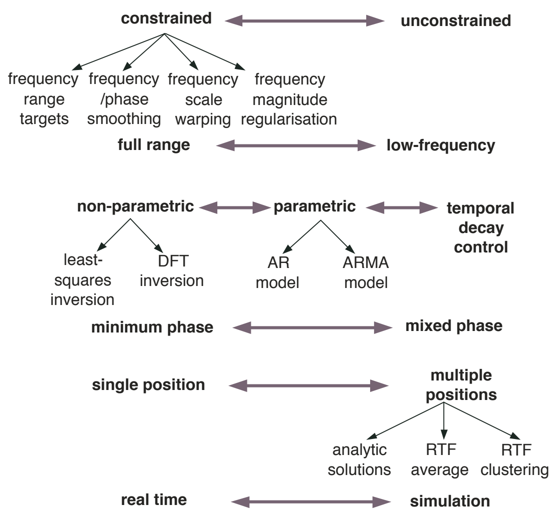

Figure 1: Alternative choices in room equalizer design.

3. ROOM RESPONSE INVERSION

The main principles of inverting a measured response can be divided into: (a) *non-parametric* (least mean squares or direct FTT inversion with or without controlled frequency resolution and with or without target response processing) [8, 11] and (b) *parametric* inversion/estimation (by AR/ARMA modeling [3, 12, 13], or even analytical RTF evaluation). A third class of equalization, where the given response as such is not inverted, is (c) *temporal decay control*. This is possible in practice only at low frequencies [4, 5].

As can be seen in Fig. 1, one way to classify room equalizers is by dividing them into *minimum-phase* and *mixed-phase* designs. Minimum-phase filters can only achieve RTF magnitude equalization affecting only the minimum-phase part of phase response, while mixed-phase designs can also affect the excess phase (all-pass) RTF component and hence potentially remove some of the room reverberation [8]. Parametric inverse filters are typically of minimum-phase type, although mixed-phase design is possible.

An important feature of room response inversion is that perceptually insignificant deep notches (transfer function zeros) should not be converted to prominent and long-ringing resonances (poles). This can be achieved by proper *smoothing techniques*. In complex smoothing [11] the basic method is to smooth the magnitude response by linear combination of frequency bins, although other weighting profiles can be applied as well. In frequency-warped FIR equalizer designs [3] the least squares criterion of AR modeling (linear prediction) corresponds to magnitude squared weighting, counteracting more strongly the effect of transfer function zeros. Another possible weighting could follow the loudness scale, being roughly approximated by the square root of magnitude.

The *frequency resolution* used in equalizer design can follow different profiles, such as the traditional 1/3-octave smoothing. Critical band scales, i.e., the ERB and the Bark scales (see comparison in [14, Fig. 1]), find motivation from the theories of hearing. Arbitrary resolution profiles can be realized by the complex smoothing of [11], while frequency-warped designs fit most easily to the Bark scale [14].

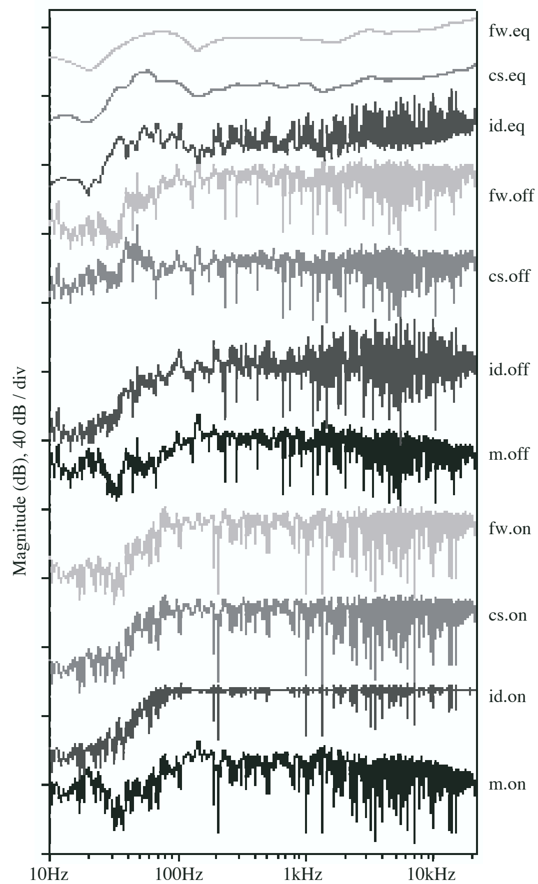

Figure 2: Magnitude spectra of measured responses ‘m.on’ and ‘m.off’ for the on and off equalization spatial position; equalizer responses of frequency warped ‘fw.eq’, complex smoothed ‘cs.eq’, and ideal ‘id.eq’ inverse filter; equalized responses for the on and off equalization position for the three equalizers: ‘fw.on’, ‘cs.on’, ‘id.on’, ‘fw.off’, ‘cs.off’ and ‘id.off’.

3.1. Magnitude response equalization

From innumerous possible choices, a nonparametric mixed-phase and a parametric minimum-phase method were selected for comparison in a practical case study in order to gain deeper understanding of their behavior. A direct inverse filter was taken as a third case. The comparison is illustrated in Figs. 2 and 3 for

- *complex smoothed* nonparametric least-squares inversion [8], denoted ‘cs’ (FIR equalizer of order 8192),
- *frequency-warped* parametric WFIR design [3] (order 12 for low- and 12 for high-frequency range), denoted ‘fw’,
- *direct inversion* [8], ‘id’, no smoothing (order 8192).

Figure 2 plots the magnitude spectra of two measured responses (‘on’ and ‘off’) for the three designed equalizer filters, and the equalized system responses in the ‘on’ and ‘off’ positions, where ‘on’ means position of equalizer design and ‘off’ is another position in the room. The loudspeaker had a lower roll-off frequency of about 80 Hz, compensated in target response design. The room was a listening room of 33 m^2 with fairly well controlled acoustics.

The performances of the complex smoothed and the WFIR equalizer are very similar in terms of the magnitude equalization achieved. The 'on'-position response is essentially flat and the 'off'-position has also good flatness for both equalizers. Thus from the viewpoint of magnitude equalization the same result is achieved by much lower WFIR filter order (24) than for the complex smoothed FIR filter (order 8192). The advantage of the latter case will be seen in Section 3.2.

Any direct inversion of a measured response $H_{\mathrm{m}}(z)$ into equalizer $H_{\mathrm{e}}(z) = 1/H_{\mathrm{m}}(z)$ is problematic in several ways [6, 7]. While the off-line direct equalization in Fig. 1 'id.on' in the measured point is almost perfect, the off-position 'id.off' is useless. Perceptually insignificant zeros with high Q-value in measured frequency response are mapped to resonances that are perceptually disturbing both from spectral and temporal points of view. Such highly resonant terms can also generate significant aliasing errors, when DFT inversion is employed [7]. Hence, least-squares based methods are more robust for direct inverse filter design.

Furthermore, in all cases, frequency ranges beyond high and low roll-off frequencies of the loudspeaker cannot stand boosting to flat response, requiring a proper roll-off target response to be specified. A further problem may be the pre-echo generated after inverse filtering that can become audible and disturbing. Especially for mixed-phase inversion, the latency introduced in order to compensate for the acausal part of inversion may become intolerable in some applications. Hence, various approaches for constraining the action of the inverse filter to avoid the above problems have been proposed over the years.

3.2. Dereverberation

Using the least-squares inversion described, it is possible to design a mixed-phase filter that, under simulated conditions, would dereverberate the signal. This brings additional latency and associated pre-signal convolutional effects due to the acausal filter response component and low-level small scale inversion / arithmetic errors due to least squares or due to finite DFT size. Such low-level pre/post echo will in many cases be inaudible. The artifacts can be seen in the time-frequency results in Fig. 3 'id.on', where about 4000 samples of pre- and post-echo appear to be around 50 dB below the recovered ("delta") direct signal.

In practice, under identical real-time test conditions (in situ), such dereverberation performance cannot be duplicated, possibly due to strong amplification of mismatched (between measurement, filter design, dereverberation test) ambient noise patterns, appearing as highly resonated pre and post echo distortions by poles compensating the mismatched RTF dips [8]. In practice, low-level pre- and post-errors will increase at least 30 dB and hence will become extremely audible and annoying. This, together with the limitation of "sweet spot" design and even some perceptual factors (plausibility for the listener), makes at present the case of full-scale dereverberation impossible for real applications.

The solution proposed by complex smoothing is to reduce both latency (acausal filter component) and also the mismatch sensitivity of the filter to such errors, at the expense of dereverberation performance, as can be seen in Fig. 3 'cs.on'. In practice, such smoothed filters achieve limited phase linearization (for short time intervals), removing only some reverberant energy, or otherwise pre-echo effects can become audible.

WFIR filters, being minimum-phase, will also remove some reverberant energy associated with RTF resonances, without introducing correction for the acausal inverse RTF components, and without introducing latency or achieving phase linearization.

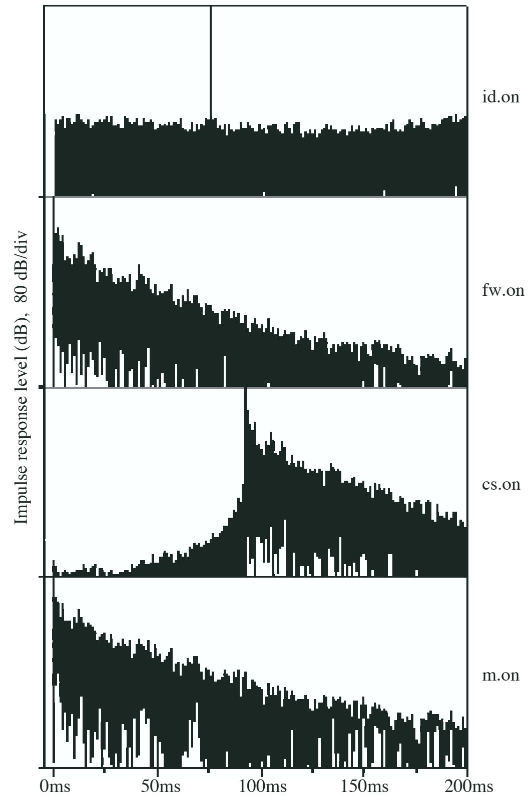

Figure 3: Impulse response level envelopes of: measured 'm.on' position response,complex smoothing equalized 'cs.on', WFIR equalized 'fw.on', and direct inversion equalized 'id.on'.

3.3. Modal equalization at low frequencies

At low frequencies, particularly around 100 Hz and below, room responses of typical listening spaces are different from mid and high frequencies. Room modes do not necessarily overlap, and they can easily be too strong in hard-walled rooms. In addition to extra level, the decay time of such modes can greatly exceed the reverberation time of mid frequencies, resulting in boomy bass response.

Instead of expensive acoustic improvement of such listening spaces, equalization of the reproduction channel can be a more attractive choice. The first to be done is magnitude equalization with proper frequency resolution, such as 1/3- or 1-octave smoothing. This may not remarkably shorten the decay times of high-Q resonances. *Modal equalization* can help more in such cases [4, 5, 9].

In modal equalization the disturbing resonances are counteracted by selective pole-zero filters or by shaping the temporal envelope of the system response. Even adding artificial modes to increase modal density is possible, as far as the boosting of low response level does not overload the speakers.

In Figure 4 we compare the low-frequency performance in a room with problematic modes for three equalizer designs: 'cs' is

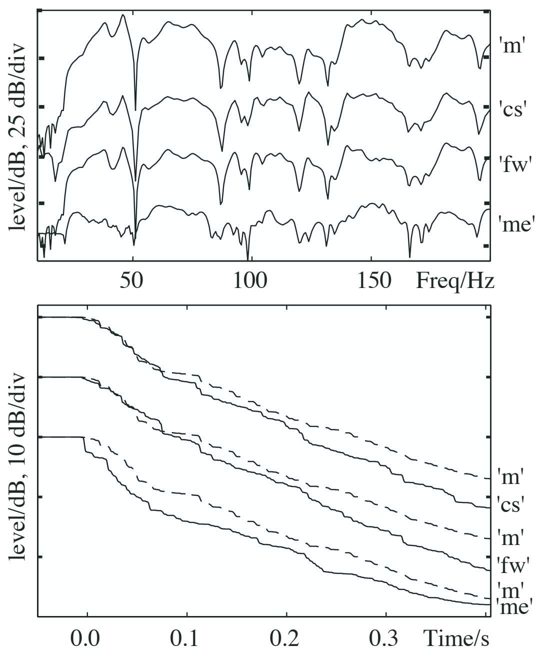

Figure 4: Performance of three room equalizers below 200 Hz: original measurement 'm', complex smoothing 'cs', WFIR design 'fw', and modal equalization 'me'. The upper pane shows magnitude responses and the lower one compares the Schroeder decay plots of equalized responses to the measured one.

a complex smoothed design of order 8192, 'fw' is a WFIR design of order 14 at low frequencies, and 'me' is a modal equalizer with 30 pole and zero pairs for modal compensation without additional magnitude equalization [4]. The 'cs' and 'fw' equalizers behave very similarly by slightly fastening the decay. The modal equalizer makes the initial decay remarkably faster, but it does not improve much the late decay. Notice that the responses are aligned so that the extra delay of 'cs' case is compensated for. All equalizations are evaluated in the same position where the measurement was carried out.

4. DISCUSSION

Given the multidimensionality of the perception of audio reproduction quality in reverberant spaces, the known approaches for assessing the performance of equalization methods is via objective and/or subjective criteria and tests. Objective methods can assess the improvement in RTF (magnitude) spectral flatness and in established acoustic criteria, by evaluating the achieved variation in the processed response in terms of the reverberation time (usually its early decay portion), the direct-to-reverberant ratio, the related clarity (C80) and definition (D50) parameters, etc.

Subjective tests may employ a number of listeners and audio examples, presented either 'off-line' (i.e., via headphones) or 'on-line' (i.e., in-situ via loudspeakers). The subjects are asked to comment on their preferences between the unprocessed and processed case, considering parameters such as 'spectral balance', 'low-frequency extension', time response (e.g. checking reproduction 'clarity' and 'transient response'), 'perceived room size', 'perceived loudness', etc.

The objective evaluation results presented by visual comparison in Figs. 2-4 of this paper, as well as earlier tests [9,10], indicate that both the complex smoothing and the WFIR equalizers can improve most of the above objective and/or subjective criteria.

From the viewpoint of future research the main challenges remain in perceptual studies, requiring extensive research on the perception of sound reproduction in different acoustic spaces. This should also result in perceptual models (such as [15]) and their application to computing objective sound quality measures.

Although equalizer design and real-time DSP are not the main problem anymore, challenges remain also there. Equalizer design techniques need to be developed to gain more detailed control over reproduced sound fields. One such direction is joint design of multichannel equalizers, searching overall optimal solutions instead of separate equalization of each channel. The ultimate from this point of view is room equalization in wave field synthesis.

Another interesting future direction to improve reproduction is signal-dependent processing. This can be for example dynamic control of equalizers, such as the amount of boosting of notches in frequency response without overdriving the loudspeakers. Psychoacoustic knowledge should also be utilized in signal-dependent ways to improve the perceived quality.

5. ACKNOWLEDGMENT

Part of the Academy of Finland projects 53537 and 105629.

6. REFERENCES

[1] J. Mourjopoulos, "On the variation ..." *J. Sound Vib.*, vol. 102, pp. 217–228, 1985.

[2] — —, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, pp. 884–900, 1994.

[3] M. Karjalainen and et al., "Comparison of loudspeaker ..." *J. Audio Eng. Soc.*, vol. 47, no. 1/2, pp. 15–31, 1999.

[4] A. Mäkivirta and al., "Modal equalization of ..." *J. Audio Eng. Soc.*, vol. 51, no. 5, pp. 324–343, 2003.

[5] M. Karjalainen and et al., "Modal equalization by ..." in *Proc. AES 23rd Int. Conf.*, Copenhagen, 2003.

[6] L. D. Fielder, "Analysis of traditional ..." *J. Audio Eng. Soc.*, vol. 51, no. 1, pp. 3–26, 2003.

[7] J. N. Mourjopoulos, "Comments on "Analysis of ..."," *J. Audio Eng. Soc.*, vol. 51, no. 12, pp. 1186–1188, 2003.

[8] P. D. Hatziantoniou and J. N. Mourjopoulos, "Errors in real-time ..." *J. Audio Eng. Soc.*, vol. 52, no. 9, pp. 883–899, 2004.

[9] M. Karjalainen and et al., "Perception of temporal ..." in *Proc. 116th AES Convention*, Berlin, 2004.

[10] J. W. Worley and et al., "Subjective assessments of ..." in *Proc. 118th AES Convention*, Barcelona, 2005.

[11] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave ..." *J. Audio Eng. Soc.*, vol. 48, no. 4, pp. 259–280, 2000.

[12] T. Paatero and M. Karjalainen, "Kautz filters and generalized ..." *J. Audio Eng. Soc.*, vol. 51, no. 1/2, pp. 27–44, 2003.

[13] M. Karjalainen and et. al., "Frequency-zooming ARMA ..." *J. Audio Eng. Soc.*, vol. 50, no. 12, pp. 1012–1029, 2002.

[14] A. Härmä and et al., "Frequency-warped signal processing ..." *J. Audio Eng. Soc.*, vol. 48, no. 11, pp. 1011–1031, 2000.

[15] J. Buchholz and et al., "Room masking: Understanding ..." in *Proc. 110th AES Convention*, Amsterdam, 2001.