AUTOMATED EQUALIZATION FOR ROOM RESONANCE SUPPRESSION

Tobias Corbach, Adrian von dem Knesebeck, Kristjan Dempwolf,
Martin Holters, Peter Sorowka, Udo Zölzer
Dept. of Signal Processing and Communications,
Helmut Schmidt University
Hamburg, Germany
audio@hsu-hh.de

ABSTRACT

Estimating room resonances in locations of big events and looking for counter-measures are normally done by sound engineers, mainly before the beginning of the event. In this paper an automation to enhance the audio quality in event rooms by suppressing the room resonances with a parametric equalizer of several high-Q peak filters is proposed. The room characteristics can be identified with few measurements in the listening area during the event, without applying an additional measuring signal (using its original sound signal). Based on this room characteristics the equalization filters are automatically designed. The results of several rooms tested with the automated equalization for room resonance suppression are presented as well as a discussion on the covered topics.

1. INTRODUCTION

When amateur or semi-professional music bands and solo entertainers are booked e.g. for a party, they typically also bring their own PA equipment which they have selected to support their characteristic sound. However, while the entertainers choose the equipment they bring to the venue, they have no influence on the venue itself. Unfortunately, many rooms which have not been acoustically treated exhibit problems that may impair the listeners' experience if no appropriate counter-measures are taken.

Commonly, the most important counter-measure is to equalize the room frequency response, especially attenuating frequency bands where room resonances lead to booming effects, i.e. overemphasis of certain bass frequencies. Many small music bands and solo entertainers, however, cannot afford a sound engineer and do not possess the necessary experience to properly setup a parametric equalizer for optimal results.

Therefore, a system is desirable which helps in the described scenario by automatically measuring the frequency response and adjusting an equalizer, as depicted in Figure 1. In order to measure the room frequency response, it is favorable not to use a special, usually annoying stimulus signal, but simply to use the original music signal that is being played back to allow measuring with the audience already present.

In the proposed system, the equalization is focused on low frequency resonances, as these are often perceived as the most annoying. Any detected resonances are attenuated with very narrow-band peak filters. The bass loss that might arise as the result is compensated with a broader peak filter with positive gain.

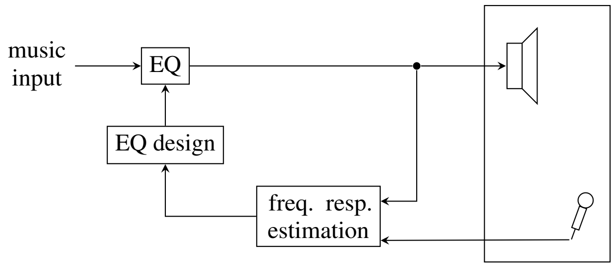

Figure 1: *Overview of the proposed system.*

2. MEASURING TECHNIQUE

In order to design filters to equalize a given room, the room frequency response has to be estimated. There are a lot of different measuring techniques to gain such a room frequency response. Taking noise as stimulus signal is quite common, as it offers a white spectrum, and it is self-evident that it reveals the channel's true transfer function when looking at the response spectrum. For this application a technique without an annoying stimulus signal like noise is needed. A well-known technique normally using white noise but also claimed to be usable with arbitrary stimuli is the "Dual Channel FFT Analysis" [1, 2].

2.1. Dual Channel FFT Analysis

This technique is based on windowing the stimulus signal $x(n)$ and the response signal $y(n)$ and applying a FFT to the M overlapping time windows, leading to their frequency domain representation $X_i(k)$ and $Y_i(k)$. The frequency response is gained by the ratio of the arithmetic mean of the instantaneous cross and auto power spectral densities (CPSD and APSD) $R_{x_iy_i}(k)$ and $R_{x_ix_i}(k)$ of corresponding frames i by

$$H(k) = \frac{\frac{1}{M}\cdot\sum_{i=0}^{M} R_{x_iy_i}(k)}{\frac{1}{M}\cdot\sum_{i=0}^{M} R_{x_ix_i}(k)} = \frac{\sum_{i=0}^{M} Y_i(k)\cdot X_i^*(k)}{\sum_{i=0}^{M} X_i(k)\cdot X_i^*(k)}. \quad (1)$$

It is not directly obvious that the approach holds for non stationary arbitrary stimuli with changing auto-correlation and APSD, as there are two independent sums in the numerator and denominator.

To show why this technique is applicable with arbitrary stimuli, a new method to derive this formula is indicated by starting at the quotient of response and stimulus spectra. The arithmetic mean

$$H(k) = \frac{1}{M}\sum_{i=1}^{M}\frac{Y_i(k)}{X_i(k)} \tag{2}$$

of the FFT block ratios itself gives an estimation of the system's frequency response. This estimation is not sufficient for arbitrary stimuli with changing and non white stimulus signal in a real application, due to additive noise in the response signal. Low stimulus magnitudes at certain frequencies and just a small amount of additive noise in the response lead to very high frequency response magnitude peaks at these frequencies. To handle this, a weighting-factor $w_i(k)$ is introduced, weighting every single frequency response frame i and every single frequency bin k of the desired frequency response

$$H(k) = \frac{1}{\sum\limits_{i=1}^{M} w_i(k)} \cdot \sum_{i=1}^{M} w_i(k)\,\frac{Y_i(k)}{X_i(k)}. \tag{3}$$

Assuming a stationary additive background noise during the measurement and a non white stimulus signal, a frequency dependent instantaneous signal-to-noise ratio arises. To suppress the mentioned errors, frames with frequencies of high SNR should be weighted stronger than frames with frequencies of low SNR. This is realized by choosing the instantaneous APSD of each stimulus frame as weighting factor

$$w_i(k) = R_{x_i x_i}(k) = X_i(k)X_i^*(k) \tag{4}$$

for the corresponding frequency response summand. In turn this leads to the dual channel FFT analysis given in (1).

As a consequence, the possibility to obtain a good estimate of the frequency response of the device under test by using this measuring technique with arbitrary signals is confirmed. It has to be noted that this frequency response is only valid for frequencies with enough energy in the stimulus signal, so the stimulus is arbitrary as long as it fulfills this requirement for the desired frequency range of the room's frequency response.

2.2. Signal Synchronization

When the latency of the device under test is not negligibly small compared to the chosen window size, significant distortions in the estimated frequency response can occur if not dealing with the introduced latency. Determining the offset between signals is a classical task in signal and system theory. The cross-correlation function $r_{xy}(n)$ has its maximum where the similarity between signal $x(n)$ and $y(n)$ is maximal. Unfortunately this is hard to interpret when it comes to periodic signals because then the cross-correlation function will have periodic maxima with a distance of the signal's period. When the response signal $y(n)$ is additionally distorted by a room transfer function and a possible small signal-to-noise ratio, $r_{xy}(n)$ will not present a clear maximum. Inserting inaudible pseudo-noise or similar signals is pretty difficult and conflicts with the requirement not to change the stimulus signal for the measurement.

Since noise has a perfectly non-periodic autocorrelation function, it may be a good idea to use the natively underlying noise in the stimulus signal for matching. A very simple approach is based on the assumption that the actual tonal parts of a music signal are mainly located in the lower bands, whereas the upper bands (e.g. above 7 kHz) are mainly filled by transients and noise. Therefore, if we apply a high pass filter on the cross-correlation $r_{xy}(n)$, just the more noise-like parts of the signals are considered, which is good for matching.

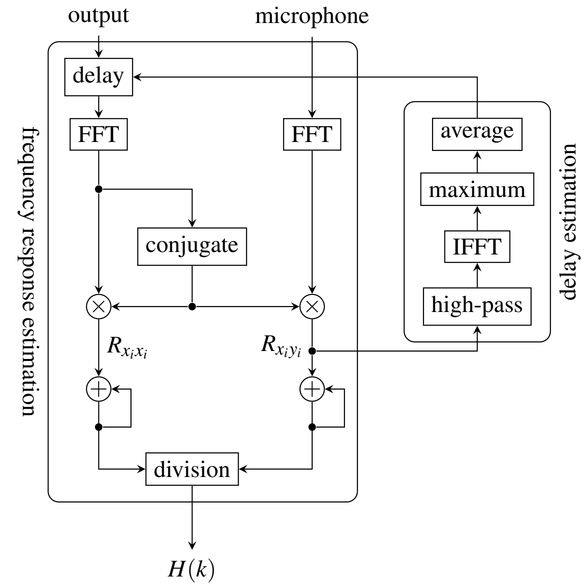

Figure 2: *The dual channel FFT analysis method with automatic synchronization.*

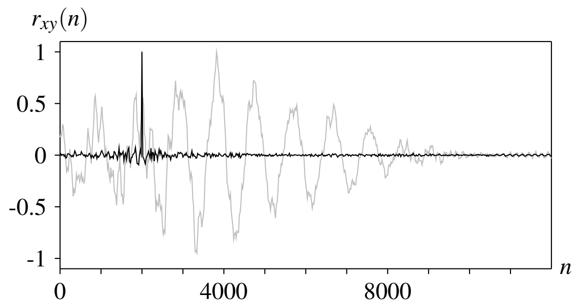

Figure 3: *Normalized cross-correlation (gray) and high pass filtered normalized cross-correlation (black).*

The right-hand part of Figure 2 shows the delay estimation of the system, which adapts the delay to get synchronized stimulus and response signal frames $x_i(n)$ and $y_i(n)$ for the frequency response estimation. Therefore, a high pass filter $H_{HP}(k)$ with a cut-off frequency k_{cut} is applied to the CPSD $R_{x_i y_i}(k)$ of the actual time frame i to get a high frequency version CPSD

$$\tilde{R}_{x_i y_i}(k) = H_{HP}(k)R_{x_i y_i}(k). \tag{5}$$

An IFFT followed by an estimation of its time domain maximum leads to the estimated frame delay

$$n_{delay} = \underset{n}{argmax}\ \left(\tilde{r}_{x_i y_i}(n)\right) \tag{6}$$

in samples. Figure 3 shows the result for one frame when a typical pop music signal is played back and a normalized cross-correlation

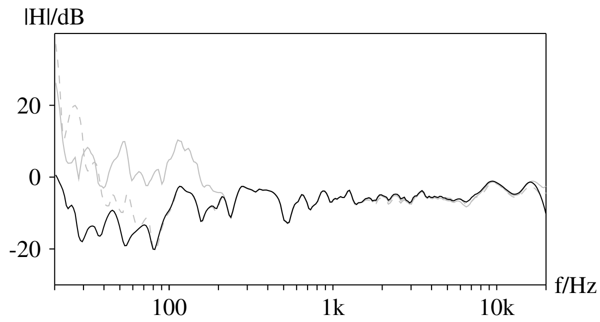

Figure 4: *Room responses with trumpet (gray), counter-bass (gray-dashed) and noise (black) as stimulus signal.*

function with and without an applied high pass pre-filter is computed. It is obvious that an easier estimation of the desired offset can be made by means of a high pass filtered normalized cross-correlation.

To get high robustness of latency estimation, only frames with the following properties are taken into account for the automatic synchronization (K denotes the FFT-size, normally equal to the frame size N in time domain) :

- The arithmetic mean of the magnitude of the CPSD above the high pass pre-filter's cutoff frequency k_{cut} exceeds a specific value ε_{hf}, that holds

$$\frac{1}{K/2-k_{cut}}\sum_{k=k_{cut}}^{K/2}\left|\tilde{R}_{x_iy_i}(k)\right| > \varepsilon_{hf}. \tag{7}$$

- The peak-to-average ratio of the computed cross-correlation function exceeds a specific value ε_{PAR}, that holds

$$\frac{\max\limits_{n}\left(\left|\tilde{r}_{x_iy_i}(n)\right|\right)}{\frac{1}{N}\sum\limits_{n=0}^{N}\left|\tilde{r}_{x_iy_i}(n)\right|} > \varepsilon_{PAR}. \tag{8}$$

Due to different microphone gains and sound card qualities (cross-talk) these parameters depend on the used hardware.

The comparison between signals of single music instruments and conventionally used noise as stimulus of the dual channel FFT analysis shows a good conformity for frequency bands with enough energy but significant differences for bands of poor energy (Figure 4). Applicable uncertainty indicators for the frequency response gained by arbitrary signals could be the variance which occurs when observing the frequency response after every new frame and/or the stimuli signals accumulated power spectral density. A further indicator may be the coherence proposed in [1]. When choosing a music signal of reasonable power spectral density (covering the whole relevant spectrum) as stimulus, the frequency responses of music dual channel FFT and measuring with logarithmic sweeps [3] show high similarities (Figure 5).

3. AUTOMATIC FILTER DESIGN

The proposed approach for room equalization is to attenuate the resonance frequencies with high-Q second order IIR peak filters. The filters are defined by the parameters center frequency, gain,

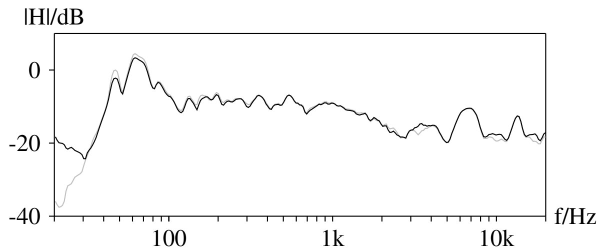

Figure 5: *Room responses measured with the sweep-method (gray) and with dual channel FFT analysis with music as stimulus (black).*

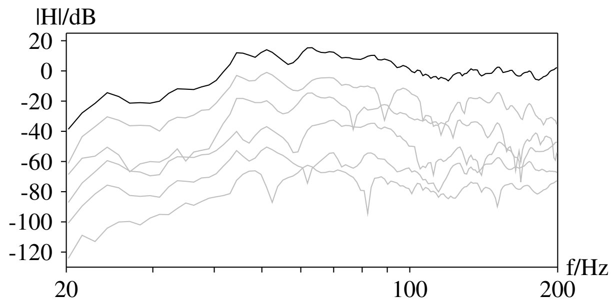

Figure 6: *Room responses (gray) separated with* 15 dB *offsets for better visibility and the resulting maximum norm H_{max} (black).*

and Q-factor and designed according to [4]. It is common to use parametric peak filters for audio equalizing applications, because the parameters are comprehensible for the user. Also a cascade of second order sections provides a flexible system which offers a dynamic order of the overall equalization filter.

Significant room resonances commonly occur in a frequency range from 20 to 200 Hz. At the resonance frequencies, standing waves build up in the room with maxima at the walls and further maxima and nodes in between. Depending on the microphone position, the recorded room response shows peaks at those resonance frequencies which have a wave maximum at that position. The resonances having a node at that position may cause dips which show a less significant influence on the response compared to the peaks caused by the resonance maxima. To ensure that all resonances are captured and may be considered in the optimization process, the room response has to be measured at multiple randomly chosen positions in the listening area, which is known as spatial averaging [5]. The magnitudes of the responses are combined using the maximum norm,

$$H_{max}(k) = \max_i |H_i(k)|. \tag{9}$$

Figure 6 shows the magnitudes of 5 room responses (gray), measured at different locations of the conference room, as described in section 5.1. The responses all have approximately the same mean power level but were separated by 15 dB offsets for better visibility. The resulting maximum norm of the 5 responses is shown by the black curve.

The resonance frequencies are determined by searching the peaks of H_{max}. The relative peak amplitudes are related to the target curve H_t. The target curve is the optimization goal, which is preferably defined by the user according to the loudspeaker system response. In case the speaker systems' response is not

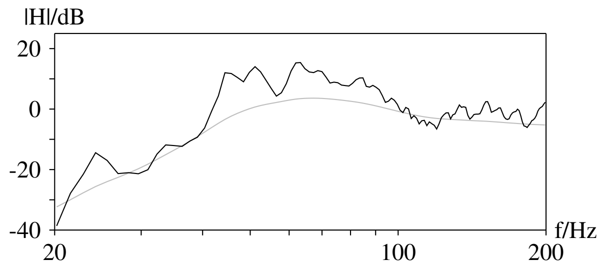

Figure 7: *Target curve H_t (black) and maximum norm H_{max} (gray).*

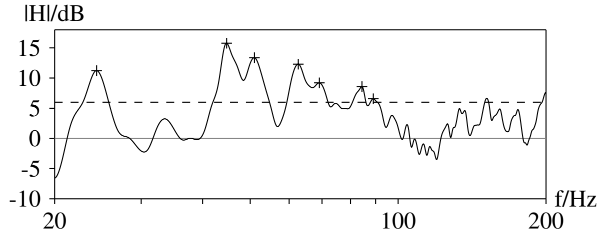

Figure 8: *Compensated response (solid) and detection threshold (dashed). The detected peaks in the range from* 20*to* 120 Hz *are marked by a* +.

known, it is possible to estimate the response by smoothing the maximum norm response. The peaks caused by the resonances are eliminated by applying a one octave smoothing [6]. A certain error is introduced, which shows up especially at frequencies where several resonances occur close to each other. At these frequencies the smoothed response slightly follows the peaks of the resonances. Figure 7 shows H_{max} and the corresponding H_t of the conference room.

The peak finding algorithm is only capable of evaluating the absolute peak amplitudes. To find the relative peaks in relation to the target curve, the not-smoothed H_{max} is compensated by the smoothed response H_t. Figure 8 shows the compensated response H_{max} by the solid curve, H_t now is the straight 0 dB line. The detection threshold is shown by the dashed line. A threshold of 6 dB was chosen as minimum relative peak amplitude. The marks in Figure 7 show the detected peaks in the user-defined optimization range from 20 to 120 Hz.

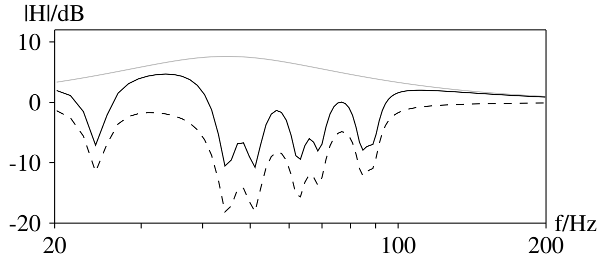

Figure 9: *Equalization curve (dashed), bass loss compensation curve (gray), and combined curve (black).*

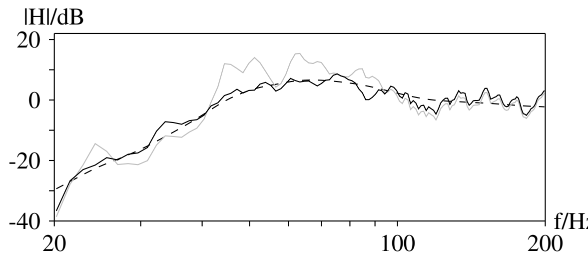

Figure 10: *H_{max} (gray), H_t (dashed), and equalized curve H_{eqd} (black) of the conference room.*

3.1. Equalization of Resonances

The frequencies of the detected peaks are used as the center frequencies of the equalization filters. The gains are determined by evaluating the deviation between not-smoothed and smoothed responses at the previously determined center frequencies. The filter gains are constrained to a maximum attenuation of 20 dB. The Q-factors of the filters are initialized with a value of 17. The optimization target is the optimization of the squared error

$$\min_{c} \left(\sum_{k} \left(|H_{max}(k)| \cdot \left|H_{filt}(c,k)\right| - |H_t(k)| \right)^2 \right), \qquad (10)$$

where $H_{filt}(c,k)$ is the equalization curve of the applied peak filters with the optimization parameters c. For the optimization of the resonance equalization filter the parameters c are the Q-factors of the peak filters. The Q-factors are constrained to a range between 15 and 20. The frequency responses are interpolated from linear to logarithmic frequency scale using splines before the calculation of the error. This way the error is weighted according to the human auditory mechanism. The optimization was done using the *lsqnonlin* function of the MATLAB Optimization Toolbox. The optimized filter curve is shown by the dashed curve in Figure 9.

3.2. Bass Loss Compensation

The presented equalization filter placement obviously results in an overall attenuation of the bass frequencies due to the overlapping of close peaking filters. To compensate this effect and to preserve the perceived bass presence, an additional filter is applied. Analogously to the optimization target of (10) the optimization target for the bass loss compensation filter $H_{basscomp}$ is

$$\min_{c} \left(\sum_{k} \left(\left|H_{filt}(k)\right| \cdot \left|H_{basscomp}(c,k)\right| - |H_t(k)| \right)^2 \right), \qquad (11)$$

with the optimization parameters gain, center frequency, and Q-factor. The center frequency is initialized with the value of the minimum of the smoothed equalized response $H_{eqd} = H_{max} \cdot H_{filt}$, where the minimum is determined in relation to H_t analog to the approach for the peak finding. The Q-factor is constrained to a range between 1 and 5. The bass compensation filter curve is shown by the gray line in Figure 9. The cascaded overall filter, consisting of H_{filt} and $H_{basscomp}$, is shown by the solid black curve. Figure 10 shows the maximum norm response H_{max} (gray), the target curve H_t (dashed), and the filtered response H_{eqd} (solid black).

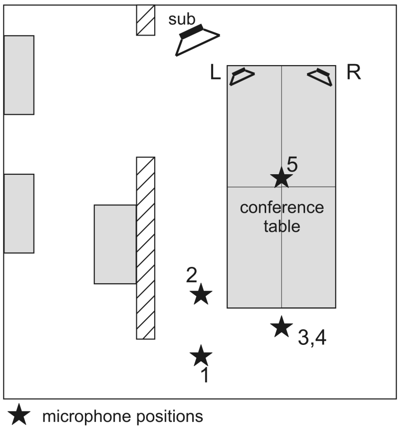

Figure 11: *The conference room.*

4. VALIDATION PROCEDURE

For the evaluation of the presented method a PC software was implemented providing all necessary functions. The measurement with both music and sweep signals and the equalization with parametric peak filters were implemented as well as the plotting of room-, filter-, and equalized room transfer functions. The automatic equalization was performed by an external MATLAB script. For the playback of music a wave player was integrated, but the insertion of the system into an existing signal chain was also used. Several rooms in the range of 100 to 300 m^3 were analyzed using the described method. All tested venues showed some acoustical problems, e.g. the arising of booming for certain bass frequencies. The procedure for the measurement was as follows:

1. Setting up the PA system at a proper position.

2. Performing several measurements at different, randomized microphone positions. As test signals both music and sweeps were applied.

3. Equalization of the frequency responses using the automatic tool.

4. Evaluation of the quality of the equalization performed with a selection of adequate music tracks (e.g. Eagles, Michael Bublé, Opeth, Porcupine Tree, Sting).

The used PA system was a compact HK-AUDIO *L.U.C.A.S smart*. This system offers a total power of 1×250 W (subwoofer) and 2×80 W (tops) and is developed for small concerts and clubs up to an audience of about 100 listeners.
It was found to be difficult to benchmark whether the automatic equalization operates successful. First of all the music playback has to be done at high levels, so that enough energy is in the room. Preferably, music tracks should be chosen which not only feature a good production and stimulate the problematic resonance frequencies, but also are well-known to the listeners. Last but not least, individual preferences play an important role for the decision whether the sound quality is enhanced or not.

5. RESULTS

5.1. Example A: Conference Room

The first venue was a rectangular conference room of about 7 m × 7 m × 2.30 m, equipped with a carpet, a conference table and a dividing wall in the middle, see the sketch in Figure 11. The ceiling of this room was equipped with acoustical absorbing panels. The acoustical characteristics can be described as fair. The loudspeakers were placed in a realistic but not ideal position: The subwoofer was set up on the floor and the tops right above the table top. Five sweep measurements were done at different positions in the area around the desk. Two further measurements were performed with music as test signal leading to results in good conformance with the sweep measurements. The positions of PA system and microphone are marked in Figure 11. The automatic tool returned a number of 7 equalization filter sections in the range 24 to 90 Hz and one bass loss compensation filter at 45 Hz, see Table 1. All frequency responses are shown and discussed in the Figures 6 and 10.

The automated equalization of the first room resulted an improvement of the sound, even though the difference at some locations was small. The main resonances of the room were clearly suppressed, the sound was less boomy.

Table 1: *The filter parameters for the conference room.*

No.	f_m/Hz	Q	G/dB
1	24.3	16	-11.2
2	44.8	20	-15.8
3	51.0	20	-13.4
4	62.6	20	-12.3
5	69.1	20	-9.2
6	84.4	20	-8.6
7	89.0	20	-6.6
8	44.8	1	+7.6

5.2. Example B: Basement

The second room under test was an empty, rectangular room of about 4.5 m × 9 m and a ceiling height of 2.2 m in a basement. From different measurements this room was known to be acoustically complicated, featuring a long reverberation time and strong room modes due to its plain floor and walls. The PA system was set up at one side of the room in an useful way. Without filtering, the sound was booming especially for bass guitar and drum sounds. The frequency responses were determined with sweep signal and music measurements at three randomly chosen positions in the listening area. The curves showed some resonances at low frequencies. The automatic equalization delivered the placement of overall 14 filter sections up to 120 Hz. Figure 12 shows the maximum norm, the target curve, and the equalized result from these measurements.

The equalization was found to be successful. Because of the acoustically poor characteristics of this room, a really good sound was not achieved but an enhancement was audible.

5.3. Example C: Seminar Room

The third venue was a big seminar room of about 8 m × 10 m, fully equipped with tables and chairs. Parts of the ceiling (height 3.40 m) and the back wall of this venue were equipped with acoustical absorbing panels. The loudspeakers were arranged in a typical

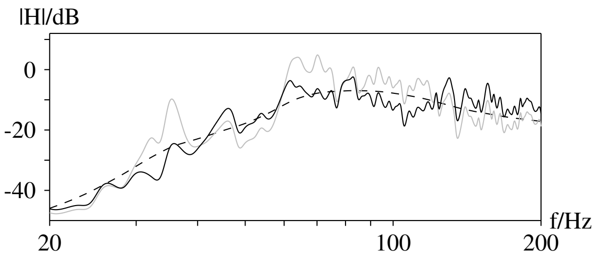

Figure 12: *H_{max} (gray), H_t (dashed), and equalized curve H_{eqd} (black) of the basement.*

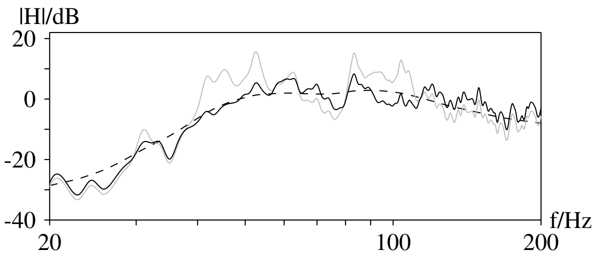

Figure 13: *H_{max} (gray), H_t (dashed), and equalized curve H_{eqd} (black) of the seminar room.*

setup with the tops placed symmetrically on the front wall and the subwoofer in the middle. The acoustic of this room offered a long reverberation time in combination with booming and vibrating at bass tones as common. Again five measurements were done with sweeps and music as test signal. The automatic tool placed 11 equalization filters in the low frequency region. One more filter section was added to compensate the bass loss. The frequency response computed with the maximum norm, the target curve and the frequency response after equalization are given in Figure 13.

The filtering resulted in an audible improvement of the sound. The room modes were clearly attenuated without loosing the bass fundament.

6. DISCUSSION

While the automatic equalizer benefited the listening experience, the listeners often also had ideas how to further improve the sound. An interesting issue is that the complete suppression of the resonances may sometimes lead to a lack of punchiness of the bass-drum. This may be caused by the bass loss due to the equalizer or to effects of the high-Q filters on the phase.

Whether the phase of the filters is audible is a matter of endless debate which we shall not get into at this point. Let us remark, however, that the maximum phase distortion of the equalizer filters is at their maximum attenuation and furthermore, as the filters are placed at room resonances, the room's phase distortion is probably much more severe.

Being caused by phase distortion or bass loss, the lack of punchiness could be reduced by doing less equalization. A pragmatic approach would be to let the user decide how much equalization he would desire with a knob ranging from 0 % to 100 %. This would also take into account that the most pleasing amount of equalization usually depends on the music material played, the playback level, and of course personal preference.

Another area for improvement is that sometimes a little boominess remained, indicating that some room resonances might have gone undetected. One possible reason is that all measurements were taken with the microphone in a node of the respective room mode. While possible, this is rather unlikely; in fact, three measurements are often sufficient to capture all relevant room characteristics [7].

More likely is that taking the maximum value across all measurements for a certain frequency and then searching for peaks in the resulting maximum spectrum is too crude. While this easily captures the most prominent resonances, weaker ones may remain hidden. An interesting alternative approach is to determine the per-frequency decay time using an energy decay relief [8], where resonances stand out with relatively long decay times, even when measured close to a node. However, this only gives an indication of the resonance frequency, not the amount of attenuation necessary to suppress any unpleasant boominess it might cause.

7. CONCLUSION

In this paper, a system for automatic room equalization especially focused on the suppression of room resonances was presented. The necessary measurements may be performed with music (or other suitable signals) as stimulus, allowing measuring with the audience present. The equalizing filter is then designed by placing sharp peak filters with negative gain on the identified resonance frequencies and compensating the resulting bass loss with a broader peak filter with positive gain.

To verify the approach, three rooms were equalized using the described method. In an informal listening test, the probands agreed that in all cases, the equalizer resulted in an improvement of the sound quality.

8. REFERENCES

[1] H. Herlufsen, "Dual channel FFT analysis (part I,II)," Tech. Rep., Brüel & Kjær Technical Review No.1, 1984, Available at http://www.bksv.com/pdf/Bv0013.pdf.

[2] S. Müller, "Measuring transfer-functions and impulse responses," in *Signal Processing in Acoustics*. 2008, vol. 1, pp. 74–76, D. Havelock, S. Kuwano and M. Vorländer (Eds.), Springer.

[3] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *108th AES Convention*, Paris, France, Feb. 19-24 2000.

[4] U. Zölzer, *Digital Audio Signal Processing*, John Wiley & Sons, Inc., New York, 1997.

[5] L.D. Fielder, "Practical limits for room equalization," in *108th AES Convention*, New York, NY, USA, Sep. 21-24 2001.

[6] P. Hatziantoniou and J. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc.*, vol. 48, no. 4, pp. 259 – 280, April 2000.

[7] J. A. Pedersen, "Loudspeaker-room adaptation for a specific listening position using information about the complete sound field," in *121st AES Convention*, 2006.

[8] J. M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, San Francisco, CA, USA, Mar. 23-26 1992, vol. 2, pp. 221–224.