Subband Based Room Impulse Response Reshaping

Radoslaw Mazur, Fabrice Katzberg, Martina Böhme, and Alfred Mertins[1]

Institute for Signal Processing, *University of Lübeck*, Ratzeburger Allee 160, 23562 Lübeck, Germany
Email: {mazur, katzberg, boehme, mertins}@isip.uni-luebeck.de

Abstract

In order to combat the degraded quality of sounds played in reverberant rooms, the methods of room impulse response equalization are used. New approaches utilize the properties of the human auditory system, such as temporal masking, for a better control of late echoes. There, a prefilter is used to modify the played signal, which renders the echoes inaudible for a given position. In case of spatial mismatch, the procedure fails and may even add additional distortions. In order to mitigate these effects, we propose to equalize subbands of the room impulse response independently. With heavy equalization of the low frequencies and only a little or none in the high frequencies, the mismatch and therefore the degradation can be reduced. The result is a bigger equalized volume where a potential human listener can move freely.

1 Introduction

When a sound is played in a closed room, the signal is reflected multiple times. Therefore, the listener receives a superposition of delayed and scaled versions of the original signal. This process can be described using the room impulse response (RIR). For a human listener, this reverberation is degrading the perceived quality. In order to combat these distortions, a prefilter can be used. This prefilter is designed in such a way that the global impulse response (GIR), the convolution of the RIR and the prefilter, has no audible distortions for a human listener [1].

The simplest approach is to design the prefilter in such a way that the GIR becomes a unit pulse, or more general, a bandpass [2, 3]. When optimizing using a quadratic criterion, the unwanted parts of the GIR are greatly reduced. Unfortunately, with this approach, the signal still contains audible echoes. However, by exploiting the properties of the human auditory system, an other approach is more feasible. The idea is not to remove the echoes completely, but rather render them inaudible for a human listener. Typically, the temporal masking curve [4] is used to describe the perceived reverberation. This relaxed approach has been very successful [5]. Additionally, the authors proposed to use a p-norm based criterion instead of the quadratic term as in [3], which allows for a better control of the late echoes. In [6], this approach has been extended with the aim of a more explicit control in the frequency domain.

In the case of a human listener, who is not able to keep completely still, these approaches are not sufficient. Even small movements change the RIR and the performance of the prefilter degrades [7]. To combat this problem, different approaches have been proposed. Generally, these can be sorted into two groups. The first idea is to use multiple positions [8, 9]. Knowing multiple RIRs in the listening area, the prefilters are designed in such a way that all these points are equalized. If these points are dense enough and fulfill the time-space sampling theorem [9], the whole volume becomes equalized. In order to get satisfactory results for bigger volumes, multiple speakers need to be used. Unfortunately, this MIMO approach is not feasible for real-world cases. It requires the measurement of all RIRs from all loudspeakers to all grid positions in the listening area, which is a very time consuming task. Recent advances are able to reduce this burden with the use of moving microphones [10]. The validity of this approach for RIR equalization has been shown in [11]. The second group of algorithms models additional errors in the optimization and adds regularizers [8, 12]. The method in [13] generates multiple hypothetical RIRs and is again able to extend the equalized volume. However, these extensions come at cost of a reduced performance at the center point. In [14] regularization was achieved by using short filters.

In this work we propose a different approach for equalizing a bigger volume. It is based on the properties of the frequency representation of the RIRs in the listening area [15]. Here, we employ the fact that for low frequencies the RIRs do not change significantly [7], and the equalization of one point leads to a whole volume being equalized. On the other hand, the high frequency parts do change significantly with spacial mismatch. Even small displacements lead to poor performance of the equalizer and may even lead to a result which is worse than doing nothing. Therefore, we propose to apply equalizers of different length, or even no equalizer, for different frequency components of the signal. The proposed algorithm is described for the single channel case, but can be extended to the multichannel case as in [8] using the same approach. Due to the limited space, we omit this discussion here.

The paper is organized as follows. In the next section, we will review the p-norm based equalization method. Then, in Section 3, the subband approach will be introduced. In Section 4, different filterbanks will be examined. In Section 5, the results for experiments for spatial mismatch will be shown. Finally, some conclusions will be given in the last section.

2 RIR Reshaping

In the following, we review the reshaping algorithm as proposed in [8]. The RIR $c(n)$ of length L_c from the loudspeaker to a position in the room has to be known. The equalizer $h(n)$ of length L_h leads to a GIR

$$g(n) = h(n) * c(n) \tag{1}$$

of length $L_g = L_c + L_h - 1$. The reshaping method from [8] uses two windows, $w_d(n)$ and $w_u(n)$, which define the desired and unwanted parts of the GIR. The desired part is obtained by $g_d(n) = w_d(n)g(n)$ and the undesired analogously by $g_u(n) = w_u(n)g(n)$. Here, the weighting windows are capturing the compromise temporal masking limit of the human auditory system [4, 5].

[1]This work has been supported by the German Research Foundation under Grants No. ME 1170/8-1 and ME 1170/10-1.

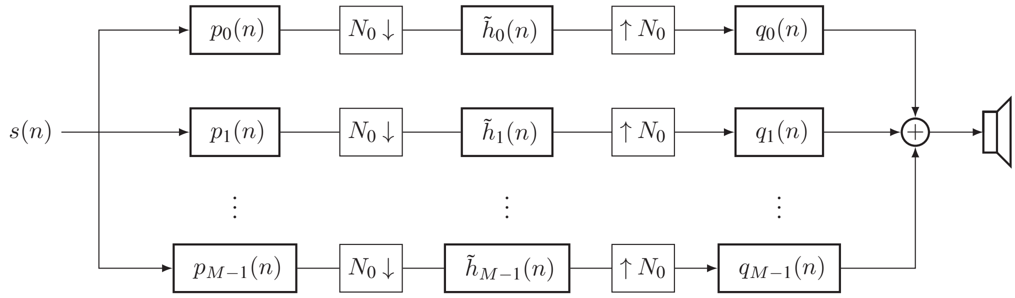

Figure 1: The subband model of the prefilter.

The prefilter can be estimated by optimizing

$$\mathrm{MIN}_h : f(\mathbf{h}) = \log_{10}\left(\frac{f_u(\mathbf{h})}{f_d(\mathbf{h})}\right) \tag{2}$$

with

$$f_d(\mathbf{h}) = ||\mathbf{g}_d||_{p_d} = \left(\sum_{k=0}^{L_g-1} |g_d(k)|^{p_d}\right)^{\frac{1}{p_d}} \tag{3}$$

and $f_u(\mathbf{h}) = ||\mathbf{g}_u||_{p_u}$, accordingly. The vectors $\mathbf{h}$, $\mathbf{g}_d$ and $\mathbf{g}_u$ contain the impulse response $h(n)$, the wanted and unwanted part, respectively. As there does not exist a closed form solution for (2), a gradient descent method needs to be used [8].

In order to achieve a smooth shaping with no outliers, the values for p_d and p_u need to be reasonably high (typically between 10 and 20). This leads to highly nonlinear cost function.

3 Subband Based Reshaping

In the case of spatial mismatch between the listener and the reference positions, the performance of the solution given by (1) degrades significantly. In [7] the average frequency-dependent error in the single channel case has been estimated as

$$F(\omega) = 2 - 2\frac{\sin(\omega\Delta/v)}{\omega\Delta/v} \tag{4}$$

with Δ being the displacement, v the speed of sound, and $\omega = 2\pi f$ the angular frequency. Equation (4) shows that for higher frequencies and displacements, the result may be even worse than no equalization at all. An example will be given in the experiments section.

In case of spatial mismatch, we propose to equalize the different frequency contents of the RIR with equalizers of different length, or no equalizers at all. This can be achieved, for example, by replacing the prefilter $h(n)$ in (1) by a filter bank shown in Fig. 1. This allows for estimating different length prefilters $\tilde{h}_m(n)$ in each subband independently, or even no prefiltering at all for higher frequencies.

In the next section, different filterbanks will be examined to see whether they are suitable for the decomposition of an RIR for equalizing. Since only a part of the RIR is being equalized in this case, the overall performance is expected to be lower compared to the full band case. The advantages of this approach are supposed to be seen mainly in the mismatch case, as it will be shown in Section 5.

4 Experiments

In this section we will introduce the baseline, i. e. the full bandwidth method. Then, different subband decomposition methods will be discussed and compared to the baseline. Here, we will focus on the basic properties – the spatial mismatch case will be discussed in the next section.

4.1 Full bandwidth

In Fig. 2 (a) a simulated RIR of length $L_c = 2000$ is shown in logarithmic representation [16]. The room size was set to $5 \times 4 \times 6$ meters, which represents a fairly big office room. The sampling frequency was chosen as $f_s = 8$kHz. With a reverberation time of $t_{60} = 400$ms there are a lot of audible echoes, as indicated by the coefficients above the temporal masking limit drawn in red. In Fig. 2 (b) the GIR $g(n) = c(n) * h(n)$ after reshaping using (2) is shown. The length of the equalizer was set to $L_h = 2000$. Here, the audible echoes are almost gone, as the coefficients are only slightly above the temporal masking limit.

In order to quantify the perceived reverberation we use the nPRQ measure from [8]. It calculates the overshot about the temporal masking curve being above the -60dB of the main peak $g_{\mathrm{os}}(n) = \max\left(\frac{1}{w_u(n)}, -60\text{dB}\right)$ as

$$\mathrm{nPRQ} = \begin{cases} \frac{1}{\|\boldsymbol{g}_E\|_0} \sum_{n=N_0}^{L_g-1} g_E(n), & \|\boldsymbol{g}_E\|_0 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

with

$$g_E(n) = \begin{cases} 20\log_{10}(|g(n)|w_u(n)), & |g(n)| > g_{\mathrm{os}}(n) \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

With all coefficients being below the temporal masking curve, indicating no reverberation, the nPRQ is equal to zero. Otherwise, higher values mean audible reverberation. In the above example, the original RIR $c(n)$ had an nPRQ value of 9.49 and the reshaped one could reduce it to 1.34. This improvement of ΔnPRQ $= -8.15$ is the maximum for this example and indicates a successful reshaping.

4.2 DCT Filterbank

The first decomposition is performed using a discrete cosine transform (DCT) filterbank: Windowing the RIR with a Hanning-window, taking the DCT-IV of the same length, and an optional subsampling. This yields a decomposition of the RIR, which has the same properties as a typical RIR, namely, a sharp peak denoting the direct sound, single smaller peaks for the early reflections, and an exponential

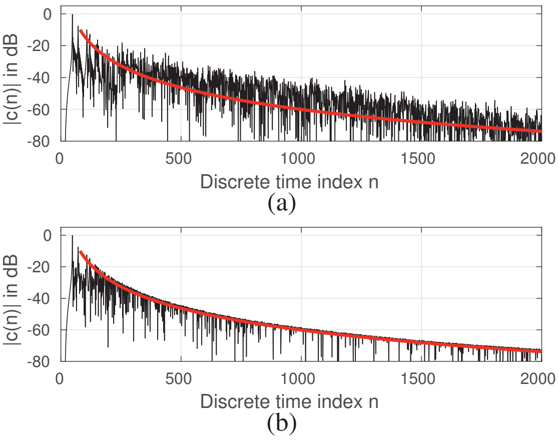

Figure 2: Equalization of an RIR. (a) The original RIR in logarithmic representation. Additionally, temporal masking limit is shown in red. (b) GIR after equalization with full bandwidth and no spacial mismatch.

decay of the late reverberant tail. Therefore, the equalization of the single subbands is usually as good as for the full bandwidth case and is comparable to the result of Fig 2.

In Fig. 3 the result after the synthesis is shown. Here, a two band decomposition has been performed and the subbands have been equalized independently. For the lower band, again, an equalizer of length 2000 has been used. For the upper band, lengths of 2000, 1000, and 1 (no equalizer) in (a), (b), and (c) have been used. The results for the nPRQ measurement are summarized in Table 1.

The results in Figure 3 (a) show that the independent equalization is almost as good as the full band approach. The result is not smooth though. This can be explained by the superposition of overlapping synthesis windows of the filterbank, with overlap in both time and frequency domain. Even when all coefficients in both subbands are below the temporal masking curve, the overlap leads to single peaks adding unevenly. The overall performance is still very good with an ΔnPRQ $= -8.12$. When the upper band is processed with a shorter equalizer the results are, as expected, worse. In case of no processing, the overall result with ΔnPRQ $= -1.11$ not satisfactory. In Table 1 the results for the four and eight channel filterbank are also shown. The general trend is that, with more subbands, the overall performance is getting worse. Again, this can be explained by the superposition effects. When applying subsampling in the subbands the situation is getting even worse, as aliasing effects dominate the result.

4.3 MLT Filterbank

In the previous section, the overlap in time and frequency domain of the synthesis windows has been identified as a major source of impairment after synthesis. Therefore, the next examined filter bank is the modulated lapped transform, which is again a DCT based filterbank, but with a prototype of double the length [17]. This leads to a better separation of the subbands.

In Table 2 the results are given. Here, we see the same general trend. With additional subbands and subsampling the performance deteriorates. On the other hand, in the two subband case, where only the lower band is equalized, there is an improvement with ΔnPRQ $= -3.44$. This case is promising for the equalization in the spatial mismatch case.

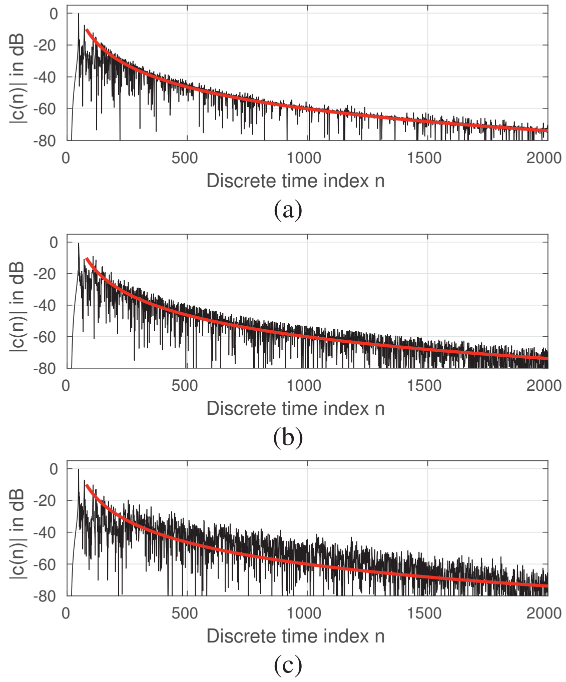

Figure 3: Subband equalization using a 2 channel dct filterbank. The lower band has been processed with an equalizer of length $L_h = 2000$. For the upper band equalizers of length 2000, 1000, and 1 (no equalization) in (a), (b), and (c) have been used.

Table 1: Comparison of the results for DCT based filterbank decomposition.

Type	Eq. length	ΔnPRQ
Reference	2000	-8.15
2-Band DCT-FB	2000 & 2000	-8.12
	2000 & 1000	-5.76
	2000 & 1	-1.11
4-Band DCT-FB	3x2000 & 2000	-6.80
	3x2000 & 1000	-6.04
	3x2000 & 1	-5.04
4-Band DCT-FB	3x1000 & 1000	-6.12
with 2x subsampling	3x1000 & 500	-5.47
	3x1000 & 1	-4.56
8-Band DCT-FB	7x2000 & 2000	-4.27
	7x2000 & 1000	-3.62
	7x2000 & 1	-3.21

Table 2: Comparison of the results for MLT filterbank.

Type	Eq. length	ΔnPRQ
Reference	2000	-8.15
2-Band MLT	2000 & 2000	-7.08
	2000 & 1000	-5.12
	2000 & 1	-3.44
4-Band MLT	3x2000 & 2000	-6.41
	3x2000 & 1000	-6.01
	3x2000 & 1	-5.39
4-Band MLT	3x1000 & 1000	-4.72
with 2x subsampling	3x1000 & 500	-4.09
	3x1000 & 1	-3.21

ISBN 978-3-8007-4767-2

Table 3: Comparison of the results for FIR filterbank.

Type	Eq. length	ΔnPRQ
Reference	2000	-8.15
2-Band FIR	2000 & 2000	-6.89
	2000 & 1000	-5.65
	2000 & 1	-4.16

A reason for the general lower performance compared to the DCT filterbank can be found in the prototype of the MLT filterbank. Here, the prototype is a cosine wave, which does not have a central location. Therefore, any single peaks of the RIR gets smeared out in the single subbands, and the shape does not look like a typical RIR. The cost function (3) is highly nonlinear and is not able to deal well with this representation.

4.4 FIR Filterbank

The last approach is to combine the properties of the previously discussed filterbanks. The design goal is to have as few subbands as possible, with as little as possible aliasing and no subsampling. On the other hand, the filters need to be located in order to avoid smearing out. This can be achieved for example with a filterbank which consists of a lowpass and a highpass with a given cutoff frequency f_c and a moderate length. In the following, we use a window of length 16. With no aliasing effects due to subsampling, the lowpass can be just a windowed sinc function with

$$h_{\text{LP}}(n) = \text{sinc}(nf_c)w(n) \tag{7}$$

where $w(n)$ is, for example, a Hamming window. The corresponding highpass can be derived accordingly [17]. The other big advantage of this approach is that the cutoff frequency can be chosen arbitrarily, for example, according to the size of the target volume.

In Table 3 the results are given. Compared to the previous filterbanks, the overall performance is the best. Especially, the performance for low and no equalization in the upper band is the best. This allows for a good equalization of a volume.

5 Spatial Robustness

In this section the spatial robustness of the equalization will be examined. In the last section, the simple FIR filterbank with two subbands has been identified being the most promising one. The approach here is to design an equalizer for a given point and estimate its performance in the vicinity of this point. In Fig. 4, a typical performance is shown. Here, a two-band filterbank with a cutoff frequency of $f_c = 0.6$, which translates to 2.4 kHz at 8 kHz sampling rate, has been used. The equalizer for the lower band had a length of 2000. The results for different equalizer lengths show a tradeoff. Especially, with no equalization in the higher band, the performance at the origin is clearly worse than the full bandwidth approach. On the other hand, already at a distance of less than 2 cm, the subband approach has the same performance. With higher distances the subband approach gives the better results.

In Fig. 5 the results for a whole area are shown. In (a) the performance for the full bandwidth is shown. The color coding indicates the improvement or deterioration in sense of nPRQ. As one can see, the center and a small volume around have been highly equalized. This comes at the price of a small volume and added reverberation outside. In (b) the subband approach has been used with equalizer length of 1000. The equalized volume has been increased. In (c) the upper band has not been equalized, which leads to a tradeoff. The center is not as good equalized, but the equalized volume is bigger. In this example the average radius of equalized volume, as indicated by the yellow color, could be doubled, which translates to an eightfold volume. In case of a human listener this approach allows for bigger movement of the head.

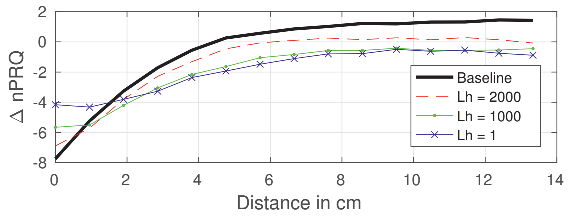

Figure 4: The equalization performance in case of spatial mismatch. With less equalization in the higher subband, the performance is lower at the origin, but allows for a bigger volume being equalized.

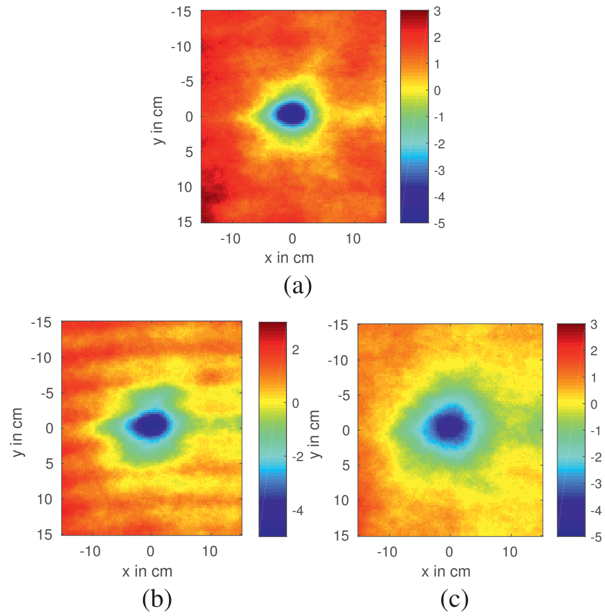

Figure 5: Equalization of a volume. In (a) a single RIR has been reshaped using full bandwidth. In (b) and (c) the new subband equalization method has been used (See text). The color codes the improvement/deterioration of the perceived echoes in terms of nPRQ. Blue and green mean improvement, yellow indicates no change and red colors show added reverberation.

6 Conclusion

In this paper we proposed a new subband based approach to room impulse response equalization. Three different subband decompositions have been examined. The approaches show a tradeoff between the performance at the origin and the equalized volume. In a typical setting, the equalized volume could be increased significantly.

References

[1] J. N. Mourjopoulos, "Digital equalization of room acoustics," *Journal of the Audio Engineering Society*, vol. 42, pp. 884–900, Nov. 1994.

[2] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *Journal of the Audio Engineering Society*, vol. 37, pp. 899–907, Nov. 1989.

[3] M. Kallinger and A. Mertins, "Room impulse response shortening by channel shortening concepts," in *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA , USA), pp. 898–902, Oct. 2005.

[4] L. D. Fielder, "Practical limits for room equalization," in *Proc. 111th Convention of the Audio Engineering Society*, pp. 1–19, Nov. 2001.

[5] A. Mertins, T. Mei, and M. Kallinger, "Room impulse response shortening/reshaping with infinity- and p-norm optimization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 249–259, Feb. 2010.

[6] J. O. Jungmann, S. Goetze, and A. Mertins, "Room impulse response reshaping by p-norm optimization based on estimates of room impulse responses," in *Proceedings of the German Annual Conference on Acoustics (DAGA)*, (Düsseldorf, Germany), pp. 611–612, Mar. 2011.

[7] B. D. Radlović, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 311–319, May 2000.

[8] J. O. Jungmann, R. Mazur, M. Kallinger, T. Mei, and A. Mertins, "Combined acoustic mimo channel crosstalk cancellation and room impulse response reshaping," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 1829–1842, Aug. 2012.

[9] T. Mei and A. Mertins, "On the robustness of room impulse response reshaping," in *Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control*, (Tel Aviv, Israel), Aug. 2010.

[10] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Sound-field measurement with moving microphones," *The Journal of the Acoustical Society of America*, vol. 141, pp. 3220–3235, May 2017.

[11] R. Mazur, F. Katzberg, H. Phan, and A. Mertins, "Room equalization based on measurements with moving microphones," in *Proc. Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, (San Francisco, USA), March 2017.

[12] J. O. Jungmann, R. Mazur, and A. Mertins, "A stochastic approach for robust listening room compensation," in *Proceedings of the German Annual Conference on Acoustics (DAGA)*, (Merano, Italy), pp. 185–186, Mar. 2013.

[13] J. O. Jungmann, R. Mazur, and A. Mertins, "Perturbation of room impulse responses and its application in robust listening room compensation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, (Vancouver, BC, Canada), pp. 433–437, May 2013.

[14] I. Kodrasi and S. Doclo, "Improving the conditioning of the optimization criterion in acoustic multi-channel equalization using shorter reshaping filters," in *EURASIP Journal on Advances in Signal Processing*, vol. 11, Feb 2018.

[15] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3790–3804, Oct. 2006.

[16] E. A. P. Habets, "Room impulse response generator." http://home.tiscali.nl/ehabets/rir_generator.html, Sept. 2010.

[17] P. Vaidyanathan, *Multirate Systems And Filter Banks*. Prentice Hall signal processing series, Pearson, 2004.

ISBN 978-3-8007-4767-2