

## OPTIMIZATION OF CASCADED PARAMETRIC PEAK AND SHELVING FILTERS WITH BACKPROPAGATION ALGORITHM

Purbaditya Bhattacharya, Patrick Nowak\* and Udo Zölzer

Department of Signal Processing and Communication  
 Helmut Schmidt University  
 Hamburg, Germany  
 bhattacp@hsu-hh.de

### ABSTRACT

Peak and shelving filters are parametric infinite impulse response filters which are used for amplifying or attenuating a certain frequency band. Shelving filters are parametrized by their cut-off frequency and gain, and peak filters by center frequency, bandwidth and gain. Such filters can be cascaded in order to perform audio processing tasks like equalization, spectral shaping and modelling of complex transfer functions. Such a filter cascade allows independent optimization of the mentioned parameters of each filter. For this purpose, a novel approach is proposed for deriving the necessary local gradients with respect to the control parameters and for applying the instantaneous backpropagation algorithm to deduce the gradient flow through a cascaded structure. Additionally, the performance of such a filter cascade adapted with the proposed method, is exhibited for head-related transfer function modelling, as an example application.

### 1. INTRODUCTION

An infinite impulse response (IIR) filter has a distinct advantage over a finite impulse response (FIR) filter for a given approximation problem. IIR filter transfer functions containing poles and zeros result in a much lower order compared to an equivalent FIR filter transfer function containing only zeros across the  $z$ -plane with all the poles at the origin. As a result, IIR filters have a considerably lesser number of coefficients to calculate and adapt, resulting in reduced computation. Additionally, parametrization of an IIR filter with control parameters like gain, bandwidth, and center / cut-off frequency provides the option to tune each parameter. The advantage of such parametric filter optimization is that the above parameters can be tuned independently while in case of non-parametric filters the entire set of filter coefficients needs to be recomputed. From an optimization perspective, experimentation with hyperparameters is relatively easier and more intuitive in case of the individual parameters compared to the set of coefficients. Hence, the design of such adaptive parametric filters is useful in multiple applications including equalizers for digital audio workstations, modelling the effects of instruments and loudspeakers, and approximating or controlling room acoustics.

Designs of first-order low-frequency and high-frequency shelving filters and second-order peak filters were presented in [1], [2]

\* Cooperation with Acoustics and Soldier Protection Group, French-German Research Institute of Saint-Louis, Saint-Louis, France

Copyright: © 2020 Purbaditya Bhattacharya et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

based on an all-pass decomposition. Second- and fourth-order parametric shelving filters were also designed in [3] and the design of higher order shelving filters was proposed in [4]. A comprehensive account about the parametric filters of different orders and their role in audio equalization has been described in [5]. However, the proposed filter structure consists of first-order shelving filters and second-order peak filters only and the filter cascade is employed in this work with the goal towards transfer function modelling as an example application. Hereby, the filter stages are sorted by frequency region of interest, thus the cascade begins with a low-frequency shelving filter followed by a series of peak filters and ends with a high-frequency shelving filter. The goal of the filter structure is to learn the parameters of the individual filters in order to match a given magnitude response. Earlier research in this area includes the work done in [6], which introduces a backpropagation based adaptive IIR filter. An approach to train a recursive filter with derivative function, for adapting a controller was introduced in a patent [7]. In the context of neural networks, a cascaded structure of FIR and IIR filters with multilayer perceptrons was used in [8] for time series modelling based on a simplified instantaneous backpropagation through time (IBPTT). A similar adaptive IIR- multilayer perceptron (IIR-MLP) network was described in [9] based on causal backpropagation through time (CBPTT). In recent years backpropagation is extensively used in convolutional and recurrent neural networks, the later being recursive in nature. However, in the aforementioned literatures the adaptation is primarily performed directly on the filter coefficients. To the best of our knowledge, optimization of equalizers with respect to the control parameters and with the help of backpropagation has not been extensively studied or illustrated.

Hence, this work contributes towards a novel approach where the necessary local gradients with respect to the control parameters are derived and the instantaneous backpropagation algorithm is applied to deduce the gradient flow through a cascaded structure of parametric filters. Finally, this method is applied in an example application of head-related transfer function (HRTF) modelling in order to show its effectiveness.

### 2. PARAMETRIC SHELVING AND PEAK FILTERS

The transfer function of a first-order shelving filter is given by

$$H(z) = 1 + \frac{H_0}{2} [1 \pm A_1(z)], \quad (1)$$

with

$$H_0 = V_0 - 1, \quad (2)$$

$$V_0 = 10^{\frac{G}{20}}, \quad (3)$$

and

$$A_1(z) = \frac{z^{-1} + a}{1 + az^{-1}}, \quad (4)$$

where  $G$  refers to the gain of the system in dB,  $a$  denotes the cut-off frequency parameter, and  $+/-$  refers to low-frequency shelving (LFS) and high-frequency shelving (HFS) filters, respectively. As seen in [10], using the same parameter  $a$  for positive and negative values of  $G$  results in an asymmetric magnitude response of the shelving filter. To make the magnitude response symmetric two different parameters  $a_B$  and  $a_C$  are used for the boost and cut cases, respectively. For an LFS these parameters are given by

$$a_B = \frac{\tan(\pi \frac{f_c}{f_s}) - 1}{\tan(\pi \frac{f_c}{f_s}) + 1} \quad (5)$$

for the boost case ( $G > 0$ ) and

$$a_C = \frac{\tan(\pi \frac{f_c}{f_s}) - V_0}{\tan(\pi \frac{f_c}{f_s}) + V_0} \quad (6)$$

for the cut case ( $G < 0$ ), where  $f_c$  denotes the cut-off frequency and  $f_s$  is the sampling frequency. To achieve a symmetric HFS, the same coefficient  $a_B$  from Eq. (5) can be used for the boost case, but the coefficient for the cut case has to be modified to

$$a_C = \frac{V_0 \tan(\pi \frac{f_c}{f_s}) - 1}{V_0 \tan(\pi \frac{f_c}{f_s}) + 1}. \quad (7)$$

The block diagram of a first-order shelving filter is shown in Fig. 1 and the corresponding difference equations can be given as

$$x_h(n) = x(n) - ax_h(n-1), \quad (8)$$

$$y_1(n) = ax_h(n) + x_h(n-1), \quad (9)$$

$$y(n) = \frac{H_0}{2} [x(n) \pm y_1(n)] + x(n). \quad (10)$$

In a similar way to that of shelving filters where a first-order low-pass or high-pass filter is added to a constant branch, a second-order peak filter can be obtained as the addition of a second-order band-pass filter to a constant branch. In  $z$ -domain the transfer function is given by

$$H(z) = 1 + \frac{H_0}{2} [1 - A_2(z)], \quad (11)$$

where  $A_2(z)$  denotes a second-order all-pass filter given by

$$A_2(z) = \frac{-a + d(1-a)z^{-1} + z^{-2}}{1 + d(1-a)z^{-1} - az^{-2}}. \quad (12)$$

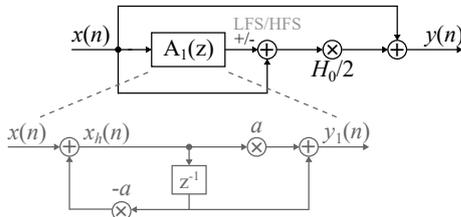


Figure 1: Block diagram of a first-order shelving filter, where  $+/-$  refers to low-frequency shelving (LFS) and high-frequency shelving (HFS), respectively.

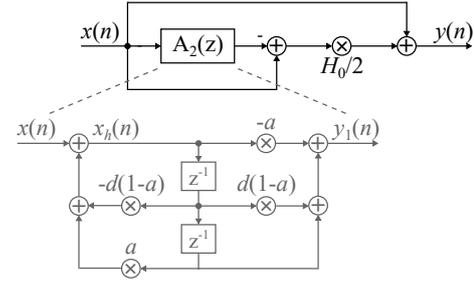


Figure 2: Block diagram of a second-order peak filter.

The block diagram of a second-order peak filter is shown in Fig. 2. Based on the block diagram the following difference equations can be derived:

$$x_h(n) = x(n) - d(1-a)x_h(n-1) + ax_h(n-2), \quad (13)$$

$$y_1(n) = -ax_h(n) + d(1-a)x_h(n-1) + x_h(n-2), \quad (14)$$

$$y(n) = \frac{H_0}{2} [x(n) - y_1(n)] + x(n). \quad (15)$$

The bandwidth related parameter  $a$  and the center frequency parameter  $d$  in the above equations can be calculated [10] as

$$d = -\cos(2\pi \frac{f_c}{f_s}), \quad (16)$$

$$a_B = \frac{\tan(\pi \frac{f_b}{f_s}) - 1}{\tan(\pi \frac{f_b}{f_s}) + 1}, \quad (17)$$

$$a_C = \frac{\tan(\pi \frac{f_b}{f_s}) - V_0}{\tan(\pi \frac{f_b}{f_s}) + V_0}, \quad (18)$$

where  $f_c$  denotes the center frequency and  $f_b$  denotes the bandwidth. Additionally,  $a_B$  and  $a_C$  denote the parameter  $a$  for boost and cut case, respectively.

In order to create an equalizer which can be controlled to shape the magnitude spectrum,  $M$  parametric filters can be cascaded. Here, the spectral regions can be controlled by tuning the gain, bandwidth, cut-off, and center frequencies. The parameters of the individual filters can be optimized to attain a desired frequency response. The block diagram provided in Fig. 3 illustrates the cascaded structure having a low-frequency shelving filter followed by  $M - 2$  peak filters and a high-frequency shelving filter. Additionally, exemplary magnitude responses with different gains  $G$  are shown for each filter stage. The predicted response  $y(n)$  is compared with a desired response  $y_d(n)$ , by a pre-defined cost function or loss function as illustrated in the block diagram in Fig. 4.

### 3. BACKPROPAGATION

In any supervised learning method a cost function between the prediction and the desired signal is necessary in order to calculate the gradient of such a function with respect to the control parameters. The negative gradient is used in a steepest descent algorithm in order to update and adapt the parameters with a goal towards optimizing the cost function. The partial derivative of a pre-defined cost function with respect to the filter parameters in a cascaded structure is not straightforward and needs to be calculated as a product of multiple local derivatives according to the chain rule.

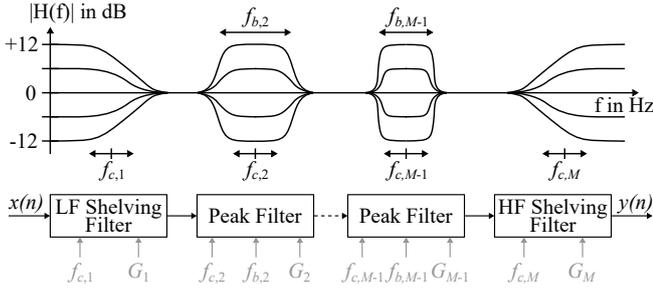


Figure 3: Block diagram of a parametric shelving and peak filter cascade along with exemplary magnitude responses of variable gain  $G$  for every filter stage.

This results in the backpropagation algorithm. Given a cascaded structure with  $M$  filters as illustrated in Fig. 3, and a global instantaneous cost or loss function  $C(n)$  for the  $n^{\text{th}}$  sample, as shown in Fig. 4, its derivative w.r.t. a parameter  $p_{M-1}$  can be written as

$$\frac{\partial C(n)}{\partial p_{M-1}} = \frac{\partial C(n)}{\partial y(n)} \frac{\partial y(n)}{\partial x_{M-1}(n)} \frac{\partial x_{M-1}(n)}{\partial p_{M-1}}, \quad (19)$$

according to the chain rule of derivatives, as illustrated in Fig. 5, where  $y(n)$  represents the predicted output of the cascaded filter structure,  $y_d(n)$  represents the desired output of the system under test,  $x_{M-1}(n)$  represents the output of the  $\{M-1\}^{\text{th}}$  filter in the cascade, and  $p_{M-1}$  represents any control parameter like gain, bandwidth, or center frequency of the  $\{M-1\}^{\text{th}}$  peak filter. This will result in a simplified instantaneous backpropagation algorithm [8]. In order to calculate the above derivative w.r.t. the instantaneous cost function it is necessary to calculate the local derivative  $\frac{\partial C(n)}{\partial y(n)}$  of the cost function, the local derivative  $\frac{\partial y(n)}{\partial x_{M-1}(n)}$  of a filter output w.r.t. its input, and the local derivative  $\frac{\partial x_{M-1}(n)}{\partial p_{M-1}}$  of a filter output w.r.t. its parameter  $p_{M-1}$ . Hence, in general the above three types of local gradients need to be calculated in order to adapt the cascaded structure. After the gradient  $\frac{\partial C(n)}{\partial p_{M-1}}$  is calculated the corresponding parameter can be updated according to gradient descent algorithm as

$$p_{M-1} := p_{M-1} - \eta \frac{\partial C(n)}{\partial p_{M-1}}, \quad (20)$$

where  $\eta$  is referred to as step-size or learning rate. If the instantaneous cost function is assumed to be a squared-error function then the expression is given by

$$C(n) = [y_d(n) - y(n)]^2, \quad (21)$$

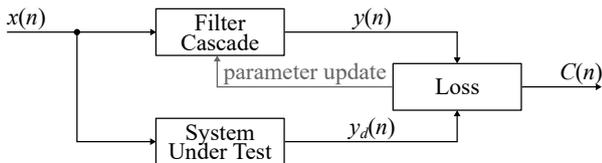


Figure 4: Block diagram of the model with predicted output  $y(n)$  and desired output  $y_d(n)$ .

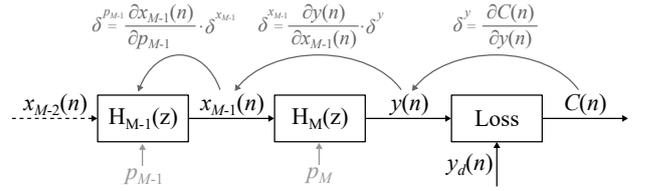


Figure 5: Block diagram illustrating the flow of the instantaneous local gradients during backpropagation.

and its local derivative w.r.t.  $y(n)$  is given by

$$\frac{\partial C(n)}{\partial y(n)} = -2 [y_d(n) - y(n)]. \quad (22)$$

In the following sections the local gradients of the filter output against the filter input and the control parameters are derived.

### 3.1. Shelving Filter

For a shelving filter, local derivatives of the filter output are calculated against the filter input, the gain and the cut-off frequency. Referring to Eq. (10) based on Fig. 1 the derivative of an LFS and an HFS output  $y(n)$  w.r.t. to its input  $x(n)$  is calculated as

$$\frac{\partial y(n)}{\partial x(n)} = \frac{H_0}{2} [1 \pm a] + 1. \quad (23)$$

The derivative of the shelving filter output w.r.t. the filter gain  $G$ , for the boost case is calculated as

$$\frac{\partial y(n)}{\partial G} = \frac{[x(n) \pm y_1(n)]}{2} \frac{\partial H_0}{\partial G}. \quad (24)$$

Given Eq. (2) and Eq. (3), the above derivation can be continued as

$$\frac{\partial y(n)}{\partial G} = \frac{[x(n) \pm y_1(n)]}{2} \frac{\partial}{\partial G} [10^{\frac{G}{20}} - 1], \quad (25)$$

$$= \frac{[x(n) \pm y_1(n)]}{40} 10^{\frac{G}{20}} \ln(10). \quad (26)$$

The derivative of the filter output w.r.t. the filter gain  $G$ , for the cut case is different from the boost case because of the dependence between the gain parameter and the parameter for cut-off frequency. With the help of Eq. (10) it can be calculated as

$$\frac{\partial y(n)}{\partial G} = \frac{[x(n) \pm y_1(n)]}{2} \frac{\partial H_0}{\partial G} \pm \frac{H_0}{2} \frac{\partial y_1(n)}{\partial G}, \quad (27)$$

$$= \frac{[x(n) \pm y_1(n)]}{40} 10^{\frac{G}{20}} \ln(10) \pm \frac{H_0}{2} \frac{\partial y_1(n)}{\partial G}, \quad (28)$$

where the expression  $\frac{\partial y_1(n)}{\partial G}$  of Eq. (28) can be extended with the help of Eq. (9) as

$$\frac{\partial y_1(n)}{\partial G} = \frac{\partial a_C}{\partial G} x_h(n) + a_C \frac{\partial x_h(n)}{\partial G} + \frac{\partial x_h(n-1)}{\partial G}, \quad (29)$$

with

$$\frac{\partial a_C}{\partial G} = \frac{-\ln(10)V_0 \tan(\pi \frac{f_c}{f_s})}{10 [\tan(\pi \frac{f_c}{f_s}) + V_0]^2}, \quad \text{(LFS)} \quad (30)$$

$$\frac{\partial a_C}{\partial G} = \frac{\ln(10)V_0 \tan(\pi \frac{f_c}{f_s})}{10 [V_0 \tan(\pi \frac{f_c}{f_s}) + 1]^2}, \quad \text{(HFS)} \quad (31)$$

$$\frac{\partial x_h(n)}{\partial G} = -\frac{\partial a_C}{\partial G} x_h(n-1) - a_C \frac{\partial x_h(n-1)}{\partial G}. \quad (32)$$

From Eq. (32),  $\frac{\partial x_h(n-1)}{\partial G}$  is calculated via IBPTT with the initialization of  $\frac{\partial x_h(k)}{\partial G}|_{k=0} = 0$ .

Finally, the derivative of the shelving filter output w.r.t. the cut-off frequency  $f_c$ , is calculated as

$$\frac{\partial y(n)}{\partial f_c} = \pm \frac{H_0}{2} \frac{\partial y_1(n)}{\partial f_c}, \quad (33)$$

where

$$\frac{\partial y_1(n)}{\partial f_c} = \frac{\partial a}{\partial f_c} x_h(n) + a \frac{\partial x_h(n)}{\partial f_c} + \frac{\partial x_h(n-1)}{\partial f_c}, \quad (34)$$

with

$$\frac{\partial a_B}{\partial f_c} = \frac{2\pi \sec(2\pi \frac{f_c}{f_s}) \left[ \sec(2\pi \frac{f_c}{f_s}) - \tan(2\pi \frac{f_c}{f_s}) \right]}{f_s}, \quad (35)$$

$$\frac{\partial a_C}{\partial f_c} = \frac{2\pi V_0 \sec(\pi \frac{f_c}{f_s})^2}{f_s \left[ \tan(\pi \frac{f_c}{f_s}) + V_0 \right]^2}, \quad (\text{LFS}) \quad (36)$$

$$\frac{\partial a_C}{\partial f_c} = \frac{2\pi V_0 \sec(\pi \frac{f_c}{f_s})^2}{f_s \left[ V_0 \tan(\pi \frac{f_c}{f_s}) + 1 \right]^2}, \quad (\text{HFS}) \quad (37)$$

$$\frac{\partial x_h(n)}{\partial f_c} = -\frac{\partial a}{\partial f_c} x_h(n-1) - a \frac{\partial x_h(n-1)}{\partial f_c}. \quad (38)$$

From Eq. (38),  $\frac{\partial x_h(n-1)}{\partial f_c}$  is calculated via IBPTT with the initialization of  $\frac{\partial x_h(k)}{\partial f_c}|_{k=0} = 0$ .

### 3.2. Peak Filter

For a peak filter, local derivatives of the filter output are calculated against the filter input, the gain, the center frequency, and the bandwidth. Referring to Eq. (15) based on Fig. 2 the derivative of a second-order peak filter output  $y(n)$  w.r.t. to its input  $x(n)$  is calculated as

$$\frac{\partial y(n)}{\partial x(n)} = \frac{H_0}{2} [1 + a_{B/C}] + 1. \quad (39)$$

As done in the case of shelving filters, the derivative of the peak filter output w.r.t. the filter gain  $G$ , for the boost case will result in

$$\frac{\partial y(n)}{\partial G} = \frac{[x(n) - y_1(n)]}{40} 10^{\frac{G}{20}} \ln(10). \quad (40)$$

For the cut case as well the derivative of the peak filter output w.r.t. the filter gain  $G$ , is derived in a similar manner. With the help of Eq. (15) the derivation leads to

$$\frac{\partial y(n)}{\partial G} = \frac{[x(n) - y_1(n)]}{40} 10^{\frac{G}{20}} \ln(10) - \frac{H_0}{2} \frac{\partial y_1(n)}{\partial G}, \quad (41)$$

and the expression  $\frac{\partial y_1(n)}{\partial G}$  from the above equation can be extended with the help of Eq. (14) as

$$\begin{aligned} \frac{\partial y_1(n)}{\partial G} &= -\frac{\partial a_C}{\partial G} x_h(n) - a_C \frac{\partial x_h(n)}{\partial G} - \dots \\ & d \frac{\partial a_C}{\partial G} x_h(n-1) + \frac{\partial x_h(n-2)}{\partial G} + \dots \\ & d(1-a_C) \frac{\partial x_h(n-1)}{\partial G}, \end{aligned} \quad (42)$$

with

$$\frac{\partial a_C}{\partial G} = \frac{-\ln(10)V_0 \tan(\pi \frac{f_b}{f_s})}{10 \left[ \tan(\pi \frac{f_b}{f_s}) + V_0 \right]^2}, \quad (43)$$

and

$$\begin{aligned} \frac{\partial x_h(n)}{\partial G} &= d \frac{\partial a_C}{\partial G} x_h(n-1) + \frac{\partial a_C}{\partial G} x_h(n-2) + \dots \\ & a_C \frac{\partial x_h(n-2)}{\partial G} - d(1-a_C) \frac{\partial x_h(n-1)}{\partial G}. \end{aligned} \quad (44)$$

From Eq. (44), the expressions  $\frac{\partial x_h(n-1)}{\partial G}$  and  $\frac{\partial x_h(n-2)}{\partial G}$  are calculated via IBPTT with the initializations of  $\frac{\partial x_h(k)}{\partial G}|_{k=0} = 0$  and  $\frac{\partial x_h(k)}{\partial G}|_{k=-1} = 0$

The derivative of the peak filter output w.r.t. the cut-off frequency  $f_c$ , leads to the expression similar to Eq. (33) given by

$$\frac{\partial y(n)}{\partial f_c} = -\frac{H_0}{2} \frac{\partial y_1(n)}{\partial f_c}, \quad (45)$$

where

$$\begin{aligned} \frac{\partial y_1(n)}{\partial f_c} &= -a \frac{\partial x_h(n)}{\partial f_c} + \frac{\partial d}{\partial f_c} (1-a)x_h(n-1) + \dots \\ & d(1-a) \frac{\partial x_h(n-1)}{\partial f_c} + \frac{\partial x_h(n-2)}{\partial f_c}, \end{aligned} \quad (46)$$

with

$$\frac{\partial d}{\partial f_c} = \frac{2\pi \sin(2\pi \frac{f_c}{f_s})}{f_s}, \quad (47)$$

and

$$\begin{aligned} \frac{\partial x_h(n)}{\partial f_c} &= -\frac{\partial d}{\partial f_c} (1-a)x_h(n-1) - \dots \\ & d(1-a) \frac{\partial x_h(n-1)}{\partial f_c} + a \frac{\partial x_h(n-2)}{\partial f_c}. \end{aligned} \quad (48)$$

From Eq. (48), the expressions  $\frac{\partial x_h(n-1)}{\partial f_c}$  and  $\frac{\partial x_h(n-2)}{\partial f_c}$  are calculated via IBPTT.

Finally, the derivative of the peak filter output w.r.t. the bandwidth  $f_b$ , is calculated as

$$\frac{\partial y(n)}{\partial f_b} = -\frac{H_0}{2} \frac{\partial y_1(n)}{\partial f_b}. \quad (49)$$

From the above equation  $\frac{\partial y_1(n)}{\partial f_b}$  is calculated as

$$\begin{aligned} \frac{\partial y_1(n)}{\partial f_b} &= -\frac{\partial a}{\partial f_b} x_h(n) - a \frac{\partial x_h(n)}{\partial f_b} - d \frac{\partial a}{\partial f_b} x_h(n-1) + \dots \\ & \frac{\partial x_h(n-2)}{\partial f_b} + d(1-a) \frac{\partial x_h(n-1)}{\partial f_b}, \end{aligned} \quad (50)$$

with

$$\frac{\partial a_B}{\partial f_b} = \frac{2\pi \sec(2\pi \frac{f_b}{f_s}) \left[ \sec(2\pi \frac{f_b}{f_s}) - \tan(2\pi \frac{f_b}{f_s}) \right]}{f_s}, \quad (51)$$

$$\frac{\partial a_C}{\partial f_b} = \frac{2\pi V_0 \sec(\pi \frac{f_b}{f_s})^2}{f_s \left[ \tan(\pi \frac{f_b}{f_s}) + V_0 \right]^2}, \quad (52)$$

and

$$\begin{aligned} \frac{\partial x_h(n)}{\partial f_b} = & d \frac{\partial a}{\partial f_b} x_h(n-1) - d(1-a) \frac{\partial x_h(n-1)}{\partial f_b} + \dots \\ & \frac{\partial a}{\partial f_b} x_h(n-2) + a \frac{\partial x_h(n-2)}{\partial f_b}. \end{aligned} \quad (53)$$

From Eq. (53), the expressions  $\frac{\partial x_h(n-1)}{\partial f_b}$  and  $\frac{\partial x_h(n-2)}{\partial f_b}$  are calculated via IBPTT.

Given the derivations, as an example, the derivative of the cost function provided in Eq. (21) w.r.t. the gain of the  $\{M-1\}^{th}$  peak filter ( $G_{M-1}$ ) in the filter cascade, and illustrated in Fig. 5, is given by

$$\frac{\partial C(n)}{\partial G_{M-1}} = \frac{\partial C(n)}{\partial y(n)} \frac{\partial y(n)}{\partial x_{M-1}(n)} \frac{\partial x_{M-1}(n)}{\partial G_{M-1}}, \quad (54)$$

according to Eq. (19). If the boost case is assumed then with the help of the equations (22), (39), and (40), respectively, the required gradient for parameter update is derived as

$$\begin{aligned} \frac{\partial C(n)}{\partial G_{M-1}} = & e(n) \left( \frac{H_0}{2} [1 + a_{B_{M-1}}] + 1 \right) \dots \\ & \dots (K [x_{M-2}(n) - y_{1_{M-1}}(n)]), \end{aligned} \quad (55)$$

where

$$e(n) = -2 [y_d(n) - y(n)], \quad (56)$$

$$K = \frac{10^{\frac{G_{M-1}}{20}} \ln(10)}{40} = \frac{V_{0_{M-1}} \ln(10)}{40}. \quad (57)$$

#### 4. HEAD-RELATED TRANSFER FUNCTION MODELLING

A head-related transfer function (HRTF) is a direction dependent transfer function between an external sound source and the entrance of the human ear canal. Thus, HRTFs account for reflections and diffractions at the human's head, torso, and pinna. These effects result in peaks and notches inside the HRTFs, which take part in the process humans use for vertical sound localization. Additionally, the interaural level difference (ILD), which is a main cue for horizontal sound localization, can be seen by comparing the HRTF of the left and right ear. Because of differences in the size and shape of human bodies, HRTFs are highly individual. The inverse Fourier transform of the HRTF is the head-related impulse response (HRIR). In spatial audio through headphones, mono signals are filtered with the corresponding HRTFs to create a virtual sound that is localized in a certain direction. In order to achieve a good resolution of the 3D space, HRTFs have to be saved for a high number of directions, resulting in a high amount of stored data. Hence, parametric IIR filters can be used to model the individual HRTFs with a lower number of saved parameters.

In [11], HRTFs were modelled with a cascade of second-order IIR filters. A localization test in the horizontal plane has shown that four to seven second-order IIR filters achieve similar results to the original FIR implementation. Furthermore, in [12] a cascade of a second-order low-frequency shelving filter and multiple peak filters was proposed for the task of HRTF modelling. There, a parameter initialization based on the error area between the model and the target, and a random search based optimization are used in order to tune the individual filters. This tuning is repeated for a

given number of peak filters and finished by post-processing triples of neighbouring filters in order to improve their interaction. In [13], this cascade of parametric filters is converted into a parallel structure of low-pass and band-pass filters. Moreover, in [14], a cascade of shelving and peak filters is used to find the minimum number of peak filters needed to model HRTFs with a given error tolerance. For this, a consecutive adding of peak filters including the initialization based on the maximum modelling error and a parameter optimization of the cascade via the Levenberg-Marquardt algorithm are used. However, the application of our proposed method to this problem is a new approach and unlike the aforementioned approaches, we adopt a simpler initialization in order to focus on the effectiveness of our method.

#### 4.1. Filter Cascade Initialization

For the initialization of the cascaded filter structure, an approximated number of required peak filters needs to be determined. Afterwards, the allocation of the initial parameter values is done. A structured initialization is very important in this case because a random initialization will make the problem extremely ill-posed and will create a random input response for the cascade with a poor correlation to the desired response. The entire process can be described by the following steps:

1. The magnitude response is smoothed, the mean of the magnitude response is subtracted, and the peaks and notches in the magnitude response are noted.
2. Certain peaks or notches are deleted based on a peak prominence threshold in dB, as defined in Matlab's peak analysis method 'findpeaks', their proximity, and magnitude differences between adjacent peaks / notches.
3. The initial center frequency of a peak filter is initialized based on the position of a peak or notch while the initial cut-off frequencies of the shelving filters depend on the slopes in the magnitude response.
4. The gain of a filter is initialized by the magnitude of the transfer function at the position of the peak or notch. In order to reduce the summation effect due to the cascade, the gain of every peak filter is scaled by a fractional factor.
5. If a notch exists in the positive half-plane, the gain is converted to a small negative random value, in order to maintain the information of being a notch. In the same way, if a peak is located in the negative half-plane, it's gain is converted to a small positive random value.
6. The bandwidth of a peak filter is initialized based on the average local gradient of the magnitude response around the peak's position.

Figure 6 shows the desired magnitude response and the initial magnitude response based on the above initialization. Here, 19 peak filters are selected based on a minimum peak prominence threshold of 0.01 dB, and a threshold of 200 Hz for the proposed peak / notch proximity along with 2 dB as magnitude difference threshold. The number of filters can be reduced or increased based on the third step. A major drawback of such an initialization is that more than an optimal number of filters will usually be proposed. Additionally the peak picking method will be insufficient in flat regions. However, the direct initialization is simple, sometimes quite close to the desired, and the simultaneous filter update improves the run-time.

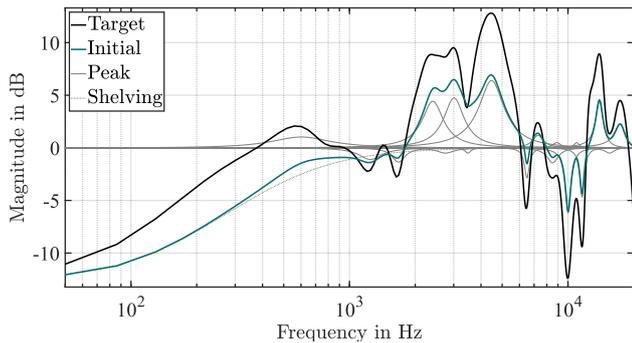


Figure 6: Magnitude responses of the desired HRTF and the initial HRTF estimate of the right ear of 'Subject-009' of the CIPIC database [15], for an azimuth  $\phi = 10^\circ$  and elevation  $\theta = 0^\circ$ .

## 4.2. Objective Function

The cascaded structure represents a minimum phase system and offers no delay compensation. Hence, a time domain loss function will be unable to approximate the desired measured impulse response and more importantly the magnitude response. Since the goal of the application is to approximate the magnitude response accurately, a loss function directly in frequency domain should be used. To adapt the cascaded structure, a log-spectral distance between the estimated and the desired magnitude responses is considered as the objective function which is given by,

$$C(Y_d(k), Y(k)) = [Y_d(k) - Y(k)]^2, \quad (58)$$

where  $Y_d(k)$  and  $Y(k)$  are the magnitude responses of the desired output  $y_d(n)$  and the estimated output  $y(n)$  in decibels, and  $k$  denotes a frequency bin. The derivative of the Fourier transform between time and frequency domain can be performed with the help of Wirtinger calculus as demonstrated in [16] and [17].

## 4.3. Results and Discussion

In this work, some HRIR examples from the benchmark CIPIC database described in [15] are chosen for the HRTF approximation. It contains HRIRs from 45 subjects. The length of the stored HRIRs is 200 samples. For every subject 1250 directions in the interaural-polar coordinate system are measured, including 25 different azimuths and 50 different elevations. In the first step, the HRIRs are converted to HRTFs and the magnitude responses are treated as the desired signal for the filter cascade. Nevertheless, before performing the discrete Fourier transform, the HRIRs are padded with zeros to a length of 1024 in order to achieve a better frequency resolution. Afterwards, the initialization described in Section 4.1 is performed. Due to differences in the HRTFs between subjects and directions, every transfer function needs a unique initialization, which can result in a different number of peak filters and initial parameter values. After the initialization of the cascaded structure, the filters are trained and updated for 100 epochs with the adam method [18], where each epoch has 1024 iterations or recurrent steps. However, in cases of many HRTFs, a smaller number of epochs is sufficient for convergence. The learning rate during the update method is selected as  $\eta = 10^{-1}$ . Additionally, there is a learning rate drop factor of 0.99 for every time the error in an epoch is higher than the error in the previous epoch. The aforementioned hyper-parameters might change

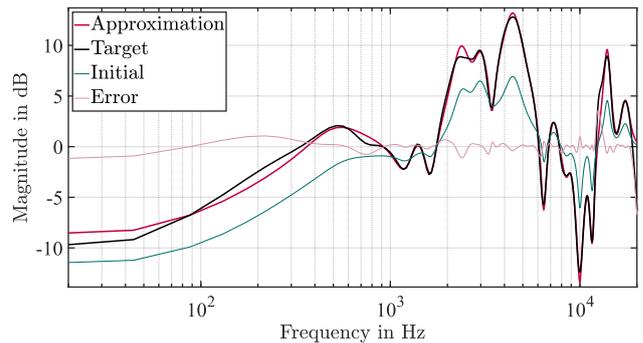


Figure 7: Magnitude responses of the desired HRTF, the initial HRTF estimate, and the final approximation of the right ear of 'Subject-009', for an azimuth  $\phi = 10^\circ$  and elevation  $\theta = 0^\circ$ .

for some cases where convergence requires different values. The entire implementation is done in Matlab. Figure 7 shows the desired magnitude response, the initial magnitude response, and the predicted response after 100 epochs corresponding to the right ear of 'Subject-009'. It can be seen that the cascaded structure has approximated the desired or target response quite well. When the minimum peak prominence threshold is increased and the other constraints are less relaxed, a lesser number of filters is usually found during initialization. Figure 8 shows the initial, desired, final approximated, and error magnitude responses corresponding to the right ear of 'Subject-008' for an azimuth angle of  $\phi = 80^\circ$  and an elevation angle of  $\theta = 0^\circ$ . With a minimum peak prominence threshold of 1 dB, eleven peak filters are used resulting in a maximum absolute error of about 4.2 dB, and a mean absolute spectral distance (MASD) of nearly 0.65 dB, calculated below 20 kHz. Lesser number of filters result in a shallower cascade and hence a faster run-time. On the other hand this could also lead to missing peak proposals and increased average approximation error. The corresponding filters of the cascaded structure are shown in Fig. 9. Here it can be seen that the gains of all filters are adapted during the update while the bandwidths and the center / cut-off frequencies are notably changed for some of these filters. Figure 10 shows the different magnitude responses corresponding to the right ear of 'Subject-008' for an azimuth angle of  $\phi = 0^\circ$  and an elevation angle of  $\theta = -22.5^\circ$  given a minimum peak prominence threshold

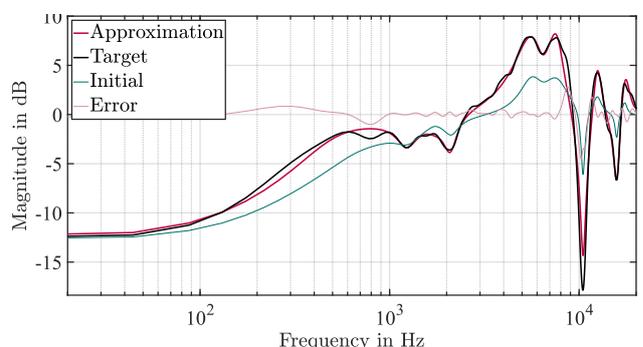


Figure 8: Magnitude responses of the desired HRTF, the initial HRTF estimate and the final approximation of the right ear of 'Subject-008', for an azimuth  $\phi = 80^\circ$  and elevation  $\theta = 0^\circ$ .

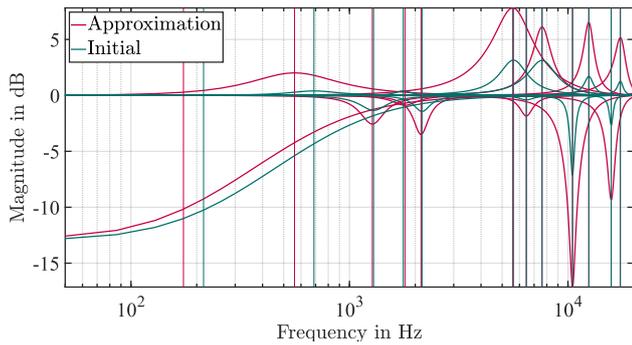


Figure 9: Responses of the individual filters after initialization and after HRTF approximation of the right ear of 'Subject-008', for an azimuth  $\phi = 80^\circ$  and elevation  $\theta = 0^\circ$ . The vertical lines indicate the positions of center / cut-off frequencies.

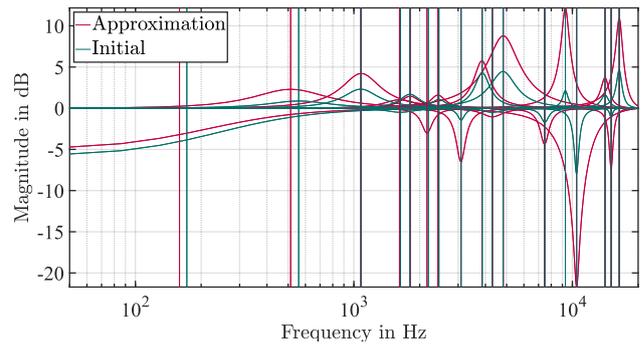


Figure 11: Responses of the individual filters after initialization and after HRTF approximation of the right ear of 'Subject-008', for an azimuth  $\phi = 0^\circ$  and elevation  $\theta = -22.5^\circ$ .

of 0.05 dB. The error response shows a maximum absolute error of about 2.5 dB and a MASD of nearly 0.45 dB. Additionally, Fig. 11 shows the individual filter responses of the 16 peak filters and the shelving filters before and after the approximation. Similar observations can be made for the left ear of the same subject for a different elevation. Figure 12 shows the corresponding magnitude responses of the left ear of 'Subject-008' for an azimuth angle of  $\phi = 0^\circ$  and an elevation angle of  $\theta = -45^\circ$ . The error response shows a maximum absolute error of nearly 2 dB and a MASD of nearly 0.3 dB. Figure 13 shows the individual filter responses of the 18 peak filters and the shelving filters before and after the approximation.

The first ten subjects are selected and simulations are performed for seven elevation angles between  $\theta = -45^\circ$  and  $\theta = 225^\circ$  with a step size of  $45^\circ$ , and for seven azimuth angles of  $\phi = -80^\circ, -55^\circ, -20^\circ, 0^\circ, 20^\circ, 55^\circ, 80^\circ$ , to evaluate an average performance of the proposed method in terms of its error minimization. The corresponding filter initialization is done with a low minimum peak prominence threshold of 0.005 dB, a separation threshold of 300 Hz, and a magnitude difference threshold of 3 dB. Table 1 displays the average performance of the method based on the above simulation, for both ears. About 72% of the examples show very good approximation with each less than 1 dB MASD, while some bad examples show difficulties in converging,

a few even having a maximum absolute spectral distance as high as 13 dB. This behavior is primarily attributed to the lack of proposals in the flat spectral regions, very steep and deep notches or large peaks in the high frequency region, and an occasional slow convergence of the high-frequency shelving filter resulting in higher errors above 18 kHz. The lowest final MASD in the experiment is about 0.08 dB while the highest final MASD is about 3.84 dB. It is also noted that the rate of change in gain and bandwidth is more predominant than the rate of change in the center or cut-off frequencies. This outcome results from the initialization, where the center or cut-off frequencies are more likely to be close to their local optimum solutions compared to the other parameters.

Table 1: Average performance based on 10 subjects and 49 directions, in terms of average MASD reduction below 20 kHz and average number of peak filters for both ears.

Ear	Init. MASD (dB)	Fin. MASD (dB)	Peak filters
Left	3.75	0.83	18
Right	3.58	0.77	18

As a future endeavor, a better initialization method and smoothing technique can be used with the goal to reduce the number of required filters, from the perspective of this particular application. This can be achieved with the help of incremental modelling of

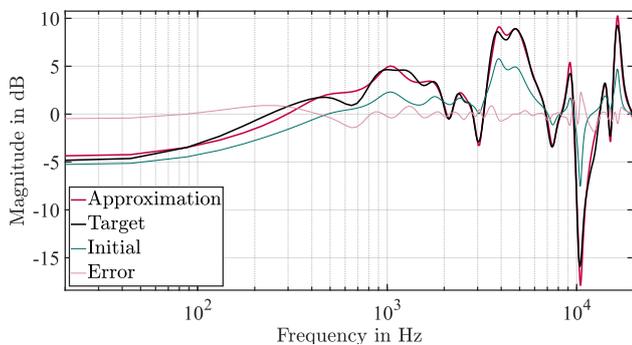


Figure 10: Magnitude responses of the desired HRTF, the initial HRTF estimate and the final approximation of the right ear of 'Subject-008', for an azimuth  $\phi = 0^\circ$  and elevation  $\theta = -22.5^\circ$ .

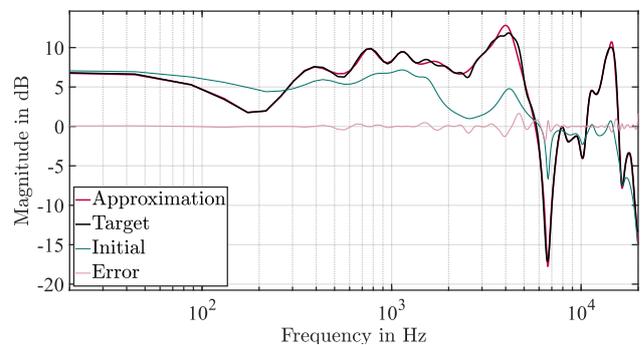


Figure 12: Magnitude responses of the desired HRTF, the initial HRTF estimate and the final approximation of the left ear of 'Subject-008', for an azimuth  $\phi = 0^\circ$  and elevation  $\theta = -45^\circ$ .

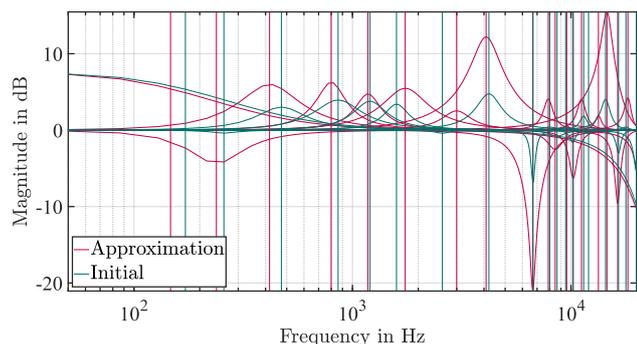


Figure 13: Responses of the individual filters after initialization and after HRTF approximation of the left ear of 'Subject-008', for an azimuth  $\phi = 0^\circ$  and elevation  $\theta = -45^\circ$ .

an HRTF where a filter is individually and sequentially adjusted and adapted. Another approach can be residual learning where the initial number of filter proposals is drastically reduced based on a strong smoothing, followed by an adaptation, a subtraction from the original HRTF and addition of a few filters based on strong residual peaks. Both of the above proposals require multiple adaptation phases and following any improvement comparisons can be made with existing methods for HRTF modelling. The gains can be properly adapted in the frequency domain as well to improve the initialization. The parameters can also be constrained and the HRTF dynamic ranges can be normalized to a fixed range in order to stabilize the performance of the method and ensure an optimal hyper-parameter setup for all examples. A deep cascaded structure might also invoke the problem of vanishing gradients during backpropagation, which should be studied. The fundamental method could replace the instantaneous backpropagation with the normal backpropagation, introduce non-linear functions in the cascade, and use an improved optimization method. Finally, the derivations can be extended to higher order parametric filters.

## 5. SUMMARY

In this paper, a new method of updating the parametric filters in a cascaded structure with the backpropagation algorithm, is described. The structure is usually comprised of a first-order low-frequency shelving filter, followed by a number of second-order peak filters and a first-order high-frequency shelving filter. The expressions for partial local derivatives of each parametric filter inside the cascade are derived w.r.t. a set of control parameters and the required gradients for iterative parameter update are calculated during backpropagation. Since only a few control parameters are updated in this structure it is computationally less expensive compared to update of filter coefficients in traditional FIR and IIR filters, particularly for high orders. Finally, the cascaded structure is used for HRTF modelling, as an example application, to illustrate its performance and discuss the scope of the proposed method.

## 6. REFERENCES

[1] P. A. Regalia and S. K. Mitra, "Tunable digital frequency response equalization filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 1, pp. 118–120, Jan. 1987.

[2] U. Zölzer and T. Boltze, "Parametric digital filter structures.," *In Proc. 99th AES Convention, Preprint No. 4099*, Oct. 1995.

[3] U. Zölzer and F. Keiler, "Parametric second- and fourth-order shelving filters for audio applications.," *In Proc. IEEE 6th Workshop on Multimedia Signal Processing*, pp. 231–234, Sep. 2004.

[4] M. Holters and U. Zölzer, "Parametric higher order shelving filters.," *14th European Signal Processing Conference*, pp. 1–4, Sep. 2006.

[5] V. Välimäki and J. D. Reiss, "All about audio equalization: solutions and frontiers," *Applied Sciences*, vol. 6, no. 5, 2016.

[6] F. X. Y. Gao and W. M. Snelgrove, "An adaptive backpropagation cascade IIR filter," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 39, no. 9, pp. 606–610, Sep. 1992.

[7] R. E. Rose, "Training a recursive filter by use of derivative function," US patent US5905659A, May 1999.

[8] A. D. Back and A. C. Tsoi, "FIR and IIR synapses, a new neural network architecture for time series modeling," *Neural Computation*, vol. 3, no. 3, pp. 375–385, 1991.

[9] P. Campolucci, A. Uncini, and F. Piazza, "Fast adaptive IIR-MLP neural networks for signal processing application," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 6, pp. 3529–3532, 06 1996.

[10] U. Zölzer, *Digital Audio Signal Processing*, John Wiley and Sons, Ltd., Chichester, UK, second edition, 2008.

[11] H. Hasegawa, M. Kasuga, S. Matsumoto, and A. Koike, "Simply realization of sound localization using HRTF approximated by IIR filter," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E83-A, no. 6, pp. 973–978, Jun. 2000.

[12] G. Ramos and M. Cobos, "Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 1735–1738, Sept. 2013.

[13] G. Ramos, M. Cobos, B. Bank, and J. A. Belloch, "A parallel approach to HRTF approximation and interpolation based on a parametric filter model," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1507–1511, Oct. 2017.

[14] P. Nowak, V. Zimpfer, and U. Zölzer, "Automatic approximation of head-related transfer functions using parametric IIR filters," in *Fortschritte der Akustik: DAGA 2020*, Mar. 2020.

[15] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, Oct. 2001.

[16] H. Caracalla and A. Roebel, "Gradient conversion between time and frequency domains using Wirtinger calculus," in *DAFx 2017*, Edinburgh, United Kingdom, Sept. 2017.

[17] X. Wang, S. Takaki, and J. Yamagishi, "Neural source-filter waveform models for statistical parametric speech synthesis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 402–415, 2020.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.