

# Psychoacoustically Motivated Filter Bank Design for Real Time Audio Systems

Asger Hansen & Jonas Dahl



Kongens Lyngby 2014

Technical University of Denmark  
Department of Applied Mathematics and Computer Science  
Matematiktorvet, building 303B,  
2800 Kongens Lyngby, Denmark  
Phone +45 4525 3351  
[compute@compute.dtu.dk](mailto:compute@compute.dtu.dk)  
[www.compute.dtu.dk](http://www.compute.dtu.dk)

# Abstract

---

DFT modulated filter banks are widely used in real time audio systems. Different prototype filter design methods have been proposed in literature. None of the methods use knowledge from psychoacoustic research to reduce the audibility of artefacts introduced by the filter bank. This thesis focus on the design of prototype filters for the DFT modulated filter bank with reduced audibility of artefacts by utilising a frequency domain masking model.

To obtain the masking model the artefacts introduced by the filter bank are quantified by a set of error functions and the psychoacoustic concepts to asses the audibility of the artefacts are discussed.

A quadratic optimisation method for prototype filter designs with and without the masking model is proposed and evaluated. The designs without the masking model shows good performance compared to classical methods while being more flexible. The designs with the masking model have poor performance compared to the designs without when evaluated by PESQ with a spectral subtraction algorithm applied in the filter bank.

The artefacts introduced by the designs with the masking model are analysed and it is concluded that the simplifications in the masking model imposed by the DFT modulated filter bank structure are too severe. Furthermore, the masking model do not account for artefacts in the modulation domain which are enhanced by applying the masking model.



# Resumé

---

DFT modulerede filterbanke er meget udbredt i tidstro lydssystemer. Forskellige designmetoder for prototypefiltre er blevet foreslået. Ingen af disse metoder bruger viden fra psykoakustisk forskning til at reducere hørbarheden af de artefakter der introduceres af filterbanken. Dette kandidatspeciale fokuserer på at designe prototypefiltre til DFT modulerede filterbanke med reduceret hørbarhed af artefakter ved hjælp af en frekvensdomænemaskeringsmodel.

Maskeringsmodellen er opnået ved at kvantificere artefakterne indført af filterbanken til nogle definerede fejl, hvis hørbarhed er vurderet ud fra forskellige psykoakustiske begreber.

En kvadratisk optimeringsmetode til design af prototypefiltre med og uden maskeringsmodellen foreslås og evalueres. Designs uden maskeringsmodellen viser gode resultater i forhold til klassiske metoder og er samtidig mere fleksible. Designs med maskeringsmodellen viser dårlige resultater i forhold til designs uden maskeringsmodellen, når de evalueres af PESQ med en spektral subtraktionsalgoritme anvendt i filter banken.

Artefakterne indført af designet med maskeringsmodellen analyseres, og det konkluderes, at simplificeringerne i maskeringsmodellen, introduceret på grund af strukturen af den DFT modulerede filterbank, er for grove. Desuden forværer maskeringsmodellen artefakter i modulationsdomænet, hvilket modellen ikke tager højde for.



# Preface

---

This thesis was prepared at Department of Applied Mathematics and Computer Science and Department of Electrical Engineering, Technical University of Denmark, to acquire a master's degree in electrical engineering.

This thesis deals with combining psychoacoustic knowledge in a DFT modulated filter bank design to reduce the audibility of the artefacts introduced in the filter bank.

The thesis is divided in 8 chapters.

Chapter 1 is an introduction to current filter bank designs and psychoacoustic knowledge already used in filter bank design. The scope of this thesis is also defined in chapter 1.

Chapter 2 is a thoroughly derivation and analysis of the DFT modulated filter bank and the efficient realisation. The artefacts introduced by the filter bank are quantified in a set of error functions.

Chapter 3 deals with quadratic minimisation of the error functions defined in chapter 2. The computational complexity of the error functions are reduced so the minimisation can be conducted on an ordinary computer.

Chapter 4 investigates the audibility of the artefacts introduced by the filter bank. A masking model is introduced to account for the artefacts introduced by the decimation and interpolation. Finally the model is applied to the error functions from chapter 2 and then minimised in the same manner as in chapter 3.

Chapter 5 show example designs obtained by the minimisation of the error function with and without the psychoacoustic model. Some of the minimisation parameters are investigated more thoroughly.

Chapter 6 evaluates the filter obtained by the proposed method. The design in chapter 3 is evaluated against classical filter bank design method by the proposed error functions and PESQ. The influence of the psychoacoustic model on the filter design is also evaluated.

Chapter 7 discuss the performance of the filter bank design with the psychoacoustic model. The artefacts are analysed and compared to the assumptions and limitations in the psychoacoustic model.

Chapter 8 presents a summary, conclusion and further work.

Kongens Lyngby, March 2014



Asger Ertmann Hansen



Jonas Amtoft Dahl

# Acknowledgements

---

We would like to thank our two supervisors Jan Larsen, Associate Professor, Department of Applied Mathematics and Computer Science, Technical University of Denmark, and Bastian Epp, Assistant Professor, Department of Electrical Engineering, Technical University of Denmark.

We would also like to thank our girlfriends and family for proofreading and support.



# Contents

---

<b>Abstract</b>	<b>i</b>
<b>Resumé</b>	<b>iii</b>
<b>Preface</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>Contents</b>	<b>ix</b>
<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>Nomenclature</b>	<b>xv</b>
<b>Notation</b>	<b>xix</b>
<b>Acronyms</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Prototype Filter Design Methods for Modulated Filter Banks	1
1.2 Other Filter Bank Concepts Utilising Psychoacoustics . . . . .	3
1.3 Scope of Thesis . . . . .	3
<b>2 Definition, Efficient Realisation &amp; Artefacts of the DFT Modulated Filter Bank</b>	<b>5</b>
2.1 Basic Concept of a Filter Bank . . . . .	5
2.2 Modulation of Prototype Filters . . . . .	7
2.3 Derivation & Analysis of the DFT Modulated Filter Bank . .	8
2.4 Efficient Realisation Using an FFT & an IFFT . . . . .	11
2.5 Filter Bank Artefacts & Error Measures . . . . .	15

<b>3</b>	<b>Prototype Filter Design by Minimisation of Error Functions</b>	<b>23</b>
3.1	Analysis Filter Design . . . . .	25
3.2	Synthesis Filter Design . . . . .	27
3.3	Prototype Filter Design with Weighted Errors . . . . .	30
<b>4</b>	<b>Audibility of Artefacts &amp; a Simple Psychoacoustic Model</b>	<b>31</b>
4.1	Audibility of Filter Bank Artefacts . . . . .	31
4.2	Definitions, Assumptions & a Simple Model of Frequency Masking in the Auditory System . . . . .	33
4.3	Psychoacoustic Model for Aliasing/Imaging Artefacts . . . . .	37
4.4	Minimisation of Psychoacoustically Weighted Errors . . . . .	40
<b>5</b>	<b>Example of Filter Designs with &amp; without Psychoacoustic Weighting</b>	<b>43</b>
5.1	Example Filter Design without Psychoacoustic Weighting . . . . .	43
5.2	Example Filter Design with Psychoacoustic Weighting . . . . .	50
<b>6</b>	<b>Evaluation of Prototype Filter Designs</b>	<b>55</b>
6.1	Methods for Evaluation of Prototype Filter Designs . . . . .	55
6.2	Evaluation of Prototype Filter Designs . . . . .	58
<b>7</b>	<b>Analysis &amp; Discussion of Artefacts Introduced by the Psychoacoustic Weighting</b>	<b>69</b>
7.1	Signal Analysis & Discussion of Assumptions in the Psychoacoustic Model . . . . .	69
7.2	Aliasing/Imaging Artefacts in the Modulation Domain . . . . .	72
<b>8</b>	<b>Summary &amp; Conclusion</b>	<b>75</b>
8.1	Further Work . . . . .	76
<b>A</b>	<b>Derivation of Matrices for Prototype Filter Design</b>	<b>77</b>
A.1	Passband Error . . . . .	77
A.2	Inband Aliasing Error . . . . .	79
A.3	Linear Response Error . . . . .	81
A.4	Aliasing/Imaging Cancellation Error . . . . .	84
A.5	Aliasing/Imaging Error . . . . .	87
	<b>Bibliography</b>	<b>91</b>

## List of Figures

---

2.1	Filter bank concept diagram . . . . .	6
2.2	Efficient realisation of a DFT modulated analysis filter bank .	13
2.3	Efficient realisation of a DFT modulated synthesis filter bank	14
2.4	Filter bank concept diagram with aliasing/imaging artifacts .	16
2.5	Illustration of aliasing and imaging through the zeroth band of a filter bank . . . . .	17
4.1	Notched-noise experiment stimulus . . . . .	36
4.2	Auditory filter bandwidth . . . . .	37
4.3	Masking of aliasing/imaging components . . . . .	38
5.1	Impulse response of filters for example design . . . . .	44
5.2	Magnitude response of prototype filter for example design . .	45
5.3	Magnitude response of $T_l$ , $T_c$ and $T_r$ for example design . . . .	46
5.4	Maximum and integrated power of the aliasing/imaging com- ponents for example design . . . . .	47
5.5	Error measures for example design . . . . .	47
5.6	Error measures as a function of $\alpha_a$ . . . . .	48
5.7	Magnitude response of prototype filter for example design with $\alpha_a = 4$ . . . . .	49
5.8	Synthesis error measures as a function of $\alpha_c$ and $\alpha_r$ . . . . .	50
5.9	Impulse response of filters for example design . . . . .	51
5.10	Magnitude response of prototype filter for example design with psychoacoustic weighting . . . . .	52
5.11	Maximum and integrated power of the aliasing/imaging com- ponents for the example design with psychoacoustic weighting	53
5.12	Error measures for example design with psychoacoustic weighting	53
6.1	MOS-LQON obtained by PESQ for four different SNR values when no speech enhancement is applied . . . . .	58

6.2	Error measures for two filter bank designed for PR and NPR	60
6.3	MOS-LQON obtained by PESQ for filter banks optimised for PR and NPR . . . . .	60
6.4	Error measures for a WOLA filter bank and a filter bank optimised by proposed method . . . . .	62
6.5	MOS-LQON obtained by PESQ for a WOLA filter bank and a filter bank optimised by the proposed method . . . . .	62
6.6	Error measures for a filter bank optimised by window method and a filter bank optimised by proposed method . . . . .	64
6.7	MOS-LQON obtained by PESQ for a filter bank optimised by the window method and a filter bank optimised by the proposed method . . . . .	64
6.8	Error surface for the proposed method without psychoacoustic weighting. . . . .	65
6.9	Error surface for the proposed method with psychoacoustic weighting. . . . .	65
6.10	Error measures for two filter banks optimised by proposed method with and without psychoacosutic weighting . . . . .	67
6.11	Psychoacoustically weighed error measures for two filter banks optimised by proposed method with and without psychoacoustic weighting . . . . .	67
6.12	MOS-LQON obtained by PESQ for filter banks optimised by proposed method with and without psychoacosutic weighting	68
7.1	Power spectral density of noise reduced signals obtained by an ideal spectral subtraction algorithm implmented in filter bank optimised by proposed method with and without psychoacoustic weighting . . . . .	70
7.2	Power spectral density estimate of the noise reduced signal by an ideal spectral subtraction and the aliasing/imaging component for $d = 1$ . . . . .	71
7.3	Modulation spectrum of noise reduced signals obtained by an ideal spectral subtraction algorithm implemented in filter bank optimised by the proposed method with and without psychoacoustic weighting . . . . .	73

## List of Tables

---

5.1	Parameters for the filter banks in the design example . . . . .	43
5.2	Parameters for the optimisation of filter banks in the design example . . . . .	44
5.3	Parameters for the optimisation of filter banks in the design example with psychoacoustic weighting . . . . .	51
6.1	Parameters for the filter banks for evaluation of PR and NPR designs . . . . .	59
6.2	Parameters for the optimisations for evaluation of PR and NPR designs . . . . .	59
6.3	Parameters for the filter banks for evaluation of proposed optimisation method against WOLA method . . . . .	61
6.4	Parameters for the optimisation for evaluation of proposed optimisation method against WOLA method . . . . .	61
6.5	Parameters for the filter banks for evaluation of proposed optimisation method against window method . . . . .	63
6.6	Parameters for the optimisation for evaluation of proposed optimisation method against window method . . . . .	63
6.7	Parameters for the filter banks with and without psychoacoustic weighting . . . . .	66
6.8	Parameters for the optimisation with and without psychoacoustic weighting . . . . .	66



# Nomenclature

---

<b>A</b>	Optimisation matrix for the passband error
$A_{p,q}$	The $p$ -th row and $q$ -th column in the optimisation matrix for the passband error <b>A</b>
<b>b</b>	Optimisation vector for the passband error
$b_p$	The $p$ -th row in the optimisation vector for the passband error <b>b</b>
<b>C</b>	Optimisation matrix for the inband aliasing error
$C_{p,q}$	The $p$ -th row and $q$ -th column in the optimisation matrix for the inband aliasing error <b>C</b>
$d$	Aliasing/imaging component index ( $d = 1, 2, \dots, D - 1$ ). $d = 0$ denotes the linear transfer
$D$	Decimation and interpolation ratio in the filter bank
<b>E</b>	Optimisation matrix for the linear response error
$E_{p,q}$	The $p$ -th row and $q$ -th column in the optimisation matrix for the linear response error <b>E</b>
$\text{ERB}(f_c)$	Equivalent Rectangular Bandwidth of auditory filter as a function of centre frequency [GM90]
<b>f</b>	Optimisation vector for the linear response error
$f_c$	Centre frequency of auditory filter in Hz
$f_k[n]$	The filtering or gain applied in the filter bank for the $k$ 'th band at time $n$
$f_{\text{mod}}$	Time domain modulation frequency in Hz
$f_p$	The $p$ -th row in the optimisation vector for the linear response error <b>f</b>
$f_s$	Sample rate in Hz
$f_{\text{shift}}[d]$	Frequency shift of the $d$ 'th aliasing/imaging component in Hz
$F_k(z)$	$\mathcal{Z}$ -transform of filtering or gain applied in the filter bank for the $k$ 'th band

$g$	Normalised distance to the centre frequency in a ROEX filter
$\mathbf{g}$	Vector representation of prototype synthesis filter $\mathbf{g} = [g_0(0), g_0(1), g_0(2), \dots, g_0(L_g - 1)]^T$
$g_0[n]$	Synthesis prototype filter, $n = 0, 1, \dots, L_g - 1$
$g_k[n]$	Synthesis filter in $k$ 'th band, $g_k[n] = g_0[n]W_K^{-nk}$
$G_k(z)$	$\mathcal{Z}$ -transform of synthesis filter in $k$ 'th band
$\mathbf{h}$	Vector representation of prototype analysis filter $\mathbf{h} = [h_0(0), h_0(1), h_0(2), \dots, h_0(L_h - 1)]^T$
$h_0[n]$	Analysis prototype filter, $n = 0, 1, \dots, L_h - 1$
$h_k[n]$	Analysis filter in $k$ 'th band, $h_k[n] = h_0[n]W_K^{-(n-\tau_k)k}$
$H_d(z)$	Desired response of passband for the analysis prototype filter
$H_k(z)$	$\mathcal{Z}$ -transform of analysis filter in $k$ 'th band
$k$	Band number in the filter bank $k = 0, 1, \dots, K - 1$
$K$	Number of bands in the filter bank
$L_g$	Length of synthesis filter
$L_h$	Length of analysis filter
$n_k[m]$	Discrete time noise signal in the $k$ 'th band
$N$	The long term power of the noise masker in the power spectrum model
$N(f)$	The long term power spectral density of a noise masker in the power spectrum model
$N^{\text{post}}$	The long term power of the noise masker in the power spectrum model weighted by auditory filter
$p$	Parameter determine bandwidth of the ROEX filter.
$\mathbf{P}$	Optimisation matrix for the aliasing/imaging error
$P_{p,q}$	The $p$ -th row and $q$ -th column in the optimisation matrix for the aliasing/imaging error $\mathbf{P}$
$P_{\text{aliasing/imaging}}^{\text{post}}[d]$	The long term power of the aliasing/imaging component weighted by auditory filter
$P_{\text{original signal}}^{\text{post}}$	The long term power of the original signal weighted by auditory filter
$P_{\text{aliasing/imaging}}[d]$	The long term power of the aliasing/imaging component
$P_{\text{original signal}}$	The long term power of the original signal
$P_s$	The long term power of the signal in the power spectrum model
$P_s^{\text{post}}$	The long term power of the signal in the power spectrum model weighted by auditory filter
$\mathbf{Q}$	Optimisation matrix for the aliasing/imaging cancellation error
$Q_{p,q}$	The $p$ -th row and $q$ -th column in the optimisation matrix for the aliasing/imaging cancellation error $\mathbf{Q}$

<b>S</b>	Optimisation matrix for the psychoacoustically weighted aliasing/imaging cancellation error
$S_{p,q}$	The $p$ -th row and $q$ -th column in the optimisation matrix for the psychoacoustically weighted aliasing/imaging cancellation error <b>S</b>
$s_k[m]$	Discrete time speech signal in the $k$ 'th band
$t_l[n]$	Impulse response of the linear response
$T_c(z)$	$\mathcal{Z}$ -transform of total aliasing/imaging transfer
$T_d(z)$	$\mathcal{Z}$ -transform of total desired response
$T_l(z)$	$\mathcal{Z}$ -transform of total transfer function of the linear response $t_l[n]$
$T_r(z)$	$\mathcal{Z}$ -transform of total aliasing/imaging without cancellation transfer
<b>U</b>	Optimisation matrix for the psychoacoustically weighted aliasing/imaging error
$U_{p,q}$	The $p$ -th row and $q$ -th column in the optimisation matrix for the psychoacoustically weighted aliasing/imaging error <b>U</b>
$w[d]$	Psychoacoustic weighting function for error functions
$W(f)$	Auditory filter
$W_{\text{roex}}$	Auditory filter approximation by ROEX filter [PNSWM82]
$\hat{W}(f)$	Auditory filter approximation by ROEX filter with bandwidth defined by $\text{ERB}(f_c)$
$y[n]$	Discrete time output signal
$y_k[n]$	Discrete time signal in the $k$ 'th band after processing in the filter bank
$\tilde{y}_k[n]$	Interpolated discrete time signal in the $k$ 'th band after processing in the filter bank
$Y(z)$	$\mathcal{Z}$ -transform of discrete time output signal $y[n]$
$Y_c(z)$	$\mathcal{Z}$ -transform of the aliasing/imaging components discrete output signal
$Y_k(z)$	$\mathcal{Z}$ -transform of discrete time signal in the $k$ 'th band after processing in the filter bank $y_k[n]$
$\tilde{Y}_k(z)$	$\mathcal{Z}$ -transform of interpolated discrete time signal in the $k$ 'th band after processing in the filter bank $\tilde{y}_k[n]$
$x[n]$	Discrete time input signal
$x_k[n]$	Discrete time signal in the $k$ 'th band before processing in the filter bank
$X(z)$	$\mathcal{Z}$ -transform of discrete time input signal $x[n]$
$X_k(z)$	$\mathcal{Z}$ -transform of discrete time signal in the $k$ 'th band before processing in the filter bank $x_k[n]$
$\tilde{X}_k(z)$	$\mathcal{Z}$ -transform of discrete time signal in the $k$ 'th band before decimation in the filter bank $\tilde{x}_k[n]$

---

$\alpha_a$	Weight of the inband aliasing error
$\alpha_c$	Weight of the aliasing/imaging cancellation error
$\alpha_r$	Weight of the aliasing/imaging error
$\alpha_{wc}$	Weight of the psychoacoustically weighted aliasing/imaging cancellation error
$\alpha_{wr}$	Weight of the psychoacoustically weighted aliasing/imaging error
$\beta$	Parameter for a kaiser window
$\beta[d]$	Threshold of audibility defined by ROEX auditory filter approximation truncated to a minimum of $-94$ dB as a function of aliasing component $d$
$\hat{\beta}[d]$	Auditory filter approximation by ROEX filter as a function of the aliasing component $d$
$\epsilon_a$	Inband aliasing error
$\epsilon_c$	Aliasing/imaging cancellation error
$\epsilon_h$	Analysis filter error
$\epsilon_l$	Liner response error
$\epsilon_p$	Passband error
$\epsilon_r$	Aliasing/imaging error
$\epsilon_t$	Total transfer error
$\epsilon_{wc}$	Psychoacoustically weighted aliasing/imaging cancellation error
$\epsilon_{wr}$	Psychoacoustically weighted aliasing/imaging error
$\epsilon_{wt}$	Psychoacoustically weighted total error
$\kappa$	The SNR where the signal is just audible according to the power spectrum model
$\tau_h$	Group delay of analysis filter
$\tau_t$	Total group delay of filter bank
$\phi_g(z)$	Delay vector of with length of $L_g$
$\phi_h(z)$	Delay vector of with length of $L_h$
$\omega_p$	Normalised upper cutoff frequency of the passband $\omega_p = \frac{\pi}{K}$
$\omega_s$	Normalised stopband frequency $\omega_s = \frac{\pi}{D}$
$\omega_{\text{shift}}[d]$	Normalised frequency shift of aliasing/imaging components

# Notation

---

$x[n]$	Discrete function of $n$
$x(n)$	Continuous function of $n$
$W_K$	Twiddle factor ( $W_K = e^{-j2\pi/K}$ )
$X(z)$	$\mathcal{Z}$ -transform of the signal $x[n]$
$\mathbf{x}$	Bold lower case is a vector
$\mathbf{X}$	Bold upper case is a matrix
$\mathbf{x}^*$	Complex conjugate of $\mathbf{x}$
$\mathbf{x}^T$	Transpose of $\mathbf{x}$
$\mathbf{z}^H$	Hermitian transpose of $\mathbf{z}$
$\mathbf{X}^\dagger$	Moore-Penrose pseudoinverse of $\mathbf{X}$
$\delta[n]$	Kronecker delta function (unit impulse)
$\Delta_K[n]$	Kronecker comb function with period $K$
$\text{princ arg}(z)$	Principle argument of $z$ in the range from $-\pi$ to $\pi$
$\text{sinc}(n)$	Normalised sinc function
$\mathcal{H}\{x[n]\}$	Hilbert transform of $x[n]$
$\gamma_x[m]$	Autocorrelation of $x[n]$ at lag $m$
$ z $	The absolute value of $z$
$\angle z$	The argument of $z$
$\Re\{z\}$	The real part of $z$
$\bar{x}[n]$	The mean value of $x[n]$ for all $n$
$\arg \min_n x[n]$	The argument $n$ where $x[n]$ is minimised
$\max\{x[n]\}$	The maximum value of $x[n]$
$\mathbb{Z}$	All integers
$\mathbb{N}^+$	All positive integers 0



# Acronyms

---

COLA	Constant Overlap-Add.
DAC	Digital-to-Analog Converter.
DFT	Discrete Fourier Transform.
ERB	Equivalent Rectangular Bandwidth.
FIR	Finite Impulse Response.
JND	Just Noticeable Difference.
MOS	Mean Opinion Score.
MOS-LQON	Mean Opinion Score Listening Quality Objective Narrowband.
NPR	Near Perfect Reconstruction.
OLA	Overlap-Add.
PESQ	Perceptual Evaluation of Speech Quality.
PR	Perfect Reconstruction.
PSD	Power Spectral Density.
ROEX	Rounded Exponential.
SNR	Signal to Noise Ratio.
WOLA	Weighted Overlap-Add.



# Introduction

---

Filter banks are widely used as a fundamental building block of the digital signal processing in embedded audio systems like hearing aids and communication devices [HS08]. One of the most used filter bank structures is the Discrete Fourier Transform (DFT) modulated filter bank because of the low computational complexity. Different methods for designing prototype filters for modulated filter banks have been proposed. Although the DFT modulated filter bank is widely used in audio applications, none of the methods use the knowledge from psychoacoustic research to reduce the audibility of the artefacts the filter bank introduces. This thesis aims to combine knowledge from psychoacoustics with a flexible prototype filter design method to obtain low complexity DFT modulated filter banks with reduced audible artefacts for use in embedded real-time audio systems.

## 1.1 Prototype Filter Design Methods for Modulated Filter Banks

Design methods for modulated filter banks can be grouped in three categories.

### **Weighted Overlap-Add (WOLA)**

WOLA is based on the Overlap-Add (OLA) method for efficient implementation of Finite Impulse Response (FIR) filters. OLA use a

rectangular analysis filter with zeropadding and a full length rectangular synthesis filter. In WOLA the rectangular filters are replaced by filters which often fulfil the time-domain Constant Overlap-Add (COLA) constraint [Smi11]. This gives Perfect Reconstruction (PR) filter banks, but with limited flexibility for the filter design. There are no well defined optimisation algorithms, so the filters are often designed by experience and intuition. Example filters can be found in [Smi11]. In [GL84] it is shown that a generalised Hamming window can be used as both the analysis and synthesis filter when oversampling with a multiplicity of four. In [CR83] multiple constraints, both in time and frequency, are defined for PR and some filters are proposed.

### **FIR Filter Design**

This category contains methods for prototype filter design based on traditional FIR filter design methods. A frequency domain specification of the desired frequency response is given to a FIR filter design method, e.g. Window method, Parks–McClellan, Equiripple, Least-squares, etc. [PB87, OSB99]. The window method is the most used for prototype filter design and is based on the frequency domain COLA constraint with a rectangular magnitude response. This results in an infinite sinc-function in time domain, which is therefore approximated by windowing the sinc. Different design methods for the windowing of the sinc has been proposed [CR83, LV98, GdHCN01, CRALMBL02, YGNT04]. Most of these are iterative and aim to minimise the linear response error of the filter bank. Because the filters are not originally designed for filter banks, PR is not obtained in any of the window based design methods, though most of them have Near Perfect Reconstruction (NPR).

### **Quadratic Optimisation**

A new and very flexible method for optimising prototype filters for DFT modulated filter banks is presented in [dH01, dHGNC01, dHGNC03]. The optimisation is based on a number of squared errors which are minimised by solving two least squares problems, one for the analysis prototype filter and one for the synthesis prototype filter. The method is very flexible and allow arbitrary filter lengths, downsampling ratio and number of bands, but because the analysis filter is designed before the synthesis filter, the optimal combination is not ensured. Furthermore, there are no error function to describe the aliasing/imaging with cancellation, which means that the method can not obtain PR. Several nonlinear iterative methods have been developed to optimise the analysis and synthesis filters together [DNCdH04, DNC05, WTRD08].

## 1.2 Other Filter Bank Concepts Utilising Psychoacoustics

None of the above methods take any psychoacoustic aspects into account. However, many filter bank designs are closely linked to psychoacoustics. Most models of the human hearing use filter banks to model frequency resolution. The spacing of the filters can be modelled by critical bands (Bark) or Equivalent Rectangular Bandwidth (ERB) [Moo12]. Different filters have been proposed such as Rounded Exponential (ROEX) [PNSWM82], Gammatone [PNSHR87], dual resonance [LPM01] and filter cascades [Lyo11]. These filters are not meant for low power real-time audio systems, but for modelling human hearing.

The spacing of the filters have been used in the design of filter banks for real-time audio systems. One approximation is the warped DFT modulated filter bank [HKS<sup>+</sup>00]. This filter bank have many of the same properties as the uniform DFT modulated filter bank, although some additional artefacts are introduced [Lö11].

The warped DFT modulated filter bank only use the psychoacoustic knowledge of frequency resolution and do not deal with the audibility of the artefacts introduced by the filter bank. In this thesis only the uniform DFT modulated filter bank is considered as the focus is on the artefacts and not on frequency resolution, though the two may influence each other.

## 1.3 Scope of Thesis

To obtain a prototype filter optimisation algorithm for a DFT modulated filter bank that reduces the audibility of the filter bank artefacts, this thesis will cover

- Definition and structure of the DFT modulated filter bank and the efficient realisation.
- Modification of the filter bank optimisation method proposed by [dH01] so the method can handle both PR and NPR optimisation while keeping the flexibility.
- Discussion of artefacts introduced by the filter bank and the psychoacoustic concepts used to quantify the audibility of these artefacts.

- Define a psychoacoustic model that can be used in the optimisation of prototype filters for a DFT modulated filter bank to reduce the audibility of filter bank artefacts.
- Apply the psychoacoustic model to the optimisation method to obtain a psychoacoustically optimised filter bank.
- Evaluation of the optimisation methods introduced in this thesis compared to designs from weighted overlap add and the window method. The evaluation is performed with an ideal spectral subtraction algorithm which is evaluated by an objective quality measure (PESQ).

## CHAPTER 2

# Definition, Efficient Realisation & Artefacts of the DFT Modulated Filter Bank

---

This chapter presents the DFT modulated filter bank. In the first section the basic concept of a filter bank is defined. In the second section the modulation of the prototype filters is presented with some comments on earlier approaches. In the third section the equations for the DFT modulated filter bank are derived and the total response equation is analysed. In the fourth section an efficient realisation using an FFT and an IFFT is presented. In the last section the artefacts introduced in the filter bank are described and a series of squared error functions to evaluate and later optimise the prototype filters are introduced.

## 2.1 Basic Concept of a Filter Bank

A filter bank consist of an analysis part and a synthesis part. The analysis part consists of a set of bandpass filters which are applied to a time domain signal in a parallel manner to obtain a time-frequency representation of the signal. The synthesis part take the time-frequency signal obtained by the analysis part and transforms it back to the time domain. Subband processing can be performed on the time-frequency signals before synthesis. A filter bank is shown in figure 2.1.

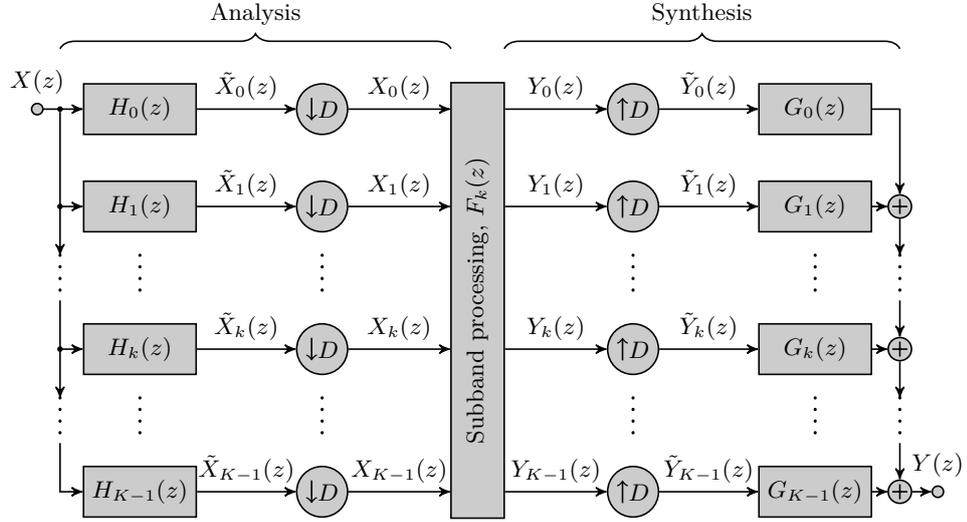


Figure 2.1: Filter bank concept diagram showing the analysis and synthesis parts. The analysis consists of a bank of decimation filters and downsamplers. The synthesis part consists of a bank of upsamplers and interpolation filters. Subband processing can be performed on the time-frequency signals between the analysis and synthesis parts.

As the input signal,  $X(z)$ , is bandlimited by the analysis filters,  $H_k(z)$ , the subband signals obtained by the filtering can be downsampled according to the bandwidth of the analysis filter in the given subband. This means the analysis filters are also used as decimation filters. As the sample rate is reduced in the subbands, the computational complexity of processing the subbands is reduced. When reconstructing the signal in the synthesis part, the signals are upsampled again, meaning that the synthesis filters,  $G_k(z)$ , are used as interpolation filters.

When some constraints for the analysis and the synthesis filters are fulfilled, the filter bank can be classified as a PR filter bank [CR83, Vai93]. This means that when no subband processing is performed the filter bank only introduces a delay and a scaling, i.e.

$$Y(z) = cX(z)z^{-\tau_t}, \quad c \neq 0 \quad (2.1)$$

where  $X(z)$  is the input signal,  $Y(z)$  is the output,  $c$  is a constant and  $\tau_t$  is the total delay of the filter bank in samples.

PR is achieved by cancellation between aliasing and imaging components generated in the decimation and interpolation processes. When processing is performed in a PR filter bank the aliasing and imaging cancellation is no longer perfect resulting in errors. The PR property can be very useful in some cases, but sets restrictions on the filter design. Due to these restrictions

many filter banks are design with NPR instead. When the PR constraints are relaxed the filters can be designed with better attenuation of aliasing and imaging to reduce the error when processing is performed [Vai93, Löl11].

## 2.2 Modulation of Prototype Filters

DFT modulated filter banks are based on a pair of prototype filters (often low-pass) that are modulated, i.e. frequency shifted, to generate a bank of bandpass filters. The modulation of the prototype filters are not well defined in literature. Most [dH01, GdHCN01, dHGCN01, dHGCN03, DNCdH04, DNC05, WTRD08] use a modulation of

$$\begin{aligned} h_k[n] &= h_0[n]W_K^{-nk}, & n = 0, \dots, L_h - 1 \\ g_k[n] &= g_0[n]W_K^{-nk}, & n = 0, \dots, L_g - 1 \end{aligned} \quad (2.2)$$

where  $h_0[n]$  is the analysis prototype filter,  $g_0[n]$  is the synthesis prototype filter,  $W_K = e^{-j2\pi/K}$ ,  $k$  is the band number,  $K$  is the number of bands and  $L_h$  and  $L_g$  are the length of the analysis and synthesis filters respectively.

In [Vai93] it is shown that for a DFT followed by an IDFT, which can be interpreted as a filter bank with rectangular filters with a length equal to the number of DFT bins, the modulation is

$$\begin{aligned} h_k[n] &= h_0[n]W_K^{-(n+1)k}, & n = 0, \dots, K - 1 \\ g_k[n] &= g_0[n]W_K^{-nk}, & n = 0, \dots, K - 1 \end{aligned} \quad (2.3)$$

[Löl11, YGNT04] use a modulation of

$$\begin{aligned} h_k[n] &= h_0[n]W_K^{-nk}, & n = 0, \dots, L_h - 1 \\ g_k[n] &= g_0[n]W_K^{-(n+1)k}, & n = 0, \dots, L_g - 1 \end{aligned} \quad (2.4)$$

[EM01] use a modulation of

$$\begin{aligned} h_k[n] &= h_0[n]W_K^{nk}, & n = 0, \dots, L - 1 \\ g_k[n] &= g_0[n]W_K^{(n-L+1)k}, & n = 0, \dots, L - 1 \end{aligned} \quad (2.5)$$

where  $L$  is the length of the analysis and synthesis filters.

Others [Mer99, PM07] do not define the filter position, and therefore also do not define the modulation offset. Some [Cro80, CR83, OSB99, Smi11] modulate and demodulate the signal instead of the filters.

This thesis use a modulation of

$$\begin{aligned} h_k[n] &= h_0[n]W_K^{-(n-\tau_t)k}, & n = 0, \dots, L_h - 1 \\ g_k[n] &= g_0[n]W_K^{-nk}, & n = 0, \dots, L_g - 1 \end{aligned} \quad (2.6)$$

where  $\tau_t$  is the desired total group delay of the filter bank. If the total group delay is set to  $\tau_t = cK - 1$  with  $c \in \mathbb{Z}$  where  $\mathbb{Z}$  denotes the set of all integers, this modulation is equal to (2.3). The choice of modulation offset, i.e.  $\tau_t$ , is discussed in section 2.3.2.

The centre frequencies of the different bandpass filters are uniformly spaced on the frequency axis, and as all filters are just modulated versions of each other, the bandwidth of the filters are the same. This means that the same downsampling rate can be used for all subbands. As the modulation is performed with a complex exponential function the filters become complex, resulting in complex subband signals even when the input signal is real. For real input signals, the negative frequencies are complex conjugates of the positive, so only the positive frequencies needs to be processed to synthesise the fullband signal, i.e.  $k = 0, 1, \dots, K/2$ .

## 2.3 Derivation & Analysis of the DFT Modulated Filter Bank

In this section the main equations for the filter bank are derived. The filter bank with signal symbols is shown in figure 2.1.

The  $\mathcal{Z}$ -transform of the modulated analysis filters is

$$\begin{aligned} H_k(z) &= \sum_{n=0}^{L_h-1} h_0[n]W_K^{-(n-\tau_t)k} z^{-n} \\ &= H_0(zW_K^k)W_K^{\tau_t k} \end{aligned} \quad (2.7)$$

The subband signals are then given by

$$\tilde{X}_k(z) = X(z)H_k(z) \quad (2.8)$$

Downsampling the subband signals by a factor of  $D$  gives

$$\begin{aligned} X_k(z) &= \frac{1}{D} \sum_{d=0}^{D-1} \tilde{X}_k(z^{1/D}W_D^d) \\ &= \frac{1}{D} \sum_{d=0}^{D-1} X(z^{1/D}W_D^d)H_k(z^{1/D}W_D^d) \end{aligned} \quad (2.9)$$

where  $d = 0$  denotes the linear part of the downsampling process and  $d = 1, 2, \dots, D - 1$  denotes the aliasing components. Assuming the subband processing is a simple filtering operation by  $F_k(z)$ , the processed subband signals are

$$Y_k(z) = F_k(z)X_k(z) \quad (2.10)$$

Upsampling the processed subband signals gives

$$\begin{aligned} \tilde{Y}_k(z) &= Y_k(z^D) \\ &= F_k(z^D) \frac{1}{D} \sum_{d=0}^{D-1} X(zW_D^d)H_k(zW_D^d) \end{aligned} \quad (2.11)$$

Filtering with the synthesis filters and summing yields

$$\begin{aligned} Y(z) &= \sum_{k=0}^{K-1} \tilde{Y}_k(z)G_k(z) \\ &= \sum_{k=0}^{K-1} F_k(z^D) \frac{1}{D} \sum_{d=0}^{D-1} X(zW_D^d)H_k(zW_D^d)G_k(z) \\ &= \sum_{d=0}^{D-1} X(zW_D^d) \frac{1}{D} \sum_{k=0}^{K-1} F_k(z^D)H_k(zW_D^d)G_k(z) \end{aligned} \quad (2.12)$$

This can be rewritten to

$$\begin{aligned} Y(z) &= X(z) \underbrace{\frac{1}{D} \sum_{k=0}^{K-1} F_k(z^D)H_k(z)G_k(z)}_{\text{Linear filtering}} \\ &+ \underbrace{\sum_{d=1}^{D-1} X(zW_D^d) \frac{1}{D} \sum_{k=0}^{K-1} F_k(z^D)H_k(zW_D^d)G_k(z)}_{\text{Aliasing/imaging}} \end{aligned} \quad (2.13)$$

By assuming no processing in the filter bank,  $F_k(z^D) = 1$ , the transfer function of the linear response can be defined as

$$T_l(z) = \frac{1}{D} \sum_{k=0}^{K-1} H_k(z)G_k(z) \quad (2.14)$$

For the aliasing/imaging part the system is not linear and a transfer function can not be obtained. To describe the transfer of aliasing/imaging a transfer

function for  $X(zW_D^d)$  can be introduced. This means that for an input of  $X(z)$  the  $d$ -th aliasing/imaging component of the output is given by

$$Y_c(z) = T_c(z)X(zW_D^d), \quad d = 1, 2, \dots, D - 1 \quad (2.15)$$

where  $T_c(z)$  is the aliasing/imaging transfer function

$$T_c(z) = \frac{1}{D} \sum_{k=0}^{K-1} H_k(zW_D^d)G_k(z), \quad d = 1, 2, \dots, D - 1 \quad (2.16)$$

The aliasing/imaging transfer function,  $T_c(z)$ , describes the amount of aliasing/imaging in the output when no processing is performed. This means that cancellation between bands can be utilised.

To obtain a measure of the transfer of aliasing/imaging when processing is performed a function with power wise summation over bands is defined

$$T_r(z) = \frac{1}{D} \sqrt{\sum_{k=0}^{K-1} |H_k(zW_D^d)G_k(z)|^2}, \quad d = 1, 2, \dots, D - 1 \quad (2.17)$$

This function describes the aliasing/imaging without cancellation between bands. It is not a normal transfer function as phase information is lost, but describes the expected magnitude transfer of aliasing/imaging components when no cancellation is assumed.

### 2.3.1 Constraints for Perfect Reconstruction

To obtain PR, it is required that the linear response is only a scaling and delay, i.e.

$$T_l(z) = cz^{-\tau t}, \quad c \neq 0 \quad (2.18)$$

and that the aliasing/imaging transfer function is zero for all  $d$ , i.e.

$$T_c(z) = 0, \quad \forall z, d \quad (2.19)$$

This is possible by designing the analysis and synthesis filters to have cancellation at specific points in time.

### 2.3.2 Modulation Revisited

In this section we will look at the modulation again, in order to show the influence of the offset in the modulation on the impulse response of the linear

part. The impulse response of  $T_l(z)$  is

$$\begin{aligned}
t_l[n] &= \frac{1}{D} \sum_{k=0}^{K-1} (h_k[n] * g_k[n]) \\
&= \frac{1}{D} \sum_{k=0}^{K-1} \sum_l h_0[l] W_K^{-(l-\tau_t)k} g_0[n-l] W_K^{-(n-l)k} \\
&= \frac{1}{D} (h_0[n] * g_0[n]) \sum_{k=0}^{K-1} W_K^{-(n-\tau_t)k} \\
&= \frac{K}{D} (h_0[n] * g_0[n]) \Delta_K[n - \tau_t]
\end{aligned} \tag{2.20}$$

where

$$\Delta_K[n] = \sum_{m=-\infty}^{\infty} \delta[n - mK], \quad n, m \in \mathbb{Z} \tag{2.21}$$

i.e. a Kronecker comb function with period  $K$ . This means that, regardless of the filters, the impulse response of the linear part will be zero except when  $n = cK + \tau_t$  with  $c \in \mathbb{Z}$ . By designing the filters so the convolution of the analysis and synthesis prototype filters is zero for all  $n = cK + \tau_t$  except when  $c = 0$ , the linear part will be a scaling and a delay of  $\tau_t$ .

If there were no offset in the modulation, i.e. (2.2), the only possible delays would be multiples of  $K$ .

For the modulation used in (2.3) and (2.4), the only possible delays are at  $n = cK - 1$  with  $c \in \mathbb{N}^+$  where  $\mathbb{N}^+$  denotes the set of all positive integers. When using symmetric filters with a length of  $L_h = L_g = cK$  with  $c \in \mathbb{N}^+$  the total group delay is  $cK - 1$ . So for symmetric filters with lengths which are multiples of  $K$  that modulation will work.

The modulation in (2.5) will work for all symmetric filters when the analysis and synthesis filters are of the same length.

## 2.4 Efficient Realisation Using an FFT & an IFFT

An efficient realisation of the DFT modulated filter bank can be obtained by polyphase decomposition and using an FFT and an IFFT. In order to simplify the calculations we only look at the case where  $L_h = L_g = L = RK$ , where  $R$  is a positive integer and  $\tau_t = cK - 1$  with  $c \in \mathbb{N}^+$ . The way to generalise the result to arbitrary  $L_h, L_g$  and  $\tau_t$  is noted in the end of each part.

### 2.4.1 Analysis Part

Writing the analysis equation (2.9) in time domain yields

$$x_k[m] = \sum_{l=0}^{L-1} x[mD - l]h_0[l]W_K^{-(l-cK+1)k} \quad (2.22)$$

with  $m$  being the downsampled time index. The sum over  $l$  can be split in two by substituting  $l = rK - 1 - v$ , where  $v = 0, \dots, K - 1$ ,  $r = 1, \dots, R$  and  $R = L/K$

$$x_k[m] = \sum_{r=1}^R \sum_{v=0}^{K-1} x[mD - rK + 1 + v]h_0[rK - 1 - v]W_K^{-(rK-1-v-cK+1)k} \quad (2.23)$$

This is equivalent to a type II polyphase decomposition with  $K$  as the number of polyphase components. By realising that  $W_K^{-(rK-cK)k} = 1$  for all  $r, c$  and  $k$ , and rearranging the sums, the following is obtained

$$x_k[m] = \sum_{v=0}^{K-1} W_K^{vk} \sum_{r=1}^R x[mD - rK + 1 + v]h_0[rK - 1 - v] \quad (2.24)$$

The outer sum over  $v$  and the  $W_K^{vk}$  is a DFT and can be efficiently implemented as an FFT. The structure is illustrated in figure 2.2.

The structure shown in figure 2.2 can be generalised to arbitrary  $L_h$  by extending it upwards and to arbitrary  $\tau_t$  by circular shift of the input to the FFT with  $\tau_t + 1$ .

### 2.4.2 Synthesis Part

Writing the synthesis equation (2.12) in time domain yields

$$y[n] = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \tilde{y}_k[n - l]g_k[l] \quad (2.25)$$

where

$$\tilde{y}_k[n] = \begin{cases} y_k[n/D], & n/D \in \mathbb{Z} \\ 0, & \text{otherwise} \end{cases} \quad (2.26)$$

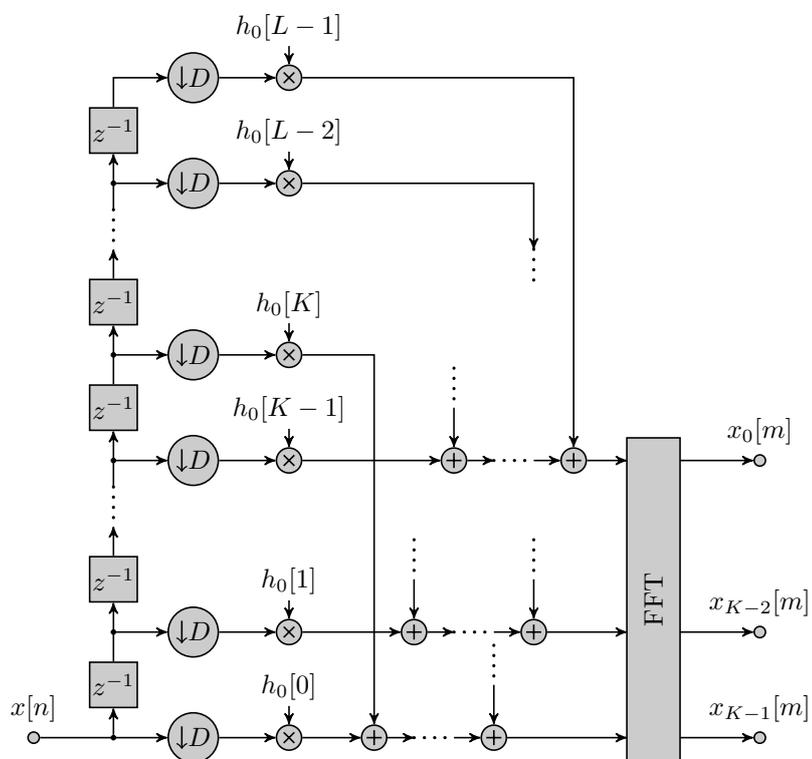


Figure 2.2: Efficient realisation of a DFT modulated analysis filter bank using an FFT for the modulation.

where  $\mathbb{Z}$  denotes the set of all integers. This can be rewritten to

$$\begin{aligned}
 y[n] &= \sum_{l=0}^{L-1} g_0[l] \sum_{k=0}^{K-1} \tilde{y}_k[n-l] W_K^{-lk} \\
 &= \sum_{l=0}^{L-1} g_0[l] \begin{cases} \sum_{k=0}^{K-1} y_k[n-l/D] W_K^{-lk}, & n-l/D \in \mathbb{Z} \\ 0, & \text{otherwise} \end{cases} \quad (2.27)
 \end{aligned}$$

Because of the periodic nature of  $W_K^{-lk}$ , the sum over  $k$  is just an IDFT of  $y_k$  at different time instances. Although not completely obvious, this is equivalent to the structure shown in figure 2.3.

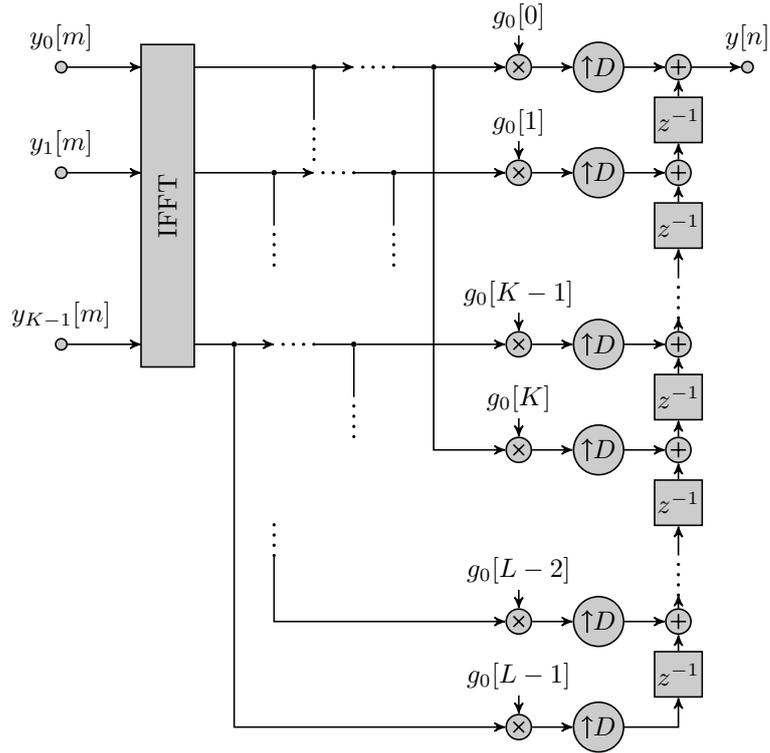


Figure 2.3: Efficient realisation of a DFT modulated synthesis filter bank using an IFFT for the modulation.

The structure shown in figure 2.3, can be generalised to arbitrary  $L_g$  by extending it downwards.

## 2.5 Filter Bank Artefacts & Error Measures

When designing filter banks one needs to consider the different artefacts introduced by the filter bank. This section will look at the artefacts introduced in both the analysis and synthesis parts of the filter bank and error measures are introduced to quantify these artefacts. The error measures are similar to the measures defined in [dH01] except an error measuring the aliasing/imaging in the output when no processing is performed.

The audibility of the artefacts are not considered in this section, but will be discussed in chapter 4.

The artefacts introduced in filter banks can be separated in a linear part and an aliasing/imaging part.

The linear part is closely linked to the linear response,  $T_l(z)$ , where an allpass filter with linear phase is desired. Any deviation from these constraints of  $T_l(z)$  results in artefacts. Constraints for the linear response of the analysis part alone can also be defined. This could be constraints like a flat passband and a linear phase.

The aliasing/imaging artefacts are introduced by the down- and upsamplers. The aliasing/imaging artefacts that are most important for the performance of the filter bank are the inband aliasing and the residual aliasing/imaging [CR83]. In figure 2.4 the aliasing/imaging artefacts are illustrated.

The inband aliasing is introduced by downsampling and is therefore closely related to the stopband attenuation of the analysis filters. In figure 2.5 a concept drawing of the zeroth band of a filter bank with four bands and a downsampling ratio of two is shown. The figure shows the inband aliasing introduced by downsampling and the imaging introduced by upsampling.

The inband aliasing contaminates the subband signals by introducing a frequency shifted error signal. To reduce this error a good stopband attenuation of the analysis filters is required [dH01].

The imaging itself is not so crucial because it does not influence the subband processing. However, it influences the residual aliasing/imaging. The residual aliasing/imaging is the amount of aliasing and imaging after cancellation in the output. By careful design of the analysis and synthesis filters the residual aliasing/imaging can be completely cancelled even though both aliasing and imaging is present inside the filter bank [Vai93].

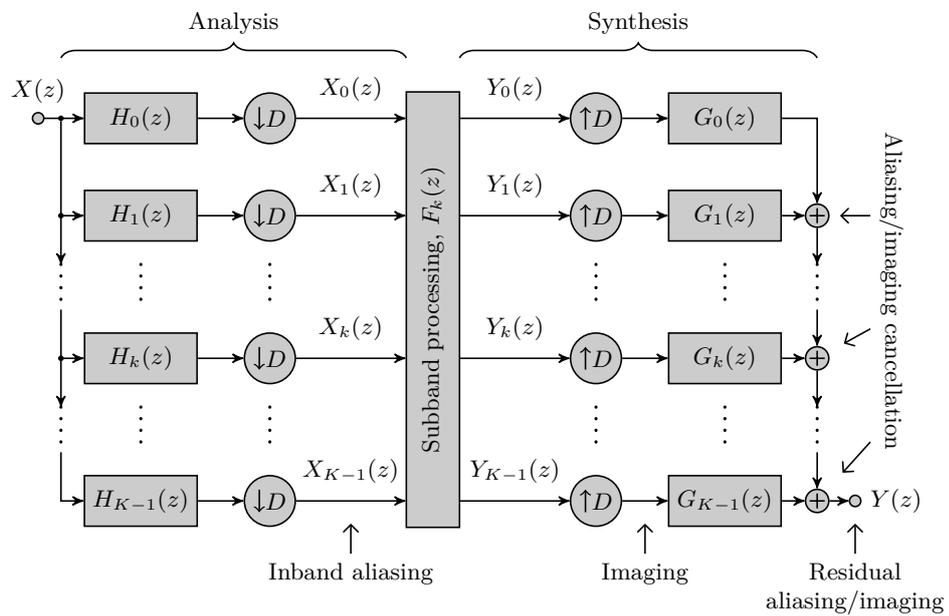


Figure 2.4: Filter bank concept diagram showing the analysis and synthesis parts. Because the decimation filters are not ideal, there will be aliasing in the frequency bands from the analysis part. Likewise, because the interpolation filters are not ideal, there will be imaging in the output of the synthesis filters. When no processing is performed in the filter bank and the PR constraints are fulfilled, the aliasing/imaging will cancel at the summation of the bands in the synthesis part. If PR is not obtained or processing is performed in the filter bank, there will be residual aliasing/imaging at the output.

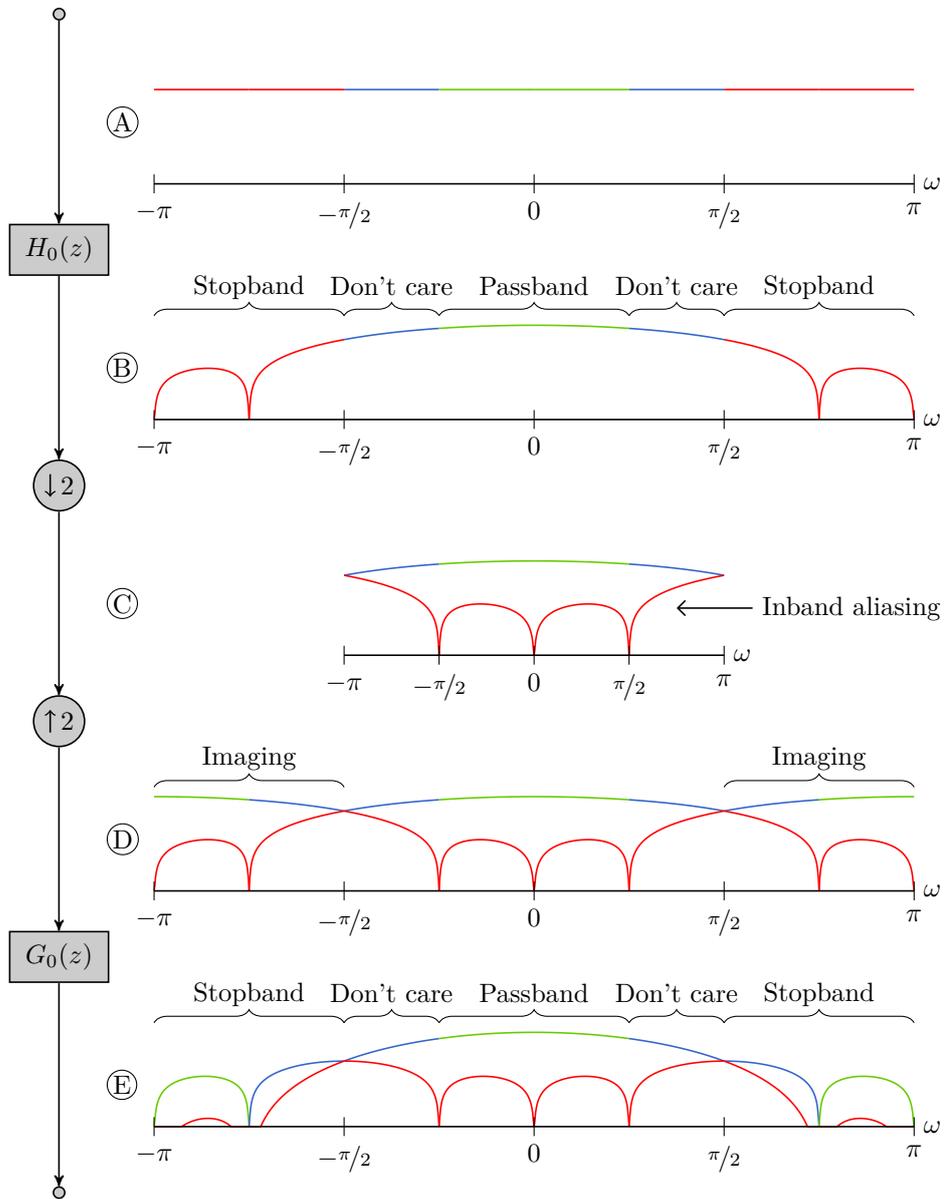


Figure 2.5: Illustration of aliasing and imaging through the zeroth band of a filter bank. The number of bands,  $K$ , is 4 and the downsampling ratio,  $D$ , is 2. Both the analysis and synthesis filter is the square root of a Hann window of length,  $L$ , equal to the number of bands, i.e. 4. The illustration shows the magnitude spectrums at different positions through the filter bank for an input signal with a flat magnitude spectrum.

When processing is performed in the filter bank, the aliasing/imaging will not cancel completely. Therefore, it is desirable to have good attenuation of the aliasing/imaging without cancellation when processing is performed [Vai93].

### 2.5.1 Error Functions for Artefacts in the Analysis Filter Bank

The artefacts of the analysis filter are described by two error functions proposed in [dH01]

- Passband response error
- Inband aliasing error

#### 2.5.1.1 Passband Response Error

Often it is desirable to have a flat magnitude spectrum and a linear phase in the passband. The bandwidth of the passband depends on the number of bands in the filter bank,  $K$ . The normalised bandwidth of the passband is set to  $\omega_p = \pi/K$ , because this is the frequency where the band edges meet between adjacent bands. This corresponds to the part denoted by Passband in part (B) of figure 2.5. If we define a desired response in the passband,  $H_d(e^{j\omega})$ , we can define an error measure for the passband response. We only need to consider the prototype filter as the response is the same for all filters, just with different frequency limits. Here we define the error as the integral of the squared distance in the frequency domain

$$\epsilon_p = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} |H_0(e^{j\omega}) - H_d(e^{j\omega})|^2 d\omega \quad (2.28)$$

where  $H_d(e^{j\omega})$  is the desired frequency response,  $H_0(e^{j\omega})$  is the frequency response of the prototype analysis filter and  $\omega_p$  is the upper cutoff frequency of the passband, i.e.  $\omega_p = \pi/K$ . The desired passband response,  $H_d(z)$ , is often chosen to be a simple delay, i.e.  $H_d(z) = z^{-\tau_h}$ , where  $\tau_h$  denotes the desired group delay of the analysis filter bank.

### 2.5.1.2 Inband Aliasing Error

High inband aliasing attenuation is important for most applications of filter banks. When using adaptive filters in the subbands, the inband aliasing will reduce the adaptation speed and limit the accuracy of the filters [LGK09]. If the stopband of the analysis filter is set to all frequencies above  $\omega_s = \pi/D$ , the inband aliasing is the signal that is passed through the stopband of the analysis filter. Thus the stopband attenuation and inband aliasing attenuation are equal. The inband aliasing error is defined as the integral of the sum of squared aliasing components

$$\epsilon_a = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D-1} \sum_{d=1}^{D-1} \left| H_0(e^{j\omega/D} W_D^d) \right|^2 d\omega \quad (2.29)$$

In figure 2.5 part © this corresponds to the red curve. An equal and maybe more intuitive definition is the integral of the squared stopband error

$$\epsilon_a = \frac{1}{2(\pi - \pi/D)} \left( \int_{-\pi}^{-\pi/D} \left| H_0(e^{j\omega}) \right|^2 d\omega + \int_{\pi/D}^{\pi} \left| H_0(e^{j\omega}) \right|^2 d\omega \right) \quad (2.30)$$

This is the parts denoted by Stopband in ② of figure 2.5.

## 2.5.2 Error Functions for Artefacts in the Synthesis Filter Bank

The artefacts of the synthesis filter is described by three error functions

- Linear response error
- Aliasing/imaging cancellation error
- Aliasing/imaging error

The first and third are proposed in [dH01]. The second error describes the aliasing/imaging when no processing is performed in the filter bank. This makes it possible to obtain PR.

### 2.5.2.1 Linear Response Error

The linear part of the total response describes the desired part of the total transfer when no processing is performed. It is often desired to have a pure

delay as the linear response. This is also the linear part of the PR constraints in section 2.3.1. Recall the linear part of the total response (2.14)

$$T_l(z) = \frac{1}{D} \sum_{k=0}^{K-1} H_k(z)G_k(z) \quad (2.31)$$

The squared error of the linear response is defined as

$$\epsilon_l = \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_l(e^{j\omega}) - T_d(e^{j\omega})|^2 d\omega \quad (2.32)$$

where  $T_d(e^{j\omega})$  is the desired total response.  $T_d(e^{j\omega})$  is chosen to be a simple delay, i.e.  $T_d(z) = z^{-\tau_t}$ , where  $\tau_t$  is the desired total group delay of the filter bank.

### 2.5.2.2 Aliasing/Imaging Cancellation Error

The aliasing/imaging cancellation error describes the amount of aliasing/imaging at the output of the filter bank when no processing is performed. The second constraint for PR, is that the aliasing/imaging transfer is zero for all  $d$  and  $z$ . Recall the aliasing/imaging transfer functions (2.16)

$$T_c(z) = \frac{1}{D} \sum_{k=0}^{K-1} H_k(zW_D^d)G_k(z), \quad d = 1, 2, \dots, D-1 \quad (2.33)$$

The squared error of the aliasing/imaging cancellation is defined as

$$\epsilon_c = \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_c(e^{j\omega})|^2 d\omega \quad (2.34)$$

When this error and the linear response error is zero the filter bank is a PR filter bank.

### 2.5.2.3 Aliasing/Imaging Error

If extensive manipulation is made in the subbands, it is often desirable to design for low residual aliasing/imaging without relying on cancellation. The aliasing/imaging with power wise summation over bands (2.17) can be used for this error function

$$T_r(z) = \frac{1}{D} \sqrt{\sum_{k=0}^{K-1} |H_k(zW_D^d)G_k(z)|^2}, \quad d = 1, 2, \dots, D-1 \quad (2.35)$$

The aliasing/imaging error is defined as

$$\epsilon_r = \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_r(e^{j\omega})|^2 d\omega \quad (2.36)$$

When the aliasing/imaging error is low, the filter bank is good for extensive manipulation in the subbands, as the amount of aliasing and imaging is low without relying on aliasing/imaging cancellation.

The name *aliasing/imaging error* may be somewhat misleading as this is not the residual at the output, but rather a measure based on the assumption of power wise summation of aliasing/imaging components. However, [dH01] describes the error this way, so to avoid confusion the name is also used here.



# Prototype Filter Design by Minimisation of Error Functions

---

To find the optimal prototype filters for the filter bank the error functions in section 2.5 are minimised.

Most of the derivations used to find the minimum of the different error functions are excluded from this chapter. The details of the derivations are available in appendix A.

The minimisation method proposed in [dH01] will be used to minimise the error functions. This method provides a simple least squares solution. All derivations of the error functions have been redone as the modulation offset of the analysis filters,  $\tau_t$ , were not applied in [dH01].

To simplify the error functions, and the minimisation of these, the  $\mathcal{Z}$ -transform of the filters can be written in vector notation

$$H_k(z) = \mathbf{h}^T \phi_h(zW_K^k) W_K^{\tau_t k} \quad (3.1)$$

$$G_k(z) = \mathbf{g}^T \phi_g(zW_K^k) \quad (3.2)$$

where

$$\begin{aligned}\mathbf{h} &= [h_0(0), h_0(1), h_0(2), \dots, h_0(L_h - 1)]^T \\ \mathbf{g} &= [g_0(0), g_0(1), g_0(2), \dots, g_0(L_g - 1)]^T \\ \phi_h(z) &= [1, z^{-1}, z^{-2}, \dots, z^{-(L_h-1)}]^T \\ \phi_g(z) &= [1, z^{-1}, z^{-2}, \dots, z^{-(L_g-1)}]^T\end{aligned}$$

The solution for minimising to minimise the error functions for the prototype filters are given by

$$\mathbf{h} = \arg \min_{\mathbf{h}} \epsilon(\mathbf{h}) \qquad \mathbf{g} = \arg \min_{\mathbf{g}} \epsilon(\mathbf{g}) \qquad (3.3)$$

where  $\epsilon(\mathbf{h})$  and  $\epsilon(\mathbf{g})$  are the error functions for the analysis and synthesis filters respectively.

Throughout the optimisation the prototype filters are constrained to only contain real values, i.e.  $\mathbf{h} \in \mathbb{R}$  and  $\mathbf{g} \in \mathbb{R}$ .

All the error functions are squared errors of the filter vectors and can be rewritten to the following matrix form

$$\epsilon(\mathbf{h}) = \mathbf{h}^T \mathbf{A} \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + c \qquad (3.4)$$

where  $\mathbf{A}$  is a matrix to form the second order term,  $\mathbf{b}$  is a vector to form the first order term and  $c$  is a constant for the zeroth order term. The same holds for the error functions for the synthesis filter.

As a squared error is convex it has a global minimum which can be found analytically by setting the derivative of the error to zero, i.e.

$$\frac{d\epsilon(\mathbf{h})}{d\mathbf{h}} = 2\mathbf{A}\mathbf{h} - 2\mathbf{b} = 0 \qquad (3.5)$$

Which gives the set of linear equations

$$\mathbf{A}\mathbf{h} = \mathbf{b} \qquad (3.6)$$

$\mathbf{A}$  is a square matrix, but not necessarily full rank, so one solution is

$$\mathbf{h} = \mathbf{A}^\dagger \mathbf{b} \qquad (3.7)$$

where  $\mathbf{A}^\dagger$  is the Moore-Penrose pseudoinverse of  $\mathbf{A}$ . The Moore-Penrose pseudoinverse give the solution with the lowest norm of  $\mathbf{h}$  which seems reasonable when  $\mathbf{h}$  is a lowpass filter.

### 3.1 Analysis Filter Design

The analysis filter is optimised according to the two error measures defined in section 2.5

- Passband response error,  $\epsilon_p$
- Inband aliasing error,  $\epsilon_a$

The passband and inband aliasing errors are solved independently and combined afterwards to a total error for the analysis filter,  $\epsilon_h$ .

Because the design method is based on minimisation of squared errors, the same filter can be obtained by other least squares FIR filter design methods, e.g. [PB87]. The MATLAB function `firls`, which is based on [PB87], can be used to design the same filter.

#### 3.1.1 Passband Response Error

Recall the passband response error

$$\epsilon_p = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \left| H_0(e^{j\omega}) - H_d(e^{j\omega}) \right|^2 d\omega \quad (3.8)$$

where  $H_d(e^{j\omega})$  is the desired total response. From now on we assume the desired response is  $H_d(e^{j\omega}) = e^{-j\omega\tau_h}$ , i.e. a magnitude response of unity and a group delay of  $\tau_h$  in the passband.

By rearranging, the following matrix form can be obtained (see appendix A.1.1)

$$\epsilon_p = \mathbf{h}^T \mathbf{A} \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + 1 \quad (3.9)$$

where  $\mathbf{A}$  is an  $L_h \times L_h$  matrix and  $\mathbf{b}$  is an  $L_h \times 1$  vector

$$\mathbf{A} = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \phi_h(e^{j\omega}) \phi_h^H(e^{j\omega}) d\omega \quad (3.10)$$

$$\mathbf{b} = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{e^{j\omega\tau_h} \phi_h(e^{j\omega})\} d\omega \quad (3.11)$$

By simplifying and solving the integrals,  $\mathbf{A}$  and  $\mathbf{b}$  equates to (see appendix A.1.2)

$$A_{p,q} = \text{sinc}\left(\frac{\omega_p}{\pi}(q-p)\right) \quad (3.12)$$

$$b_p = \text{sinc}\left(\frac{\omega_p}{\pi}(\tau_h - p)\right) \quad (3.13)$$

where  $A_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{A}$ ,  $b_p$  is the  $p$ -th row in  $\mathbf{b}$  and

$$\text{sinc}(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x}, & x \neq 0 \\ 1, & x = 0 \end{cases} \quad (3.14)$$

### 3.1.2 Inband Aliasing Error

Recall the inband aliasing error

$$\epsilon_a = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D-1} \sum_{d=1}^{D-1} \left| H_0(e^{j\omega/D} W_D^d) \right|^2 d\omega \quad (3.15)$$

By rearranging, the following matrix form can be obtained (see appendix A.2.1)

$$\epsilon_a = \mathbf{h}^T \mathbf{C} \mathbf{h} \quad (3.16)$$

where  $\mathbf{C}$  is an  $L_h \times L_h$  matrix

$$\mathbf{C} = \frac{1}{2\pi(D-1)} \sum_{d=1}^{D-1} \int_{-\pi}^{\pi} \phi_h(e^{j\omega/D} W_D^d) \phi_h^H(e^{j\omega/D} W_D^d) d\omega \quad (3.17)$$

By simplifying and solving the integral we get (see appendix A.2.2)

$$C_{p,q} = \frac{1}{D-1} (D\Delta_D[q-p] - 1) \text{sinc}\left(\frac{1}{D}(q-p)\right) \quad (3.18)$$

where  $C_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{C}$  and

$$\Delta_D[n] = \sum_{m=-\infty}^{\infty} \delta[n - mD], \quad n, m \in \mathbb{Z} \quad (3.19)$$

i.e. a Kronecker comb function with period  $D$ .

### 3.1.3 Minimising the Total Error of the Analysis Filter

The total error for the analysis filter,  $\epsilon_h$ , is obtained by adding the passband error and the inband aliasing error

$$\begin{aligned} \epsilon_h &= \epsilon_p + \epsilon_a \\ &= \mathbf{h}^T \mathbf{A} \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + 1 + \mathbf{h}^T \mathbf{C} \mathbf{h} \\ &= \mathbf{h}^T (\mathbf{A} + \mathbf{C}) \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + 1 \end{aligned} \quad (3.20)$$

By minimising the total error for the analysis filter the optimal analysis filter is obtained

$$\begin{aligned}\mathbf{h} &= \arg \min_{\mathbf{h}}(\epsilon_h) \\ &= (\mathbf{A} + \mathbf{C})^\dagger \mathbf{b}\end{aligned}\quad (3.21)$$

## 3.2 Synthesis Filter Design

The synthesis filter is optimised according to the three error measures defined in section 2.5

- Linear response error,  $\epsilon_l$
- Aliasing/imaging cancellation error,  $\epsilon_c$
- Aliasing/imaging error,  $\epsilon_r$

Like the analysis filter the errors are simplified independently and then combined to obtain the optimal synthesis filter. It is important to note that the synthesis filter depends on the analysis filter. This means that the optimal synthesis filter will change if the analysis filter is changed.

### 3.2.1 Linear Response Error

Recall the linear response error

$$\epsilon_l = \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_l(e^{j\omega}) - T_d(e^{j\omega})|^2 d\omega \quad (3.22)$$

where  $T_d(e^{j\omega})$  is the desired total response. From now on we assume the desired response is  $T_d(e^{j\omega}) = e^{-j\omega\tau_t}$ , i.e. a magnitude response of unity and a total group delay of  $\tau_t$ .

This can be written in matrix form (see appendix A.3.1)

$$\epsilon_l = \mathbf{g}^T \mathbf{E} \mathbf{g} - 2\mathbf{g}^T \mathbf{f} + 1 \quad (3.23)$$

where  $\mathbf{E}$  is an  $L_g \times L_g$  matrix and  $\mathbf{f}$  is an  $L_g \times 1$  vector

$$\mathbf{E} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \quad (3.24)$$

$$\mathbf{f} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Re \left\{ e^{j\omega \tau_t} \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \right\} d\omega \quad (3.25)$$

By solving the integrals  $\mathbf{E}$  and  $\mathbf{f}$  simplifies to (see appendix A.3.2)

$$E_{p,q} = \frac{K^2}{D^2} \sum_{c=-\infty}^{\infty} h_0[cK + \tau_t - p] h_0[cK + \tau_t - q] \quad (3.26)$$

$$f_p = \frac{K}{D} h_0[\tau_t - p] \quad (3.27)$$

where  $E_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{E}$ ,  $f_p$  is the  $p$ -th row in  $\mathbf{f}$  and  $h_0[n] = 0$  when  $n \neq 0, 1, \dots, L_h - 1$ .

### 3.2.2 Aliasing/Imaging Cancellation Error

Recall the aliasing/imaging cancellation error

$$\epsilon_c = \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_c(e^{j\omega})|^2 d\omega \quad (3.28)$$

This can be written in matrix form (see appendix A.4.1)

$$\epsilon_c = \mathbf{g}^T \mathbf{Q} \mathbf{g} \quad (3.29)$$

where  $\mathbf{Q}$  is an  $L_g \times L_g$  matrix

$$\mathbf{Q} = \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l W_D^d) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \quad (3.30)$$

This simplifies to (see appendix A.4.2)

$$Q_{p,q} = E_{p,q}(D\Delta_D[p-q] - 1) \quad (3.31)$$

where  $Q_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{Q}$  and  $E_{p,q}$  is from (3.26).

### 3.2.3 Aliasing/Imaging Error

Recall the aliasing/imaging error

$$\epsilon_r = \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_r(e^{j\omega})|^2 d\omega \quad (3.32)$$

Inserting  $T_r(e^{j\omega})$  and expanding the quadratic form gives

$$\begin{aligned} \epsilon_r = \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \int_{-\pi}^{\pi} & H_k(e^{j\omega} W_D^d) G_k(e^{j\omega}) \\ & H_k^*(e^{j\omega} W_D^d) G_k^*(e^{j\omega}) d\omega \end{aligned} \quad (3.33)$$

This can be written in matrix form (see appendix A.5.1)

$$\epsilon_r = \mathbf{g}^T \mathbf{P} \mathbf{g} \quad (3.34)$$

where  $\mathbf{P}$  is an  $L_g \times L_g$  matrix

$$\begin{aligned} \mathbf{P} = \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} & \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\ & \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^k W_D^d) W_K^{-\tau_t k} \phi_g^H(e^{j\omega} W_K^k) d\omega \end{aligned} \quad (3.35)$$

This simplifies to (see appendix A.5.2)

$$P_{p,q} = \frac{K}{D^2} \gamma_{h_0}[p-q](D\Delta_D[p-q] - 1) \quad (3.36)$$

where  $P_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{P}$  and  $\gamma_{h_0}[m]$  is the raw autocorrelation of  $h_0[n]$  at lag  $m$ , i.e.

$$\gamma_{h_0}[m] = \sum_n h_0[n] h_0[n-m] \quad (3.37)$$

### 3.2.4 Total Error of Synthesis Filter

The total error for the synthesis filter,  $\epsilon_t$ , is obtained in a similar manner as the analysis filter. I.e.

$$\begin{aligned}\epsilon_t &= \epsilon_l + \epsilon_r + \epsilon_i \\ &= \mathbf{g}^T \mathbf{E} \mathbf{g} - 2\mathbf{g}^T \mathbf{g} + 1 + \mathbf{g}^T \mathbf{Q} \mathbf{g} + \mathbf{g}^T \mathbf{P} \mathbf{g} \\ &= \mathbf{g}^T (\mathbf{E} + \mathbf{Q} + \mathbf{P}) \mathbf{g} - 2\mathbf{g}^T \mathbf{f} + 1\end{aligned}\quad (3.38)$$

The optimal synthesis filter can be obtained by minimising  $\epsilon_t$

$$\begin{aligned}\mathbf{g} &= \arg \min_{\mathbf{g}} (\epsilon_t) \\ &= (\mathbf{E} + \mathbf{Q} + \mathbf{P})^\dagger \mathbf{f}\end{aligned}\quad (3.39)$$

## 3.3 Prototype Filter Design with Weighted Errors

When designing filter banks, they are often designed for a specific purpose. E.g. in hearing aids that do large spectral modifications, the aliasing/imaging cancellation error is not as important as the inband aliasing and aliasing/imaging error. Therefore, it can be desirable to weight the different errors in the design of the prototype filters.

For the analysis filter the  $\mathbf{A}$  matrix and  $\mathbf{b}$  vector are matched, so we take the passband response as the baseline for the weighting. Thus the inband aliasing error can be weighted compared to the passband response error. The solution for the analysis filter is then

$$\mathbf{h} = (\mathbf{A} + 10^{\alpha_a} \mathbf{C})^\dagger \mathbf{b}\quad (3.40)$$

where  $\alpha_a$  is a weighting of the inband aliasing error. Negative values of  $\alpha_a$  shift the prototype filter to have a better passband by compromising the inband aliasing, while positive values will give better inband aliasing attenuation by compromising the passband response.

A similar weighting can be applied to the synthesis filter design. In the synthesis filter design  $\mathbf{E}$  and  $\mathbf{f}$  are matched, so the weights are applied to the two other errors. The solution for the synthesis filter is then

$$\mathbf{g} = (\mathbf{E} + 10^{\alpha_r} \mathbf{Q} + 10^{\alpha_c} \mathbf{P})^\dagger \mathbf{f}\quad (3.41)$$

where  $\alpha_c$  is the weight of the aliasing/imaging cancellation error and  $\alpha_r$  is the weight of the aliasing/imaging error.

# Audibility of Artefacts & a Simple Psychoacoustic Model

---

The error measures defined in section 2.5 do not take the aspects of human hearing into account. In this chapter we look at the audibility of the artefacts and modify the error measures to be psychoacoustically motivated.

The chapter consists of three parts. The first part discuss artefacts in filter banks and the psychoacoustic concepts used to assess the audibility of the artefacts. The next part set up a simple model for frequency domain masking. In the last part, this model is applied to the relevant error measures and optimisation functions in order to obtain a filter bank design with reduced audibility of the aliasing/imaging artefacts.

## 4.1 Audibility of Filter Bank Artefacts

In section 2.5, the artefacts introduced by the filter bank were described. The audibility of these artefacts are described in this section. Only the artefacts in the total response are considered as the analysis artefacts are not directly audible.

### 4.1.1 Audibility of Linear Response Errors

The linear part of the total response can introduce artefacts in the magnitude and phase response. The magnitude error can be seen as a modulation of the magnitude over frequency. The modulation is periodic over the spacing between filters,  $f_s/K$  where  $f_s$  is the sampling frequency. The audibility of the magnitude error could be assessed by Just Noticeable Difference (JND) for magnitude response modulation. To our knowledge JND for magnitude response modulation has not been investigated.

Another approach would be to look at the magnitude modulation in the time domain. Modulation in magnitude is equivalent to addition of “side bands” in the time domain. This can also be seen in eq. (2.20) where the comb function introduces “side bands” if the convolution of  $h_0[n]$  and  $g_0[n]$  is not zero when  $n = cK + \tau_t$ , with  $c \in \mathbb{Z}$  except when  $n = 0$ . The audibility of the “side bands” could then be assessed in psychoacoustics by the time domain masking concept [Pla05, Moo12, ZF99].

The phase response of the linear part of the total transfer is probably better viewed as a group delay. The minimum audible delay and the minimum audible delay difference across frequency has to be quantified psychoacoustically.

The maximum tolerable delay depends highly on the use case and might be frequency dependent. In hearing aids with open fitting the tolerable group delay is quite low because of the interference between the direct and processed sounds [Kat04].

### 4.1.2 Audibility of Aliasing/Imaging Errors

Recall the aliasing/imaging transfer in equation (2.15) and (2.16)

$$\begin{aligned} Y_c(z) &= T_c(z)X(zW_D^d), & d = 1, 2, \dots, D-1 \\ T_c(z) &= \frac{1}{D} \sum_{k=0}^{K-1} H_k(zW_D^d)G_k(z), & d = 1, 2, \dots, D-1 \end{aligned} \quad (4.1)$$

The aliasing/imaging in the output is the sum of  $Y_c(z)$  over  $d$ .  $X(zW_D^d)$  can be viewed as a frequency shifted version of  $X(z)$  with a shift of

$$\omega_{\text{shift}}[d] = \text{princ arg}(W_D^d) \quad (4.2)$$

where  $\text{princ arg}(z)$  is the argument of  $z$  in the range from  $-\pi$  to  $\pi$ . For  $d = 1$  the frequency shift is

$$\omega_{\text{shift}}[1] = \text{princ arg}(W_D) = -2\pi/D \quad (4.3)$$

The audibility of a frequency shifted signal can be assessed by frequency domain masking [Pla05, Moo12, ZF99]. This will be considered further in the next section.

Another approach would be to look at the aliasing/imaging components in time domain. Frequency shifting by  $W_D^d$  is an amplitude modulation of the time domain signal by  $W_D^{-nd}$ . The modulation depth at a specific frequency and  $d$  is determined by the transfer function  $T_c(z)$ . To assess the audibility of the amplitude modulation, modulation detection thresholds could be utilised [Moo12, DKK97a].

The above analysis is only valid when no processing is performed in the filter bank, i.e. when  $F_k(z) = 1$ . When processing is performed, there will no longer be cancellation between bands, so we need to look at the estimate of magnitude transfer of the aliasing/imaging components without cancellation, i.e.  $T_r(z)$ .

The same principles apply for  $T_r(z)$  as for  $T_c(z)$ . We can either look at the frequency shifted input signals or the time domain amplitude modulated signals.

### **4.1.3 Approach for Psychoacoustic Optimisation of the DFT Modulated Filter Bank**

Different psychoacoustic concepts can be used for assessing the audibility of the artefacts. In this thesis we focus on frequency domain masking to model the audibility of the aliasing/imaging artefacts. In the next section a simple model based on the power spectrum model of masking with ROEX filters is introduced. Afterwards, the model is used to weight the optimisation functions to obtain a filter bank with less audible aliasing/imaging artefacts.

## **4.2 Definitions, Assumptions & a Simple Model of Frequency Masking in the Auditory System**

In this section we first go through the power spectrum model and the ROEX filter to approximate the shapes of the auditory filters. Afterwards, ERB and

ROEX are combined to obtain a simple auditory filter model. This model is used to weight the aliasing/imaging errors in the design of prototype analysis and synthesis filters for a DFT modulated filter bank.

Frequency resolution in the human hearing has been estimated by various masking experiments. The results show that a sound is most effectively masked by other sounds containing frequencies close to the original and that the masking pattern changes with frequency [Moo12, Pla05, ZF99]. This has led to the concept of a non-uniformly distributed bank of filters to model the observed behaviour. These filters are called auditory filters [Moo12].

#### 4.2.1 The Concept of Masking

The limitations of the human hearing has been an active research topic for many years. When dealing with the limitations of the human hearing from a psychoacoustic perspective the concept of masking is essential. Masking has been defined in [AAUA60] as

1. Masking is the process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound.
2. Masking is the amount by which the threshold of audibility of a sound is raised by the presence of another (masking) sound. The unit customarily used is the decibel.

#### 4.2.2 The Power Spectrum Model of Frequency Masking

When detecting a tone in noise an auditory filter with centre frequency close to the tone is used, i.e. the filter with the best Signal to Noise Ratio (SNR). Some of the noise will eventually also pass through the auditory filter. Only the noise passing through the filter will contribute to the masking of the tone. This concept is defined as the power spectrum model [Pat86].

Depending on the listener and the experiment a certain SNR between the test tone and the sound passing the auditory filter is required in order to hear the test tone. This can be described as

$$\kappa = \frac{P_s^{post}}{N^{post}}$$

where  $P_s$  and  $N$  are the long term power of the signal and noise respectively,  $\kappa$  is the SNR where the signal is just audible and the superscript *post* denotes

that the values are after filtering by the auditory filter.  $\kappa$  is typically around 0.4, but varies from one person to another [Moo12].

The long term power required for a signal to be audible can be defined by  $k$  and the amount of noise in the auditory filter

$$P_s^{post} = \kappa \int_0^\infty W(f)N(f) df \quad (4.4)$$

where  $W(f)$  is the frequency dependent weighting by the auditory filter and  $N(f)$  is the power spectral density of the noise at frequency  $f$ .

If we define the normalisation of the filter to have unity gain at the frequency of the signal then the power of the signal is the same before and after filtering, i.e.  $P_s = P_s^{post}$ . We will assume this normalisation from now on. This means that the power required for the signal to be just audible is

$$P_s = \kappa \int_0^\infty W(f)N(f) df \quad (4.5)$$

The power spectrum model use the long-term power of both signal and masker. This means that the model can not be used for modelling masking with fluctuations in the SNR.

### 4.2.3 Simple Auditory Filter Shape

Different filters have been proposed to model the shape and behaviour of the auditory filters. Some of the most used are ROEX [PNSWM82], Gammatone [PNSHR87], dual resonance [LPM01] and filter cascades [Lyo11]. In this thesis the most simple ROEX filter will be used due to its simplicity. The ROEX has some limitations compared to other filters, but as a proof of concept it is sufficient.

The one parameter ROEX filter is given by

$$W_{roex}(g) = (1 + pg)e^{-pg}, g \geq 0 \quad (4.6)$$

where  $g$  is the normalised distance to the centre frequency,  $f_c$ , i.e.  $g = \frac{|f-f_c|}{f_c}$ . The parameter  $p$  can be fitted to measured data and determines the width of the filter, i.e. the bandwidth and slope of the skirts.

As the frequency response of the ROEX filter is defined by the normalised distance to the centre frequency the bandwidth of the filters increases proportionally with the centre frequency.

#### 4.2.4 Bandwidth of Auditory Filters across Centre Frequency

As with the ROEX-filter, the bandwidth of the auditory filters increase as a function of centre frequency [Moo12]. Usually when dealing with bandwidths of auditory filters the ERB is used. ERB is defined as the bandwidth of a rectangular filter with the same power transfer as an auditory filter. The ERB is related to the centre frequency by [GM90]

$$ERB(f_c) = 24.7 + 0.108f_c \quad (4.7)$$

ERB is usually measured by the notched noise method using the stimuli shown in figure 4.1 [GM90]. The main reason for using notched noise is to avoid off-frequency listening. Off-frequency listening occurs when an auditory filter which is not centred at  $f_c$  has a better SNR than the auditory filter centered at  $f_c$ . Furthermore, the shape of the auditory filter is assumed to be symmetric. This assumption is widely accepted when measuring at low levels [Pat86]. For higher levels the filters widen at the low frequency side [GM00]. This results in a raised masking threshold for frequencies higher than the test tone, i.e. upwards spread of masking. For the rest of the report we assume that the filters are symmetric and measured at low levels as this is the situation with least masking, i.e. the worst case situation for audibility of filter bank artefacts.

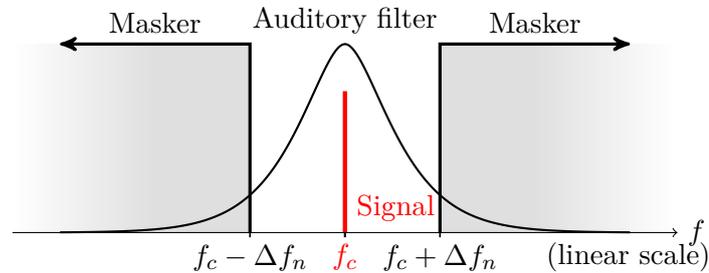


Figure 4.1: Stimulus used in the Notched-noise method. Wideband noise masker with a notch centered around the test signal frequency. The notched-noise method is used to measure the shape of auditory filters. The filters are assumed symmetric and a notched noise (instead of one-sided) is used to avoid off-frequency listening.

In figure 4.2, ERB is compared to the bandwidth of  $1/6$ -octave spaced filters. If a bank of ROEX filters are made with a constant  $p$  value, the bandwidth will increase linearly, thus give logarithmic spaced filters. According to the ERB, this seems to hold for the auditory filters at high frequencies. For low frequencies the frequency resolution of the  $1/6$ -octave filters will approach infinite precision, which is not the case for auditory filters. Although the bandwidth of the auditory filters do not follow the bandwidth of the ROEX-filters with constant  $p$ , the shape of the filters are still correct. The change

in bandwidth can thus be modelled by making  $p$  frequency dependant with lower values at lower frequencies.

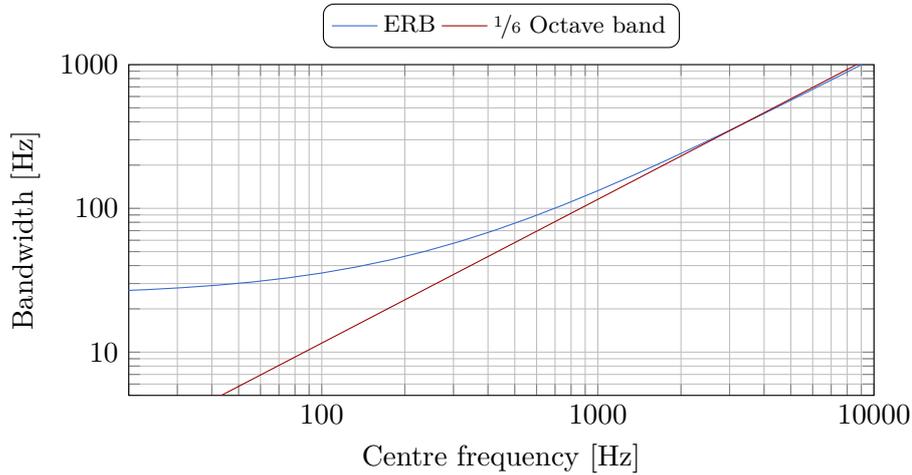


Figure 4.2: Auditory filter bandwidth as a function of centre frequency.

#### 4.2.5 Simple Auditory Filter Model

To obtain a simple model of the auditory filters which accounts for both the shape and the change in bandwidth as a function of frequency the ROEX filter and the ERB can be combined. The ROEX filters are related to ERB by [PNSWM82]

$$ERB(f_c) = \frac{4f_c}{p} \quad (4.8)$$

An auditory filter with centre frequency  $f_c$ , can therefore be modelled by

$$\hat{W}(f) = \left(1 + \frac{4|f - f_c|}{24.7 + 0.108f_c}\right) e^{-\frac{4|f - f_c|}{24.7 + 0.108f_c}} \quad (4.9)$$

### 4.3 Psychoacoustic Model for Aliasing/Imaging Artefacts

In order to apply the model to the aliasing/imaging artefacts, the frequency shift of the aliasing/imaging components need to be translated to a threshold of audibility. As mentioned in section 4.1.2, the aliasing/imaging components

are shifted by  $\text{princ arg}(W_D^d)$ . The frequency shift in Hertz are given by

$$f_{\text{shift}}[d] = \text{princ arg}(W_D^d) \frac{f_s}{2\pi} \quad (4.10)$$

The auditory filter as a function of  $d$  is then

$$\hat{\beta}[d] = \left(1 + \frac{4|f_{\text{shift}}[d]|}{24.7 + 0.108f_c}\right) e^{-\frac{4|f_{\text{shift}}[d]|}{24.7+0.108f_c}} \quad (4.11)$$

Assuming the input signal to the filter bank is a tone and no off-frequency listenings occurs, we can describe the audibility of each aliasing/imaging component by the auditory filter when using the input signal as masker. This is illustrated in figure 4.3.

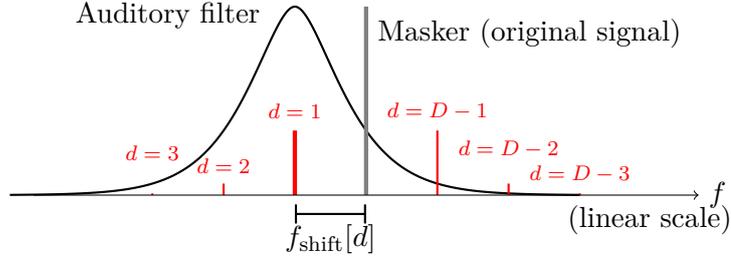


Figure 4.3: Masking of aliasing/imaging components (red) by the original signal. The aliasing/imaging components are considered separately. This illustration is for the aliasing/imaging component at  $d = 1$ .

According to (4.2.2), the aliasing/imaging component for a specific  $d$  is on the limit of audibility when the SNR of the aliasing/imaging to original signal is equal to  $\kappa$ . I.e.

$$\kappa = \frac{P_{\text{aliasing/imaging}}^{\text{post}}[d]}{P_{\text{original signal}}^{\text{post}}} = \frac{P_{\text{aliasing/imaging}}[d]}{P_{\text{original signal}} \hat{\beta}[d]} \quad (4.12)$$

This means that the aliasing/imaging component is on the limit of audibility when the power of it is

$$P_{\text{aliasing/imaging}}[d] = \kappa \hat{\beta}[d] P_{\text{original signal}} \quad (4.13)$$

In practice the above threshold of audibility,  $\kappa \hat{\beta}[d]$ , is not realistic as the human auditory system have a limited dynamic range. One way to circumvent this problem is to incorporate the absolute threshold of audibility, but as the signal level is unknown, this is not easily done. Another approach is to use a more complex model for the ROEX filter like the two or three parameter ROEX [PNSWM82].

The dynamic range of the system where the filter bank is to be used could also be considered. In most embedded real-time audio systems Digital-to-Analog Converters (DACs) of 16 bit precision is used. A 16 bit DAC has a theoretic dynamic range of approximately 98 dB, though in practice it is often lower [PM07]. The threshold of audibility is therefore truncated to never go below  $-94$  dB. I.e.

$$\beta[d] = \max(\kappa\hat{\beta}[d], 10^{-94/10}) \quad (4.14)$$

As the DFT modulated filter bank is uniform, the model has to be simplified further to make a single masking threshold for all filters. Therefore, a single centre frequency,  $f_c$ , is chosen. Auditory filters below that frequency will be too wide and those above will be too narrow.

#### 4.3.1 Applying Weights to Error Functions

The aliasing/imaging error functions defined in section 2.5.2 can be weighted with the threshold of audibility to make perceptual error functions. Because the threshold of audibility is build on the power spectrum model of masking, it describes the threshold of audibility for the power of the signals, so when applying the model to the error functions, it should be done in power, i.e. on the squared errors.

The errors are weighted with the inverse of the threshold of audibility because when the threshold is low the audibility is high, i.e.

$$w[d] = \frac{1}{\beta[d]} \quad (4.15)$$

where  $w[d]$  is the weighting of the  $d$ -th aliasing/imaging component error. Furthermore, when the weighting is applied it is normalised to have a mean value of unity to retain the internal weighting between the different error functions

$$\epsilon_{wc} = \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_c(e^{j\omega})|^2 d\omega \quad (4.16)$$

$$\epsilon_{wr} = \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_r(e^{j\omega})|^2 d\omega \quad (4.17)$$

where  $\epsilon_{wc}$  is the weighted aliasing/imaging cancellation error,  $\epsilon_{wr}$  is the weighted aliasing/imaging error without cancellation and  $\bar{w}$  is the mean value of  $w[d]$  over  $d$ .

## 4.4 Minimisation of Psychoacoustically Weighted Errors

The minimisation of the psychoacoustically weighted errors is done in similar manner as the original error functions. For details see chapter 3 and A.

### 4.4.1 Psychoacoustically Weighted Aliasing/Imaging Cancellation Error

The psychoacoustically weighted aliasing/imaging cancellation error is defined as

$$\epsilon_{wc} = \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_c(e^{j\omega})|^2 d\omega \quad (4.18)$$

this can be rewritten to matrix form

$$\epsilon_{wc} = \mathbf{g}^T \mathbf{S} \mathbf{g} \quad (4.19)$$

where  $\mathbf{S}$  is an  $L_g \times L_g$  matrix

$$\begin{aligned} \mathbf{S} = & \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ & \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\ & \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l W_D^d) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \end{aligned} \quad (4.20)$$

which simplifies to

$$S_{p,q} = E_{p,q} \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} W_D^{(p-q)d} \quad (4.21)$$

where  $S_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{S}$  and  $E_{p,q}$  is from (3.26).

### 4.4.2 Psychoacoustically Weighted Aliasing/Imaging Error

The psychoacoustically weighted aliasing/imaging error is defined as

$$\epsilon_{wr} = \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_r(e^{j\omega})|^2 d\omega \quad (4.22)$$

this can be rewritten to matrix form

$$\epsilon_{wr} = \mathbf{g}^T \mathbf{U} \mathbf{g} \quad (4.23)$$

where  $\mathbf{U}$  is an  $L_g \times L_g$  matrix

$$\begin{aligned} \mathbf{U} = & \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ & \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\ & \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^k W_D^d) W_K^{-\tau_t k} \phi_g^H(e^{j\omega} W_K^k) d\omega \end{aligned} \quad (4.24)$$

which simplifies to

$$U_{p,q} = \frac{K}{D^2} \gamma_{h_0}[p-q] \sum_{d=1}^{D-1} \frac{w[d]}{\bar{w}} W_D^{(p-q)d} \quad (4.25)$$

where  $U_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{U}$  and  $\gamma_{h_0}[m]$  is the raw autocorrelation of  $h_0[n]$  at lag  $m$ .

#### 4.4.3 Total Psychoacoustically Weighted Error of Synthesis Filter

The total psychoacoustically weighted error for the synthesis filter,  $\epsilon_{wt}$ , is obtained by substituting the original errors in eq. (3.38) with the psychoacoustically weighted equivalents

$$\epsilon_{wt} = \mathbf{g}^T (\mathbf{E} + \mathbf{S} + \mathbf{U}) \mathbf{g} - 2\mathbf{g}^T \mathbf{f} + 1 \quad (4.26)$$

The psychoacoustically weighted optimised prototype synthesis filter can then be calculated as

$$\mathbf{g} = (\mathbf{E} + \mathbf{S} + \mathbf{U})^\dagger \mathbf{f} \quad (4.27)$$

The weighted error function is then defined as

$$\mathbf{g} = (\mathbf{E} + 10^{\alpha_{wr}} \mathbf{U} + 10^{\alpha_{wc}} \mathbf{S})^\dagger \mathbf{f} \quad (4.28)$$

where  $\alpha_{wc}$  is the weight of the aliasing/imaging cancellation error with psychoacoustic weighting and  $\alpha_{wr}$  is the weight of the aliasing/imaging error with psychoacoustic weighting.



# Example of Filter Designs with & without Psychoacoustic Weighting

---

In this chapter example designs obtained by the optimisation algorithm with and without psychoacoustic weighting are presented. The influence of the  $\alpha$  values for the optimisation is also investigated.

## 5.1 Example Filter Design without Psychoacoustic Weighting

To give an idea of the filters designed by the proposed method an example design with the parameters listed in table 5.1 and 5.2 is shown in figure 5.1, 5.2, 5.3, 5.4 and 5.5.

Filter Bank Parameter	$K$	$D$	$L_h$	$L_g$	$\tau_h$	$\tau_t$	$f_s$
Value	16	$\frac{K}{2}$	$2K$	$2K$	$\frac{L_h-1}{2}$	$\frac{L_g-1}{2} + \tau_h$	8 kHz

Table 5.1: Parameters for the filter bank design example. The optimisation parameters for this design example are shown in table 5.2.

In the example design the target group delay is set so both the analysis

#### 44 Example of Filter Designs with & without Psychoacoustic Weighting

Optimisation Parameters	$\alpha_a$	$\alpha_r$	$\alpha_c$	$f_c$	$\kappa$	$\alpha_{wr}$	$\alpha_{wc}$
Value	0	0	0	-	-	-	-

Table 5.2: Parameters for the optimisation of filter bank design example. The filter bank parameters for this design example are shown in table 5.1.

and synthesis filter become symmetric and therefore are linear phase<sup>1</sup>, i.e.  $\tau_h = L_h - 1/2$  and  $\tau_t = L_g - 1/2 + \tau_h$ . The impulse responses of the analysis and synthesis filters are shown in figure 5.1. They are symmetric as expected with a group delay of  $L_h - 1/2$  and  $L_g - 1/2$ . Changing the group delay will result in filters that are not symmetric. This can be desirable in situations where the total delay of the filter bank is constrained, but where the increased filter length still provides better frequency selectivity.

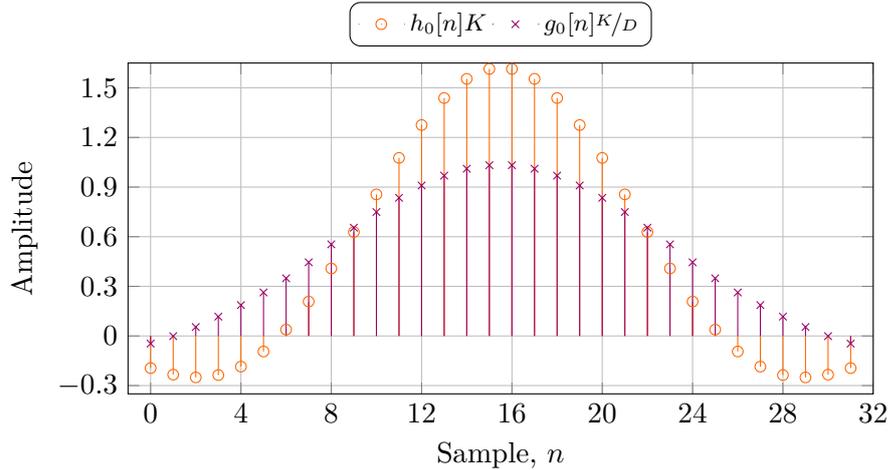


Figure 5.1: Impulse response of prototype analysis and synthesis filters with the parameters given in table 5.1 and 5.2.

Figure 5.2 show the magnitude responses of the analysis and synthesis filters. Both filters have the expected lowpass shape, with a relatively good stopband attenuation. The analysis filter has a relatively flat response in the passband while the synthesis filter is much narrower. This is because the error function for the analysis filter includes the passband response, while the error functions for the synthesis filter only focus on the total response. Thus the synthesis filter have the response needed to make the linear response flat after summation over bands.

Figure 5.3 show the magnitude response of the linear part of the total transfer function,  $T_l(f)$  (2.14), the aliasing/imaging part of the total transfer

<sup>1</sup>A symmetric FIR filter with an even number of taps,  $L$ , has a group delay of  $L-1/2$ .

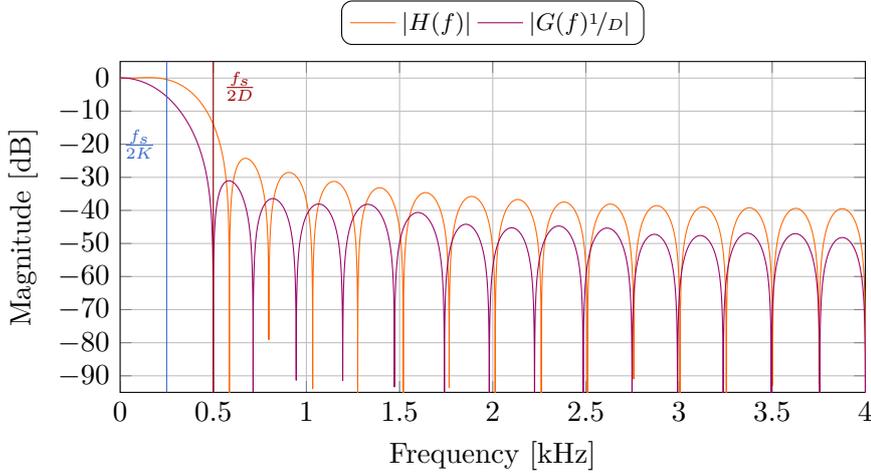


Figure 5.2: Magnitude response of prototype analysis and synthesis filters with the parameters given in table 5.1 and 5.2.

function for each  $d$ ,  $T_c(f)$  (2.16), and the power wise summation of the aliasing/imaging part of the total transfer function for each  $d$ ,  $T_r(f)$  (2.17). The plots only show one period of the spectrum as it repeats itself for every  $f_s/K$ . The linear part has a transfer of approximately 0 dB.

The aliasing/imaging transfer functions describe the transfers for frequency shifted versions of the input signal. The frequency shift is given by (4.2). The aliasing/imaging with cancellation is generally lower than the aliasing/imaging without cancellation. It also exhibits typical cancellation dips in the response, but do not cancel at all frequencies. This means that PR is not obtained in this situation. This is because of the compromise between cancelling aliasing/imaging and attenuating aliasing/imaging. For applications with heavy processing in the filter bank, attenuation is more important than cancellation and vice versa. As will be shown in the next section, the compromise can be controlled by the  $\alpha_c$  and  $\alpha_r$  parameters.

To give an overview of the aliasing/imaging transfer for different frequency shifts, the mean and peak transfer are shown in figure 5.4. The cancellation transfer is generally lower than the transfer without cancellation. Both transfers are highest for small frequency shifts. This is because the filters attenuates most far from the passband.

The overall errors are shown in figure 5.5. The total error,  $\epsilon_t$ , is clearly controlled by the aliasing/imaging error without cancellation,  $\epsilon_r$ . The linear error,  $\epsilon_l$ , is quite low, which we could also see in first plot in figure 5.3.

## 46 Example of Filter Designs with & without Psychoacoustic Weighting

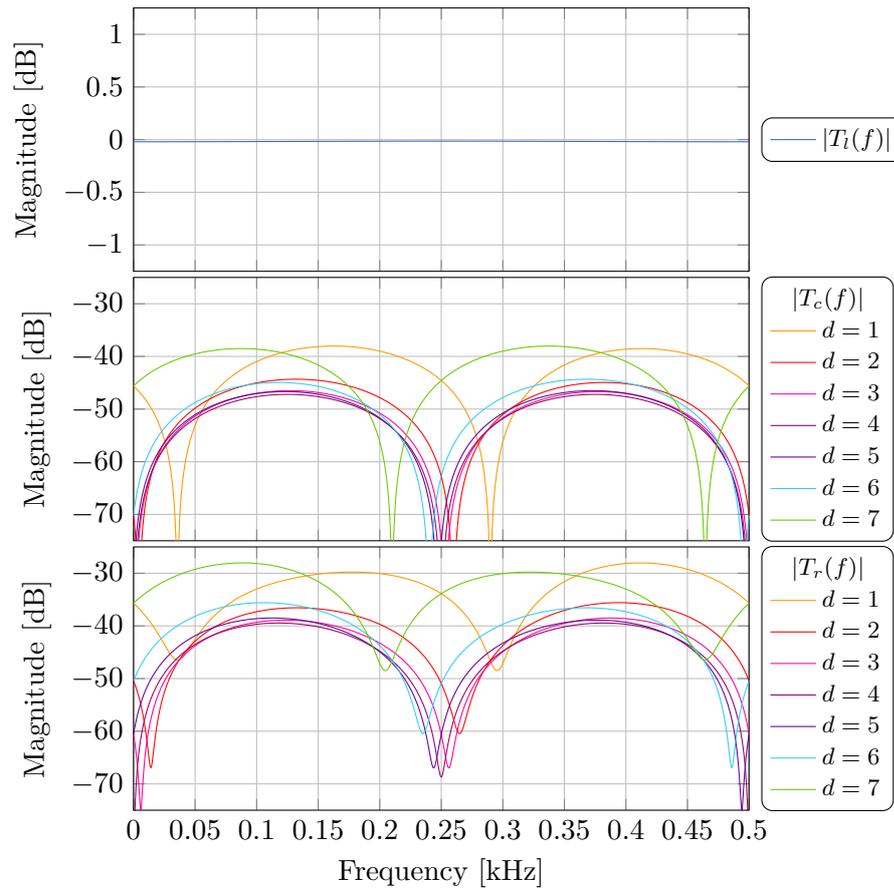


Figure 5.3: Magnitude response of the linear part of the total transfer function,  $T_l(f)$  (2.14), the aliasing/imaging part of the total transfer function for each  $d$ ,  $T_c(f)$  (2.16), and the power wise summation of the aliasing/imaging part of the total transfer function for each  $d$ ,  $T_r(f)$  (2.17). The plots only show one period of the spectrum as it repeats itself for every  $f_s/K$ . The parameters for the design is shown in table 5.1 and 5.2.

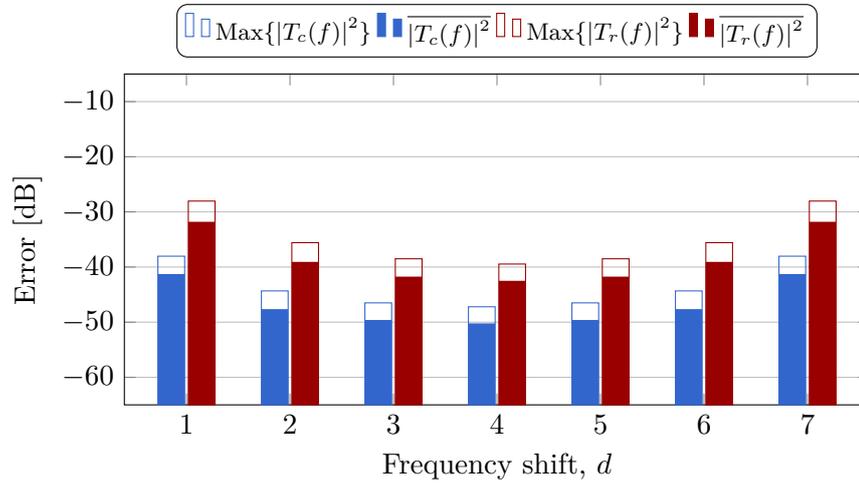


Figure 5.4: Maximum and integrated power of the aliasing/imaging components for each  $d$ . The aliasing/imaging with cancellation is lower than without. The optimisation find the lowest overall power of the aliasing/imaging components, but do not distribute them evenly. This is because it is “easier” to attenuate frequencies far away from the passband.

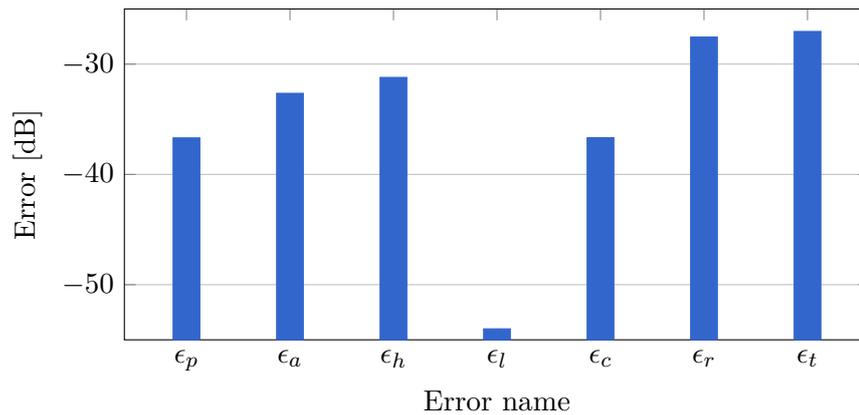


Figure 5.5: Error measures for filter bank design with the parameters in table 5.1 and 5.2. The total analysis error,  $\epsilon_h$ , is mainly controlled by the inband aliasing,  $\epsilon_a$ . The total error,  $\epsilon_t$ , is mainly controlled by the aliasing/imaging without cancellation,  $\epsilon_r$ . The linear response error,  $\epsilon_l$ , is very low.

## 48 Example of Filter Designs with & without Psychoacoustic Weighting

### 5.1.1 Influence of Optimisation Parameters

The above example was with all  $\alpha$  parameters set to zero. In figure 5.6 the errors are shown for different values of  $\alpha_a$ , all other parameters are the same as in table 5.1 and 5.2. In the upper plot the minimum of the analysis error,  $\epsilon_h$ , is at  $\alpha_a = 0$ . This is expected as this is the error that is minimised by the analysis filter design method. Lower  $\alpha_a$  lowers the passband error, but increases the stopband error and vice versa. The passband error approaches 0 dB for very high  $\alpha_a$  values. This is caused by reducing the overall gain of the analysis filter.

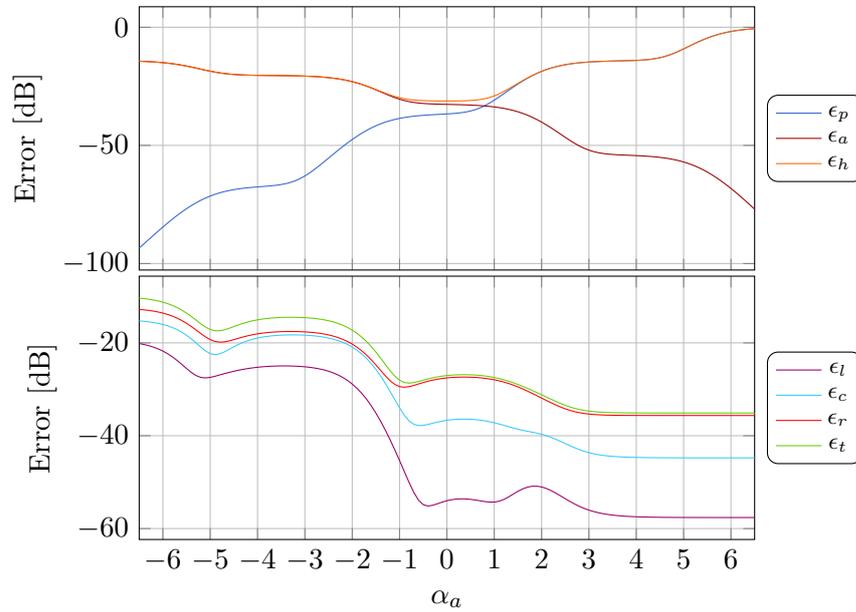


Figure 5.6: Error measures as a function of  $\alpha_a$ . The total analysis error,  $\epsilon_h$ , is lowest at  $\alpha_a = 0$ . The passband error,  $\epsilon_p$ , approaches zero dB for  $\alpha_a \rightarrow \infty$ . This is because, to attenuate aliasing, the gain of the analysis filter is reduced. The total error,  $\epsilon_t$ , is lowest for high values of  $\alpha_a$ . The gain of the synthesis filter compensates for the attenuation in the analysis filter.

In the lower plot of figure 5.6 the errors for the synthesis filter design are shown for different values of  $\alpha_a$ . Note that  $\alpha_a$  is only used in the analysis filter design, and therefore only has an indirect effect on the synthesis filter design errors. It is worth noting that the total error is lowest at higher  $\alpha_a$  values. In that situation the passband of the analysis filter is compromised for increased aliasing attenuation.

When adjacent filter bands add the total response becomes higher in the crossovers if this is not compensated for in the synthesis filters. Thus to make

the magnitude of the linear part of the total response flat, the passband of the synthesis filters are reduced to compensate for the response of the analysis filters. This is also the reason for some window design methods, where the analysis and synthesis filters are the same, to aim for 3 dB attenuation at the crossover frequencies [CRALMBL02]. Power complementary, and therefore also paraunitary, filters inherently have this property. It can even be argued that a narrower analysis filter, e.g. magnitude complementary, and a wider synthesis filter makes a better compromise because the imaging originating from the don't care region of the analysis filter is reduced, i.e. the blue lines in figure 2.5. For a discussion of this see [CR83, sec. 7.3.2].

In figure 5.7 the filters obtained with  $\alpha_a = 4$  is shown. The analysis filter is now considerably narrower than the one shown in figure 5.2, and the synthesis filter is wider. The stopband attenuation of the synthesis filter is uneven with dips around even multiples of the Nyquist frequency of the downsampled signal,  $f_s/2D$ , and bumps around the odd multiples of the Nyquist frequency of the downsampled signal. Because of the increased don't care attenuation of the analysis filter, the imaging components will have dips at odd multiples of the downsampled Nyquist frequency and therefore the synthesis filter need not attenuate these regions as much. So a better compromise is found by attenuate the peaks found around the even multiples of the downsampled Nyquist frequency.

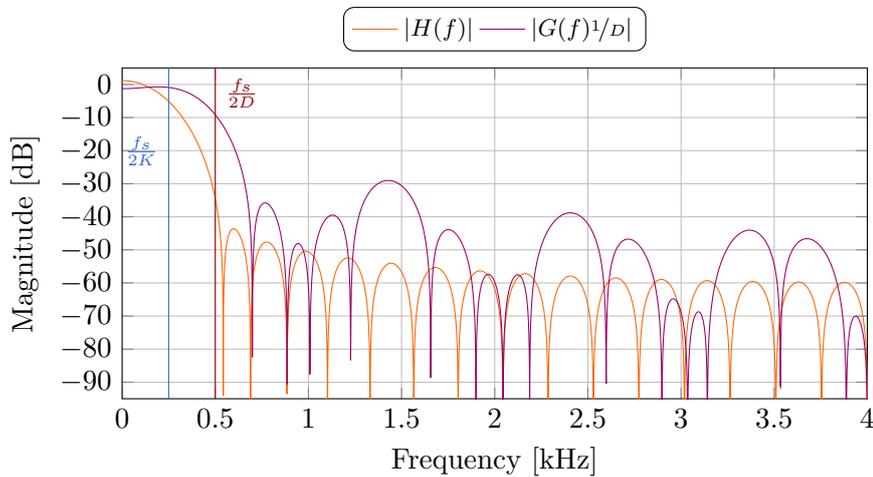


Figure 5.7: Magnitude response of prototype analysis and synthesis filters with the parameters given in table 5.1 and 5.2 except for  $\alpha_a$  which is 4.

In figure 5.8 the influence of  $\alpha_r$  and  $\alpha_c$  on the synthesis filter errors are shown. There is a trade-off between aliasing/imaging cancellation,  $\epsilon_c$ , and aliasing/imaging attenuation,  $\epsilon_r$ , directly influenced by the two parameters  $\alpha_c$  and  $\alpha_r$ . The error functions used in [dH01] corresponds to setting the

## 50 Example of Filter Designs with & without Psychoacoustic Weighting

weighting of the cancellation error,  $\alpha_c$ , to  $-\infty$ , i.e. removing the error function from the optimisation. This results in filters with lower aliasing/imaging error,  $\epsilon_r$ , but higher aliasing/imaging cancellation error,  $\epsilon_c$ . The aliasing/imaging error assumes power wise summation of alising/imaging components between bands. When  $\epsilon_c$  is higher than  $\epsilon_r$  it shows that this assumption is actually wrong. Although not clearly visible in figure 5.8, this is the case when  $\alpha_r$  is substantially higher than  $\alpha_c$ . So when only using the error functions used in [dH01], the aliasing/imaging will actually be worse than  $\epsilon_r$  predicts when no processing is performed.

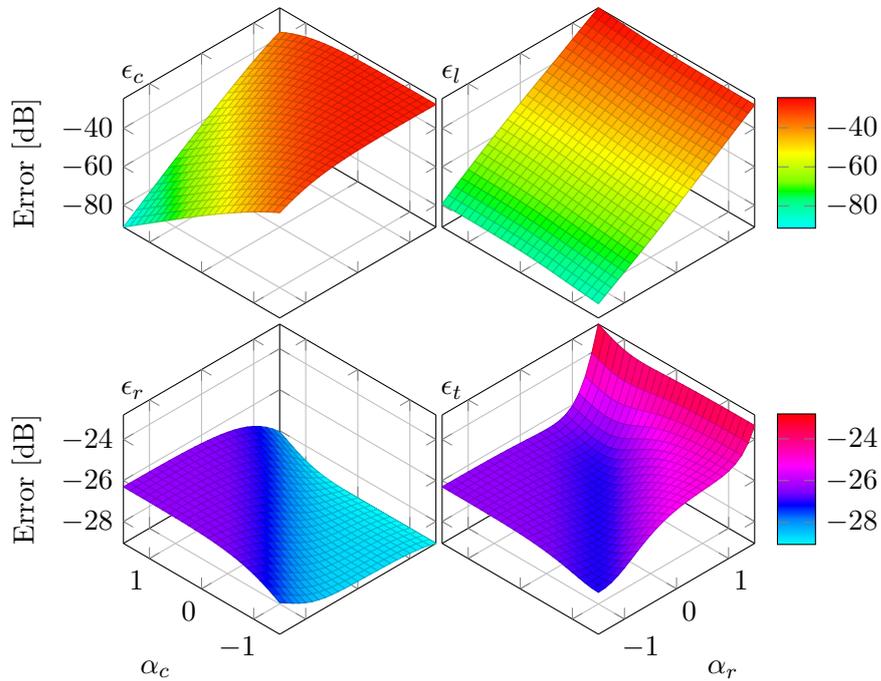


Figure 5.8: Synthesis error measures as a function of  $\alpha_c$  and  $\alpha_r$ . In the diagonal,  $\alpha_c = \alpha_r$ , the linear error,  $\epsilon_l$  is traded for a slight decrease in the aliasing/imaging error,  $\epsilon_r$ . Higher values of  $\alpha_c$  do not increase the aliasing/imaging error,  $\epsilon_r$ , nearly as much as it decreases the aliasing/imaging cancellation error,  $\epsilon_c$ . Therefore, it seems like a good tradeoff to increase  $\alpha_c$  in situations where PR is desired, but not required.

## 5.2 Example Filter Design with Psychoacoustic Weighting

An example design of the proposed method with psychoacoustic weighting is presented to give an idea of the influence of the weighting. The filter bank parameters for the design are the same as in the example design without

psychoacoustic weighting. The optimisation parameters are listed in table 5.3 and the design is shown in figure 5.9, 5.10, 5.11 and 5.12.

Optimisation Parameters	$\alpha_a$	$\alpha_r$	$\alpha_c$	$f_c$	$\kappa$	$\alpha_{wr}$	$\alpha_{wc}$
Value	0	-	-	8 kHz	0.4	0	0

Table 5.3: Parameters for the optimisation of filter bank design example with psychoacoustic weighting. The filter bank parameters for this design example are shown in table 5.1.

The centre frequency for the auditory filter in the masking model,  $f_c$ , is set very high in order to emphasise the impact of the psychoacoustic model. This is necessary because the small number of bands compared to the sample rate results in frequency bands wider than auditory filters at low frequencies.

In figure 5.9 the impulse responses of the prototype filters for the psychoacoustically optimised filter bank are shown. The filters are still linear phase with the same group delay as in the example design without psychoacoustic weighting. The analysis filter is the same as in the example design without psychoacoustic weighting, as the optimisation of the analysis filter is not weighted. The synthesis filter look quite different compared to the other design.

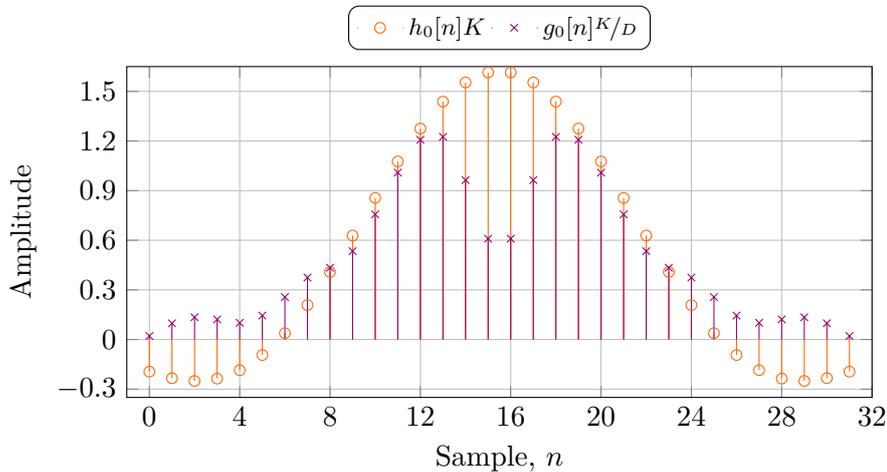


Figure 5.9: Impulse response of prototype analysis and synthesis filters with the parameters given in table 5.1 and 5.3.

In figure 5.10 the magnitude spectrums of the prototype analysis and synthesis filters are shown. The analysis filter is still the same as in the design without psychoacoustic weighting. The synthesis filter is a lot different. Instead of the lowpass shape obtained in the other design the synthesis filter is like a staircase function. For an ordinary interpolation filter this design is not good

## 52 Example of Filter Designs with & without Psychoacoustic Weighting

due to the limited stop band attenuation. According to the psychoacoustic model the artefacts shifted far in frequency are more audible than artefacts with small frequency shifts. The synthesis filter attenuate the far end more than the design without psychoacoustic weighting. This extra attenuation is achieved by compromising the stopband close to the passband, but these artefacts should not be as audible.

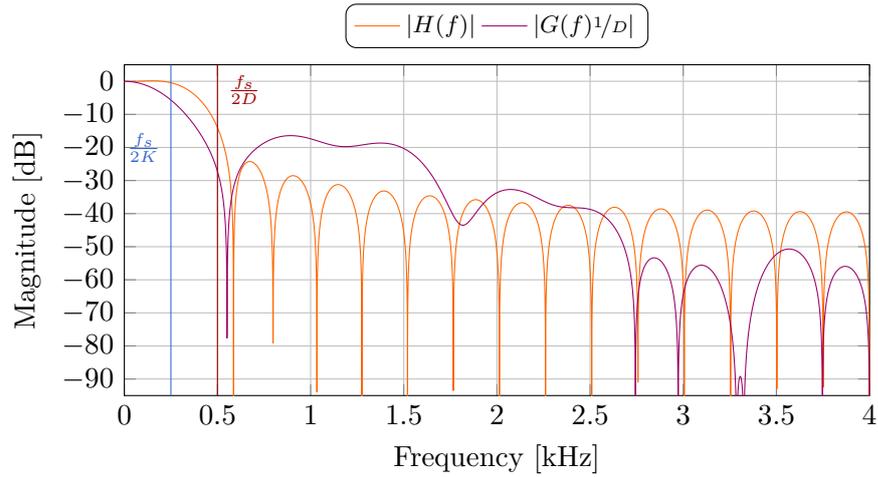


Figure 5.10: Magnitude response of prototype analysis and synthesis filters with the parameters given in table 5.1 and 5.3. The synthesis filter focus on attenuating frequencies far away from the passband most.

In figure 5.11 the peak and average transfer of the different aliasing/imaging components are shown. The threshold of audibility for the aliasing/imaging components according to the psychoacoustic model,  $\beta[d]$ , are also shown. The distribution of the aliasing components are quite different from the example design without psychoacoustic optimisation (figure 5.4). The design is optimised with the weighting defined by the threshold of the aliasing/imaging components so the components should be attenuated accordingly, which also seems to be the case. In general the aliasing/imaging components are very close to the threshold of audibility.

In figure 5.12 the error measures for the design are shown. The errors for the analysis filter  $\epsilon_l$ ,  $\epsilon_a$  and  $\epsilon_h$  are the same as for the design without psychoacoustic weighting. The errors for the synthesis filter design are the psychoacoustically weighted errors. The total error,  $\epsilon_{wt}$ , is lower than the total error for the design without psychoacoustic weighting. This indicates that the psychoacoustically weighted synthesis filter is easier to design. The reason is that it is easier to attenuate frequencies far from the passband than frequencies close to the passband.

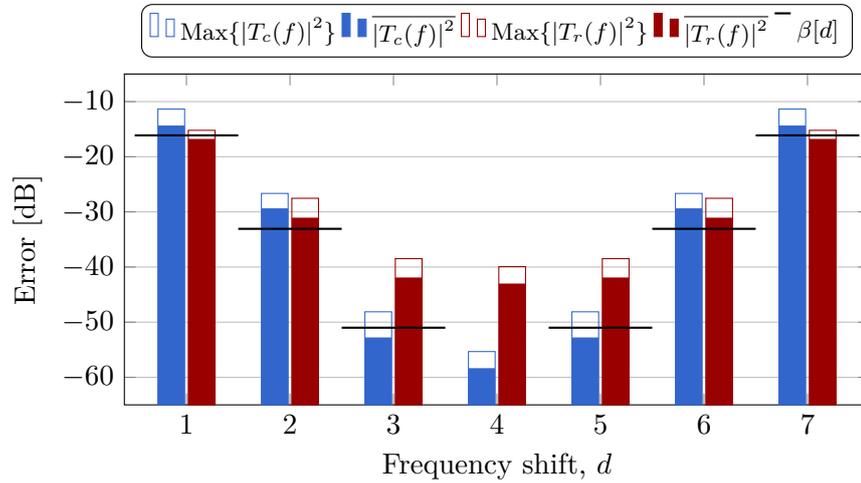


Figure 5.11: Maximum and integrated power of the aliasing/imaging components for each  $d$  for the example design with psychoacoustic weighting. The aliasing/imaging components are optimised compared the threshold of audibility. Therefore, the components with large frequency shifts are attenuated most while the components with small frequency shifts are attenuated less.

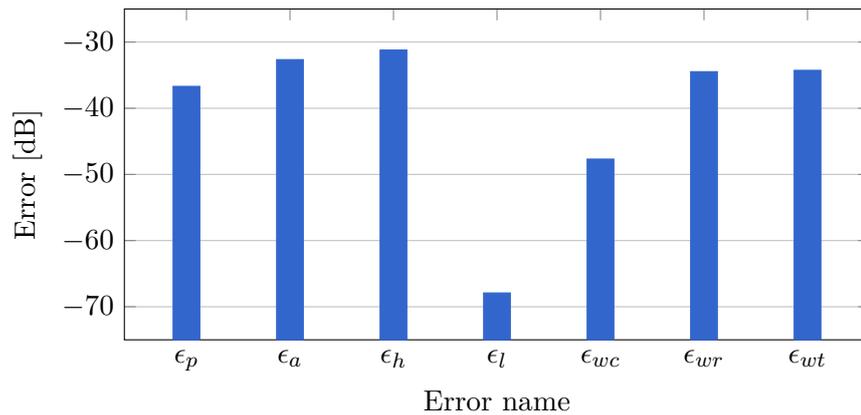


Figure 5.12: Error measures for filter bank design with the parameters in table 5.1 and 5.3. The analysis errors are the same as without psychoacoustic weighting. The total error,  $\epsilon_{wt}$ , is lower than without psychoacoustic weighting although the power of the aliasing/imaging components are higher.

## 54 Example of Filter Designs with & without Psychoacoustic Weighting

# Evaluation of Prototype Filter Designs

---

In this chapter different prototype filter designs are evaluated. First, the methods used for the evaluation is described. Afterwards, the results from the evaluation are presented.

## 6.1 Methods for Evaluation of Prototype Filter Designs

The evaluation of filter optimisation methods is done by the error measures defined in section 2.5 and an objective evaluation of speech quality for a spectral subtraction algorithm applied in the filter bank. The spectral subtraction algorithm is used to evaluate the filter design when processing is performed.

### 6.1.1 Ideal Spectral Subtraction Algorithm

Spectral subtraction is a widely used single channel speech enhancement algorithm [Loi13]. The algorithm used for the evaluation is an ideal power spectral subtraction. Usually when performing spectral subtraction only the

noisy speech signal is available and the speech and noise power need to be estimated from that signal. In an ideal spectral subtraction, the speech and noise power are estimated from the speech and noise signals separately and the gain is applied on the noisy speech signal.

The noisy speech signal is defined as

$$x_k[m] = s_k[m] + n_k[m] \quad (6.1)$$

where  $s_k[m]$  is the speech signal in band  $k$  at time  $m$  and  $n_k[m]$  is the noise signal.

The noise power is subtracted from the noisy speech power to obtain the estimated speech power

$$|y_k[m]|^2 = |s_k[m]|^2 + |n_k[m]|^2 - \widehat{|n_k[m]|^2} \quad (6.2)$$

where  $y_k[m]$  is the denoised output signal and  $\widehat{|n_k[m]|^2}$  is the estimated noise power. When using ideal estimators the estimated noise power and the noise power are equal, so only the speech power is transferred to the output

$$|y_k[m]|^2 = |s_k[m]|^2 \quad (6.3)$$

The noise reduced output signal is the estimated speech power and the phase of the noisy speech signal, i.e.

$$y_k[m] = \sqrt{|s_k[m]|^2} \angle x_k[m] \quad (6.4)$$

This can be written as a gain rule for the time varying band dependant gain coefficients [Loi13]

$$f_k[m] = \sqrt{1 - \frac{|n_k[m]|^2}{|s_k[m]|^2 + |n_k[m]|^2}} \quad (6.5)$$

where  $f_k[m]$  is the gain applied in band  $k$  at time  $m$ .

### 6.1.2 Objective Evaluation of Speech Quality

In [HL08] various objective speech quality evaluation methods have been tested on a set of speech enhancement algorithms. The results were compared to the overall quality Mean Opinion Score (MOS) [ITU06a] obtained from subjective evaluation in compliance with [ITU06b] of the same speech enhancement algorithms with the same audio samples [HL07]. The MOS

from the objective and subjective evaluation were compared by Pearson's correlation coefficient. Although Perceptual Evaluation of Speech Quality (PESQ) [ITU01] was not originally designed to evaluate speech enhancement algorithms, the correlation coefficient was  $\rho = 0.89$  [HL08]. With the exception of composite measures, PESQ was the method with highest correlation for the evaluation of the speech enhancement algorithms tested in [HL07, HL08]. Therefore, PESQ is used for the objective speech quality evaluation.

The PESQ algorithm take the original speech signal and the degraded speech signal, i.e. the noise reduced signal, as input. The algorithm consist of some preprocessing, time-alignment and an auditory transform.

In general the preprocessing and time-alignment accounts for the overall gain variations and time variation between the clean signal and the degraded signal. After the preprocessing and time-alignment both signals are transformed in frames by Zwicker's loudness model on the bark scale [ZF99].

The loudness difference between the original and degraded signal is then used in combination with a simple masking model to compute a disturbance distribution. The disturbance distribution is multiplied by an asymmetric factor to penalise negative and positive loudness differences differently.

The disturbance distribution and the asymmetric disturbance distribution is averaged over bark and frames, while bad frames are recalculated to ensure the time-alignment is correct. Both the averaged disturbance distribution and asymmetric disturbance distribution is used to calculated the PESQ score.

In this evaluation the PESQ score is translated to Mean Opinion Score Listening Quality Objective Narrowband (MOS-LQON) in compliance with [ITU03, ITU06a]. MOS is between 1 and 5 and describe the overall speech quality (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent).

### 6.1.3 Speech & Noise Samples for the Evaluation

The NOIZEUS database<sup>1</sup> [HL07] was used for the evaluation. NOIZEUS is a speech and noise corpus made for the evaluation of speech enhancement algorithms.

---

<sup>1</sup><http://ecs.utdallas.edu/loizou/speech/noizeus/>

In short NOIZEUS consist of 30 different phonetically balanced sentences obtained from 3 male and 3 female speakers. The speech samples have a duration of 3 s and is sampled at 8 kHz.

There are 8 different noise types all with the same duration and sample rate as the speech samples. These noises are originally from the AURORA database [HP00]. The noise types are: babble, car, exhibition hall, restaurant, street, airport, train station and train.

The noise and speech signals are added to obtain SNR of 0 dB, 5 dB, 10 dB, 15 dB. SNR values are calculated in compliance with the active speech level defined in [ITU11].

In the evaluation the score for all speakers and noise types are averaged to obtain one score for each SNR.

In figure 6.1 an objective evaluation of noisy speech samples is shown. This is the MOS-LQON obtained with PESQ when no speech enhancement is applied.

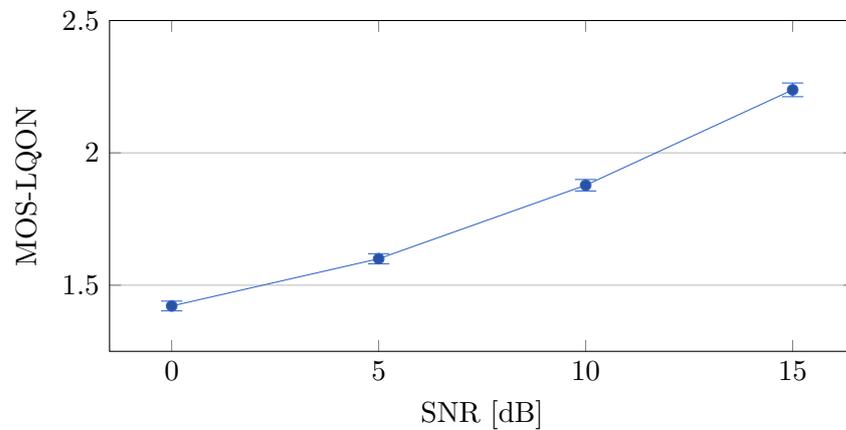


Figure 6.1: MOS-LQON obtained by PESQ with sound samples from NOIZEUS database when no speech enhancement is applied. Four different SNR values are used and the score is averaged across speakers and noise types. The error bars represent 1.96 times the standard error of the mean to each side.

## 6.2 Evaluation of Prototype Filter Designs

In this section the proposed method for filter optimisation without psychoacoustic weighting is evaluated against some standard filter optimisation

methods. Finally, the proposed method with and without psychoacoustic optimisation are compared.

### 6.2.1 Evaluation of Optimisation Method for Designing Perfect Reconstruction & Near Perfect Reconstruction Filter Banks

In table 6.1 and 6.2 the parameters for a filter bank with two different filter optimisations are shown. Both filter sets are optimised by the proposed method, but with focus on either PR or NPR. This is obtained by either setting  $\alpha_r$  or  $\alpha_c$  to  $-\infty$ .

Filter Bank Parameter	$K$	$D$	$L_h$	$L_g$	$\tau_h$	$\tau_t$	$f_s$
Both Designs	128	$\frac{K}{2}$	$2K$	$2K$	$\frac{L_h-1}{2}$	$\frac{L_g-1}{2} + \tau_h$	8 kHz

Table 6.1: Parameters for the filter bank. The optimisation parameters for this filter bank design is shown in table 6.2. The filters obtained for the filter bank is evaluated in figure 6.2 and 6.3.

Optimisation Parameters	$\alpha_a$	$\alpha_r$	$\alpha_c$
Proposed Method PR	-0.5	$-\infty$	0
Proposed Method NPR	-0.5	0	$-\infty$

Table 6.2: Parameters for the optimisation method. The filter bank parameters for this filter bank design is shown in table 6.1. The filters obtained for the filter bank is evaluated in figure 6.2 and 6.3. The two filter sets for the filter bank is designed by the proposed optimisation method, one set with focus on PR and the other filter set with focus on NPR. The optimisation with focus on PR has  $\alpha_r = -\infty$  while the optimisation with focus on NPR has  $\alpha_c = -\infty$ .

The error measures are shown in figure 6.2. Both  $\epsilon_l$  and  $\epsilon_c$  are  $-\infty$  dB for the PR filter bank which is also the requirement for PR. The aliasing/imaging error,  $\epsilon_r$ , is larger for the PR filter bank than the NPR filter bank.

In figure 6.3, MOS-LQON obtained by PESQ are shown. For an SNR of 0 db the NPR design is best while for SNR of 15 dB the PR design is best. This is in line with the expectation that a small  $\epsilon_r$ , i.e. NPR, is desired when extensive processing is performed in the filter bank while PR is desired when only minor processing is performed.

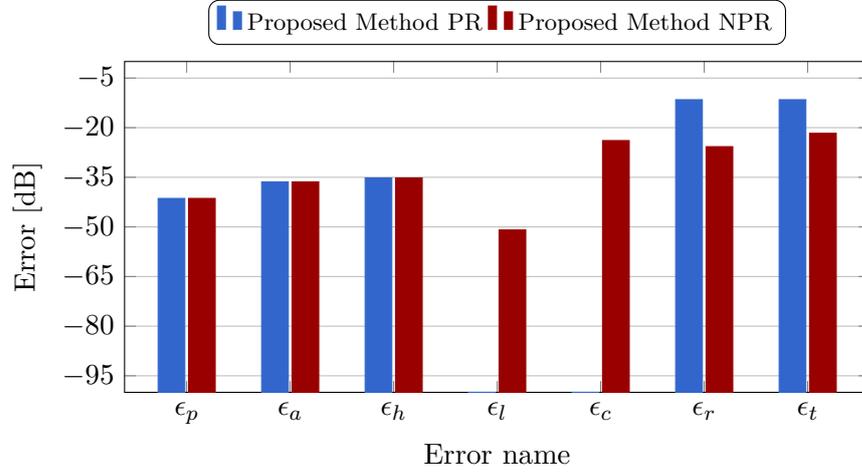


Figure 6.2: Error measures for the two filter bank designs defined in table 6.1 and 6.2. The filters obtained are evaluated by PESQ in figure 6.3. The two filter sets for the filter bank are designed by the proposed optimisation method, one set with focus on PR and the other filter set with focus on NPR. For the filter bank optimised for PR both  $\epsilon_l$  and  $\epsilon_c$  are  $-\infty$  dB while having a larger  $\epsilon_r$  than the NPR filter bank.

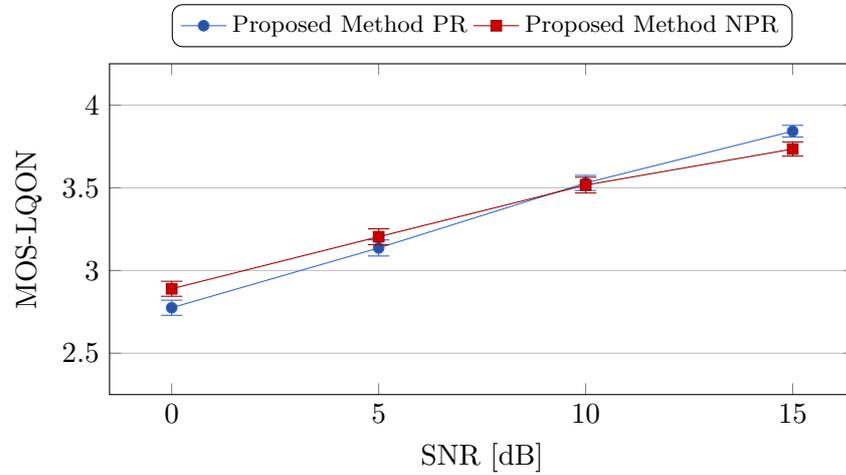


Figure 6.3: MOS-LQON obtained by PESQ with sound samples from NOIZEUS database when ideal spectral subtraction is applied in the filter bank. The designs are defined in table 6.1 and 6.2. The error measures for the filter banks are shown in figure 6.2. The two filter sets are designed by the proposed optimisation method, one set with focus on PR and the other with focus on NPR. The filter bank optimised for PR has the best MOS-LQON for high SNR while the filter bank optimised for NPR has the best MOS-LQON for low SNR. The error bars represent 1.96 times the standard error of the mean to each side.

### 6.2.2 Evaluation of the Proposed Method Against the WOLA Method

In this section the proposed method is evaluated against the WOLA method [Loi13, CR83, Smi11]. The windows used for the WOLA design is a Hann window as analysis filter and a rectangular window as the synthesis filter. The setup is shown in table 6.3. The proposed method is designed with the optimisation parameters in table 6.4. The  $\alpha_a$  value is obtained by sweeping the parameter and choosing the value where the total error,  $\epsilon_t$ , is lowest.

Filter Bank Parameter	$K$	$D$	$L_h$	$L_g$	$\tau_h$	$\tau_t$	$f_s$
Both Designs	128	$\frac{K}{2}$	$K$	$K$	$\frac{L_h-1}{2}$	$\frac{L_g-1}{2} + \tau_h$	8 kHz

Table 6.3: Parameters for the filter bank. The optimisation parameters for this filter bank design is shown in table 6.4. The filters obtained for the filter bank is evaluated in figure 6.4 and 6.5. The two filter sets for the filter bank is designed by either proposed optimisation method or a classic WOLA method.

Optimisation Parameters	$\alpha_a$	$\alpha_r$	$\alpha_c$
WOLA Method	-	-	-
Proposed Method	1	0	0

Table 6.4: Parameters for the optimisation method. The filter bank parameters for this filter bank design is shown in table 6.3. The filters obtained for the filter bank is evaluated in figure 6.4 and 6.5. The two filter sets for the filter bank is designed either by the proposed optimisation method or the classic WOLA method.

The error measures for both designs are shown in figure 6.4. The WOLA method designs a PR filter bank by compromising the aliasing/imaging error compared to the proposed method.

In figure 6.5 the MOS-LQON for the two designs are shown. The proposed method is slightly better than the WOLA design.

### 6.2.3 Evaluation of the Proposed Method Against the Window Method

In this section the proposed method is evaluated against the window method proposed in [CRALMBL02]. The proposed method is constrained to fit the filter bank parameters where the window method has a good performance, i.e. long filters compared to the number of bands [CRALMBL02, LV98]. The setup is shown in table 6.5. The proposed method is designed by the

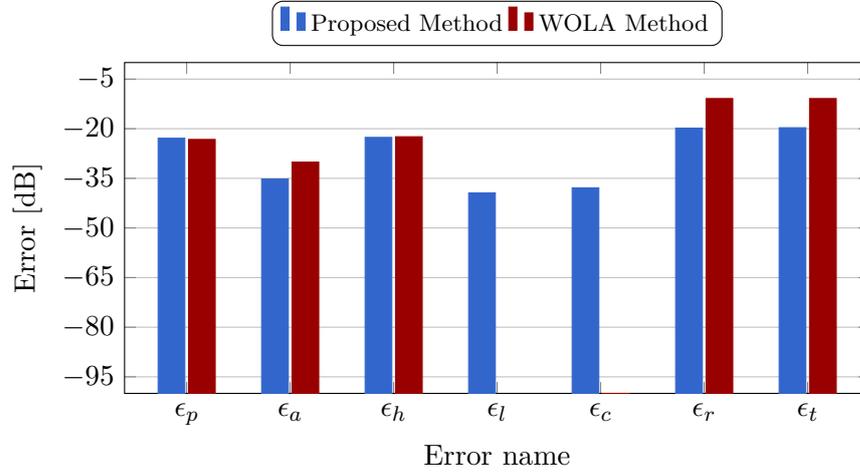


Figure 6.4: Error measures for the two filter bank designs defined in table 6.3 and 6.4. The filter banks are evaluated by PESQ in figure 6.5. The two filter sets are designed by the proposed optimisation method and the WOLA method [Loi13, CR83, Smi11]. The most significant difference between the two designs is the PR property of the WOLA design which is achieved by compromising  $\epsilon_r$ .

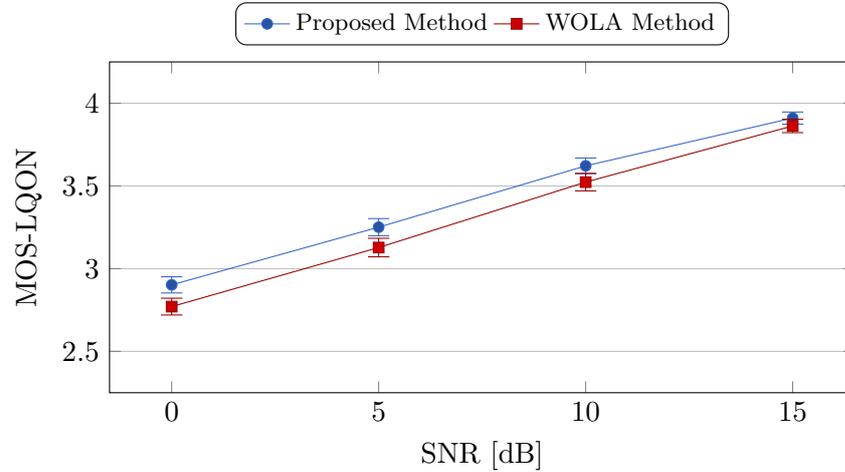


Figure 6.5: MOS-LQON obtained by PESQ with sound samples from NOIZEUS database when ideal spectral subtraction is applied in the filter bank. The designs are defined in table 6.3 and 6.4. The error measures for the filter banks are shown in figure 6.4. The two filter sets are designed by the proposed optimisation method and the WOLA method [Loi13, CR83, Smi11]. The filter bank optimised by the proposed method scores slightly better than the filter bank obtained by the WOLA method. The difference is most pronounced at 0 dB SNR which matches the findings in section 6.2.1 that NPR tends to be better when extensive manipulation is performed. The error bars represent 1.96 times the standard error of the mean to each side.

optimisation parameters in table 6.6. The  $\alpha_a$  value is obtained by sweeping the parameter and choosing the value where the total error,  $\epsilon_t$ , is lowest.

Filter Bank Parameter	$K$	$D$	$L_h$	$L_g$	$\tau_h$	$\tau_t$	$f_s$
Both Designs	64	$\frac{K}{2}$	$4K$	$4K$	$\frac{L_h-1}{2}$	$\frac{L_g-1}{2} + \tau_h$	8 kHz

Table 6.5: Parameters for the filter bank. The optimisation parameters for this filter bank design is shown in table 6.6. The filters obtained for the filter bank is evaluated in figure 6.6 and 6.7. The two filter sets for the filter bank is designed by either proposed optimisation method or the window method [CRALMBL02].

Optimisation Parameters	$\alpha_a$	$\alpha_r$	$\alpha_c$	$\beta$
Window Method	–	–	–	10.06126
Proposed Method	4.2	0	0	–

Table 6.6: Parameters for the optimisation method. The filter bank parameters for this filter bank design is shown in table 6.5. The filters obtained for the filter bank is evaluated in plot 6.6 and 6.7. The two filter sets for the filter bank is designed either by the proposed optimisation method or the window method [CRALMBL02]. The window method used requires a  $\beta$  parameter which defines the  $\beta$  for the Kasier window used. The value used is the one proposed in the example design in [CRALMBL02].

The errors for both designs are shown in figure 6.6. The proposed method scores better than the window method designs across all errors. Both designs are NPR as neither  $\epsilon_l$  nor  $\epsilon_c$  is zero in any of the designs.

In figure 6.7 the MOS-LQON for the two designs are shown. The proposed method scores slightly better than the window method.

#### 6.2.4 Evaluation of the Proposed Method with Psychoacoustic Weighting

In this section the influence of the psychoacoustic weighting is investigated. Two filter banks are designed, one with psychoacoustic weighting and one without, both with symmetric filters with the same total group delay. The optimisation parameters are obtained by sweeping the inband aliasing parameter,  $\alpha_a$ , and the filter length of the analysis filter,  $L_h$ . To obtain the same total group delay and symmetric filters the synthesis filter was shortened by the same amount as the analysis filter was extended. The surface of the total error,  $\epsilon_t$ , obtained from the sweeps of  $\alpha_a$  and  $L_h$  is shown in figure 6.8 and 6.9.

The two error surfaces are quite different and contain multiple valleys. It

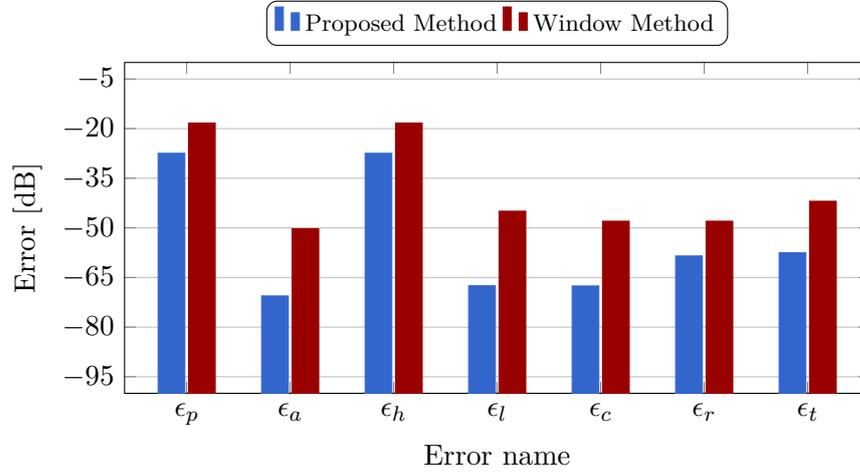


Figure 6.6: Error measures for the two filter bank designs defined in table 6.5 and 6.6. The filter bank are evaluated by PESQ in figure 6.7. The two filter sets for the filter bank are either designed by the proposed optimisation method or by the window method [CRALMBL02]. The proposed method has a lower error on all error measures compared to the design by the window method.

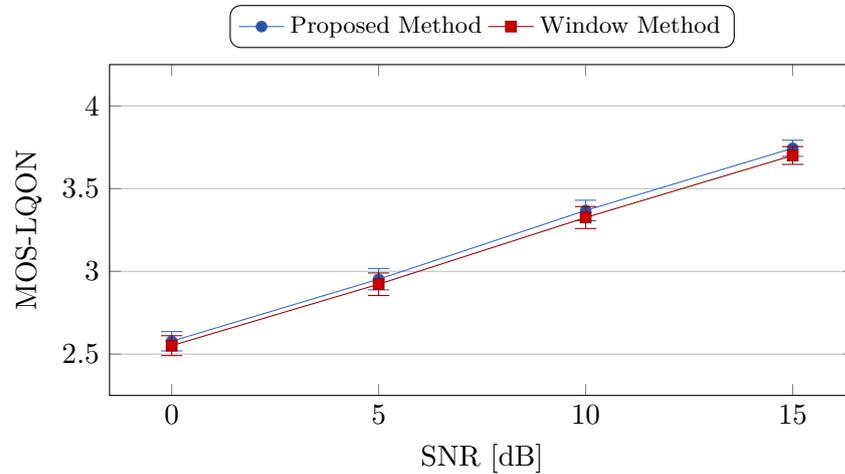


Figure 6.7: MOS-LQON obtained by PESQ with sound samples from NOIZEUS database when ideal spectral subtraction is applied in the filter bank. The designs are defined in table 6.5 and 6.6. The error measures for the filter banks are shown in figure 6.6. The two filter sets are either designed by the proposed optimisation method or by the window method [CRALMBL02]. The filter banks have almost identical scores with the proposed method being slightly better. The error bars represent 1.96 times the standard error of the mean to each side.

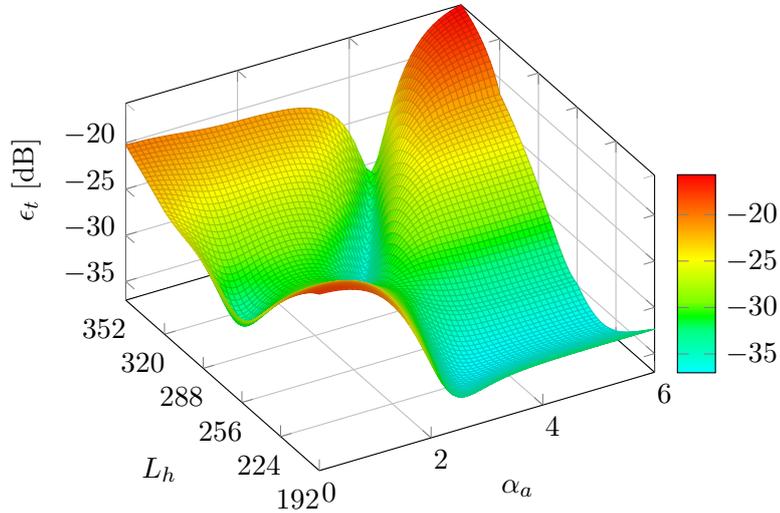


Figure 6.8: Total error,  $\epsilon_t$ , surface for the proposed method without psychoacoustic weighting. The sweep parameters are  $\alpha_a$  and  $L_h$ . The total group delay  $\tau_t$  is held constant. Both the analysis and synthesis filter is linear phase which means that an increase in  $L_h$  results in a decrease in  $L_g$ . The rest of the filter bank and optimisation parameters are shown in table 6.7 and 6.8.

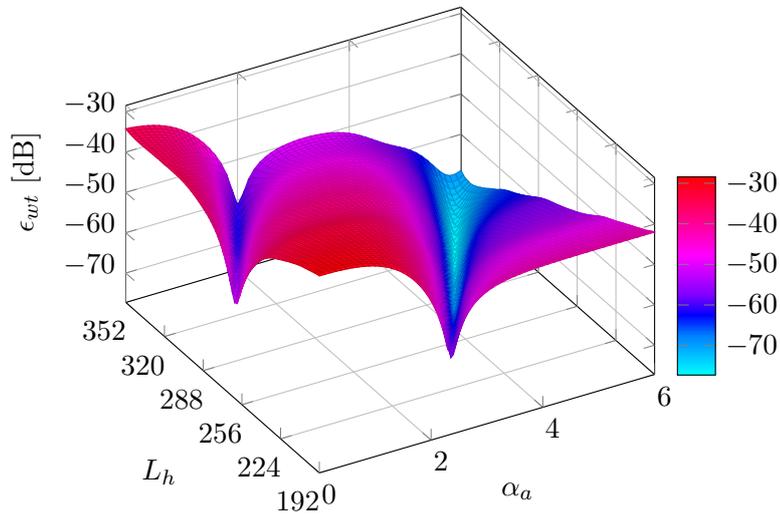


Figure 6.9: Total error,  $\epsilon_{wt}$ , surface for the proposed method with psychoacoustic weighting. The sweep parameters are  $\alpha_a$  and  $L_h$ . The total group delay  $\tau_t$  is held constant. Both the analysis and synthesis filter is linear phase which means that an increase in  $L_h$  results in a decrease in  $L_g$ . The rest of the filter bank and optimisation parameters are shown in table 6.7 and 6.8.

seems that a simple expectation maximisation algorithm would not be a good choice for finding optimisation parameters. Therefore, we have used this brute force approach and only optimised two parameters. The chosen parameters are where the total error is lowest in each surface. The parameters for the filter banks and the optimisations are shown in table 6.7 and 6.8.

Filter Bank Parameter	$K$	$D$	$L_h$	$L_g$	$\tau_h$	$\tau_t$	$f_s$
Proposed Method	128	$\frac{K}{2}$	$2K - 24$	$2K + 24$	$\frac{L_h - 1}{2}$	$\frac{L_g - 1}{2} + \tau_h$	8 kHz
Proposed Method (Psychoacoustics)	128	$\frac{K}{2}$	$2K + 62$	$2K - 62$	$\frac{L_h - 1}{2}$	$\frac{L_g - 1}{2} + \tau_h$	8 kHz

Table 6.7: Parameters for the filter bank. The optimisation parameters for this filter bank design is shown in table 6.8. The filters obtained for the filter bank is evaluated in plot 6.10 and 6.12. The two filter sets for the filter bank is designed by the proposed optimisation method with and without the psychoacoustic optimisation.

Optimisation Parameters	$\alpha_a$	$\alpha_r$	$\alpha_c$	$f_c$	$\kappa$	$\alpha_{wr}$	$\alpha_{wc}$
Proposed Method	3.7	0	0	–	–	–	–
Proposed Method (Psychoacoustics)	5.16	–	–	500 Hz	0.4	0	0

Table 6.8: Parameters for the optimisation method. The filter bank parameters for this filter bank design is shown in table 6.7. The filters obtained for the filter bank is evaluated in figure 6.10 and 6.12. The two filter sets for the filter bank is designed by the proposed optimisation method with and without the psychoacoustic optimisation.

The error measures for the two filter optimisations are shown in figure 6.10 and 6.11. As expected the non weighted filter bank scores best on the non weighted errors, while the psychoacoustically weighted filter bank is best on the psychoacoustically weighted errors.

In figure 6.12 the MOS-LQON scores obtained by PESQ are shown. The results show that the filter bank without psychoacoustic weighting in the optimisation is far superior. The score is very high also compared to the other designs obtained in the evaluation against other methods.

To verify that the difference in the MOS-LQON score is representing an audible difference an informal subjective listening test were conducted. The result were in line with the PESQ result and it were decided that no further psychoacoustic listening experiment were required to conclude that the psychoacoustically weighted filter bank performed worse than the filter bank without psychoacoustic weighting.

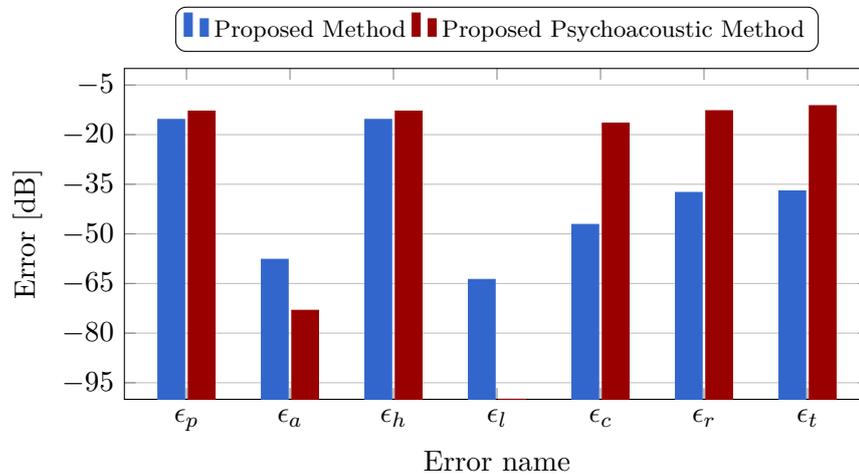


Figure 6.10: Error measures for the two filter bank designs defined in table 6.7 and 6.8. The filter banks are evaluated by PESQ in figure 6.12. The two filter sets for the filter bank are designed by the proposed optimisation method with and without psychoacoustic weighting. The filter bank optimised with the psychoacoustic weighting has the largest error. This makes sense as this filter bank is optimised to reduce the weighted errors which are shown in figure 6.11.

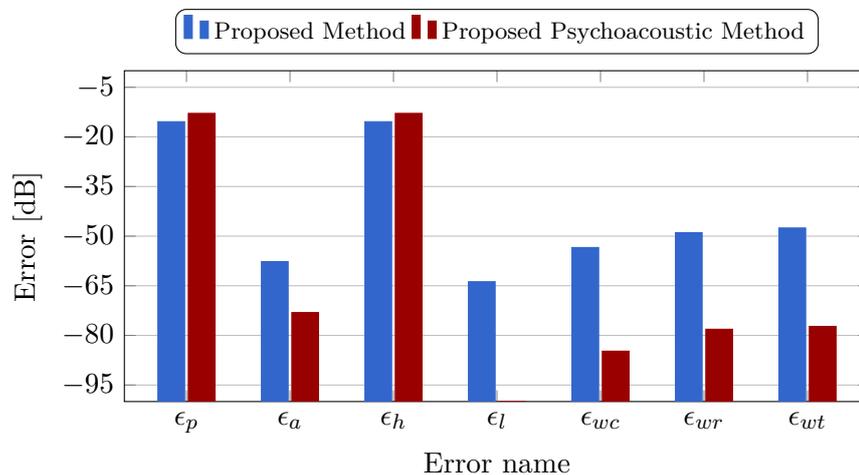


Figure 6.11: Psychoacoustically weighed error measures for the two filter bank designs defined in table 6.7 and 6.8. The filter banks are evaluated by PESQ in figure 6.12. The two filter sets for the filter banks are designed by the proposed optimisation method with and without psychoacoustic weighting. The filter bank optimised with the psychoacoustic weighting has the smallest error. This makes sense as the filter bank is optimised to reduce the weighted errors and not the nonweighted errors shown in figure 6.10.

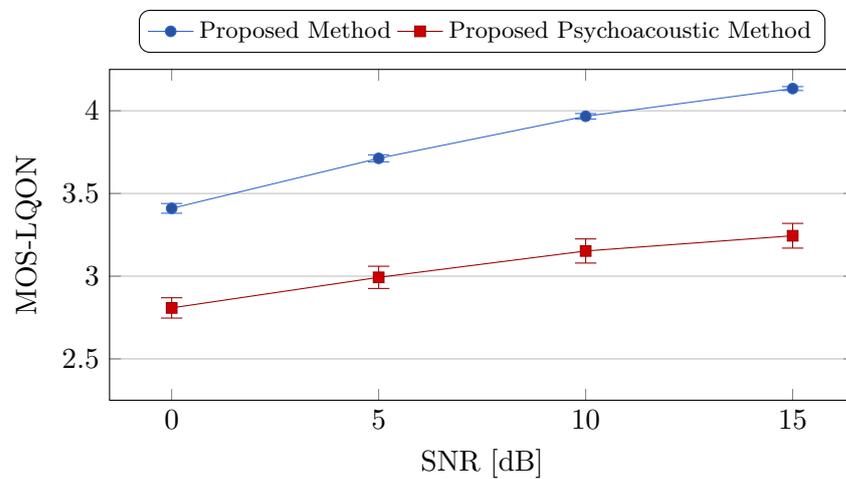


Figure 6.12: MOS-LQON obtained by PESQ with sound samples from NOIZEUS database when ideal spectral subtraction is applied in the filter bank. The designs are defined in table 6.7 and 6.8. The error measures for the filter banks are shown in figure 6.10 and 6.11. The two filter sets are designed by the proposed optimisation method with and without psychoacoustic weighting. The filter bank optimised without psychoacoustic weighting is superior to the design with psychoacoustic weighting. The error bars represent 1.96 times the standard error of the mean to each side.

# Analysis & Discussion of Artefacts Introduced by the Psychoacoustic Weighting

---

In the previous chapter we saw that the filter bank performed worse with the psychoacoustic weighting than without. In this chapter, an analysis of the output signals are performed to see if the optimisation did as expected and the assumptions made in the psychoacoustic model are discussed. Afterwards, artefacts in the modulation domain are discussed.

## 7.1 Signal Analysis & Discussion of Assumptions in the Psychoacoustic Model

To investigate why the psychoacoustically weighted filter bank performs worse than the one without psychoacoustic weighting a brief analysis of the noise reduced signals is presented. The low frequency part of Power Spectral Density (PSD) estimates for a female speaker (speaker 14 from NOIZEUS database) degraded by the car noise with an SNR of 15 dB are shown in figure 7.1. The female speaker was chosen as the artefacts are most pronounced on the female speakers compared to the male speakers.

From the PSD it can be seen that the signals processed with the psychoa-

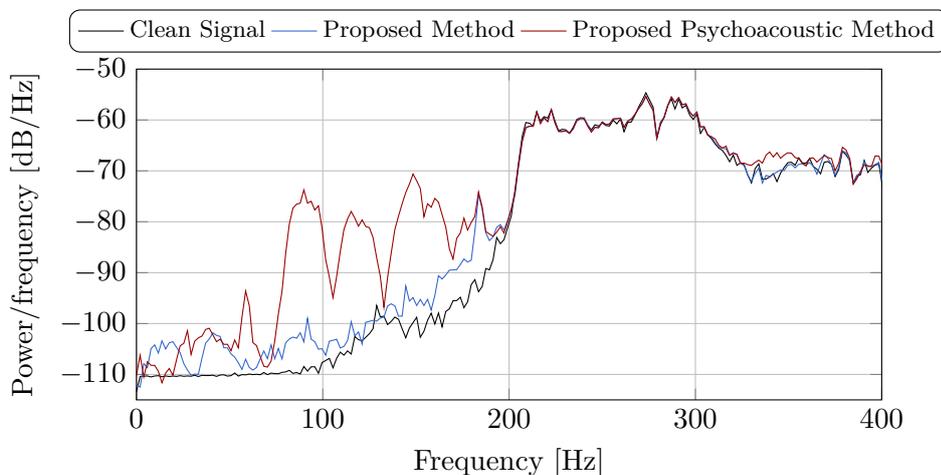


Figure 7.1: PSD of noise reduced signals obtained by an ideal spectral subtraction algorithm implemented in filter bank optimised by proposed method with and without psychoacoustic weighting. The signal is speaker 14 (female) from the NOIZEUS database degraded by a car noise at 15 dB SNR. The clean speech PSD without noise is plotted to estimate the target curve for the noise reduced signals. The pPSD are obtained by the Welch method with a 4096 tap Hamming window, 50% overlap and a DFT with the same number of bins as the length of the window.

oustically weighted filter bank has a larger power than the clean speech signal at frequencies below 200 Hz. The extra power is an artefact as it is not present in the clean speech. Compared to the peak power of the clean speech between 200 Hz and 300 Hz the power of the artefacts are only attenuated approximately 15 dB.

The aliasing/imaging components with least attenuation is at  $d = 1$  (and  $d = D - 1$ ). To verify that the observed power is an aliasing/imaging artefact the aliasing/imaging component for  $d = 1$ , i.e.  $Y_c(f)|_{d=1}$ , is shown in figure 7.2. The aliasing/imaging component for  $d = 1$  is shifted by  $-125$  Hz according to (4.10). The component fits very well with the low frequency artefacts. This indicates that the artefacts are generated by the aliasing/imaging component with a frequency shift of  $-125$  Hz.

According to the psychoacoustic model, the artefacts should be less audible than the artefacts generated by the filter bank without psychoacoustic optimisation. This is not the case because of the assumptions made in the psychoacoustic model.

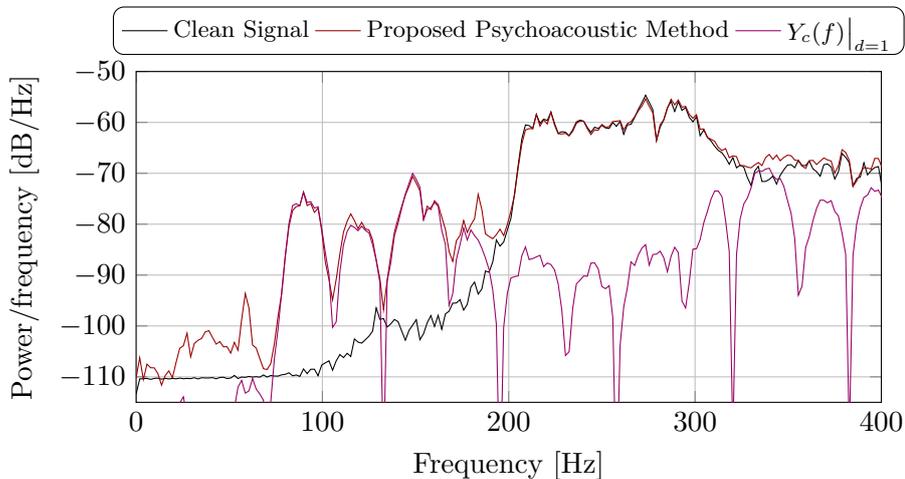


Figure 7.2: PSD estimate of the noise reduced signal by an ideal spectral subtraction and the aliasing/imaging component for  $d = 1$  for the filter bank optimised with psychoacoustic weighting. The signal is speaker 14 (Female) from the NOIZEUS database degraded by a car noise at a SNR = 15 dB. The clean speech PSD without noise is plotted to estimate the target curve for the noise reduced signals. The pPSD are obtained by the Welch method with a 4096 tap Hamming window, 50% overlap and a DFT with the same number of bins as the length of the window.

### 7.1.1 Assumptions in the Psychoacoustic Model

The psychoacoustic model is based on a single auditory filter width, i.e.  $f_c = 500$  Hz. This makes the masking curves too wide below 500 Hz and too narrow above 500 Hz. If the model used a filter with  $f_c = 90$  Hz (the lowest frequency peak in figure 7.2), the gain of aliasing/imaging power required for it to be inaudible would be  $\kappa\beta[1]|_{f_c=90\text{ Hz}} \approx -51$  dB. Using an  $f_c$  of only 90 Hz would result in very narrow masking curves for all frequencies, so the psychoacoustically weighted filter bank would approach a filter bank without psychoacoustic weighting. Another and better way to solve the issue is to change the model to use auditory filters with different bandwidth for different frequencies. This is not possible in a DFT modulated filter bank, but could probably be incorporated in a warped DFT modulated filter bank.

Another issue with the psychoacoustic model is that off-frequency listening is not accounted for. The model assumes that the auditory filter is centered at the aliasing/imaging component. By using off-frequency listening the aliasing/imaging components could still be audible because the signal is not positioned symmetrically around the aliasing/imaging component. Furthermore, the model only look at the audibility of one aliasing component at a time assuming that the only other signal present is the original. According

to the power spectrum model, the sum of the aliasing components in an auditory filter should be compared to the original signal.

Another issue is that the psychoacoustic model only look at masking of single frequency components. It is assumed that the aliasing/imaging components are only masked by the signal that generate them. All frequencies in the original signal that is passed through the linear response will contribute to the masking of the aliasing/imaging. For a broadband input signal the aliasing/imaging components that are far away in frequency from the generating frequency component are masked by other frequencies in the original signal. Assuming the input signal has a uniform distribution of power per frequency, the optimal way to reduce the audibility of aliasing/imaging according to the power spectrum model is to reduce the overall power of the aliasing/imaging. This is what the optimisation method without psychoacoustic weighting do.

## 7.2 Aliasing/Imaging Artefacts in the Modulation Domain

As touched upon in section 4.1.2, the audibility of the aliasing/imaging artefacts could also be assessed in the modulation domain. The modulation spectrum<sup>1</sup> of the same signal as used in the previous section is shown in figure 7.3.

The signal processed with the psychoacoustically weighted filter bank has a peak in the modulation spectrum at 125 Hz that is not observed in the clean speech signal. To explain the extra modulation at 125 Hz the aliasing/imaging transfer has to be reinterpreted. The aliasing/imaging transfer is defined in equation (2.15) as

$$Y_c(z) = T_c(z)X(zW_D^d), \quad d = 1, 2, \dots, D - 1 \quad (7.1)$$

In the frequency domain we see  $W_D^d$  as a frequency shift, but in time domain it could be viewed as a modulation of the signal. By simplifying the transfer to  $T_l(z) = 1$  and  $T_c(z) = 1$  the output of the filter bank can be written in a

---

<sup>1</sup>The modulation spectrum is the PSD estimate of the Hilbert envelope. The Hilbert envelope is the absolute value of the analytic signal obtained by  $|x_a[n]| = |x[n] + j\mathcal{H}\{x[n]\}|$  where  $\mathcal{H}\{x[n]\}$  is the hilbert transform of  $x[n]$ . The PSD is estimated with the Welch method with a 4096 tap Hamming window, 50% overlap and a DFT with the same number of bins as the length of the window.

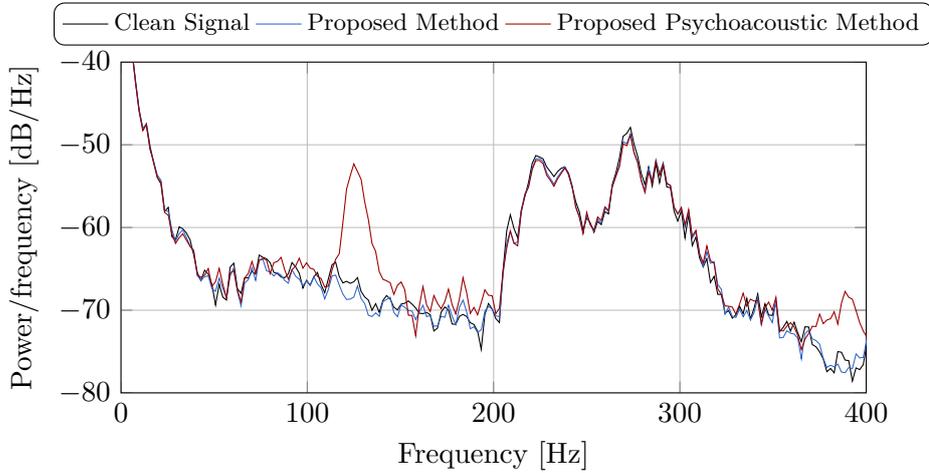


Figure 7.3: Modulation spectrum<sup>1</sup> of noise reduced signals obtained by an ideal spectral subtraction algorithm implemented in filter bank optimised by the proposed method with and without psychoacoustic weighting. The signal is speaker 14 (Female) from the NOIZEUS database degraded by a car noise at a SNR = 15 dB. The clean speech modulation spectrum without noise is plotted to estimate the target curve for the noise reduced signals.

very simple way

$$y[n] = \sum_{d=0}^{D-1} x[n]W_D^{-nd} \quad (7.2)$$

The input  $x[n]$  is modulated by  $D$  complex exponentials. Looking at the linear part and the aliasing/imaging components for  $d = 1$  and  $d = D - 1$  we get

$$\begin{aligned} y[n] &= x[n] + x[n]W_D^{-n} + x[n]W_D^{-n(D-1)} \\ &= x[n](1 + 2\cos(2\pi n/D)) \end{aligned} \quad (7.3)$$

This is cosine modulation with a modulation frequency of

$$f_{mod} = \frac{f_s}{D} = 125 \text{ Hz} \quad (7.4)$$

where  $f_{mod}$  is the modulation frequency and  $f_s$  is the sampling frequency. This indicates that the extra modulation observed in figure 7.3 is caused by the aliasing/imaging components at  $d = 1$  and  $d = D - 1$ .

$T_l(z)$  and  $T_c(z)$  introduce different filtering of the carrier  $x[n]$  and the sidebands  $x[n]W_D^{-nd}$ , which results in frequency dependent modulation depth and phase, but they do not change the modulation frequency.

The audibility of modulation have been investigated by various experiments [Moo12]. Studies show that the detection of modulation is highly dependent

on the carrier bandwidth, i.e.  $X(z)$ , [DKK97a]. To account for the detection of modulation a model incorporating a modulation filter bank has been proposed in [DKK97a, DKK97b]. This model implies that modulation detection is performed individually for each auditory filter meaning that the sidebands need to be in the same auditory filter as the carrier. If the sidebands and carrier are in different auditory filters, frequency domain masking should be used to assess the audibility instead [DKK97a].

In order to mask the aliasing/imaging components, the psychoacoustic weighting try to move the power of the aliasing/imaging components to the same auditory filter as the original signal. This means that by applying the psychoacoustic weighting, the psychoacoustic interpretation of the aliasing/imaging components are moved from frequency domain to modulation domain.

## Summary & Conclusion

---

Different methods have been proposed for the design of the prototype filters for DFT modulated filter banks. None of these methods use knowledge from psychoacoustics to reduce the audibility of the artefacts introduced by subsampling in the filter bank. In this thesis the quadratic optimisation method proposed in [dH01] was modified to incorporate frequency domain masking and an additional error for optimising for perfect reconstruction was introduced.

An overview of the disagreement in literature for the offset in the modulation of the analysis and synthesis filters was presented and a new modulation offset was proposed. This offset enables the filter bank to be designed with an arbitrary group delay.

The artefacts introduced by filter banks were quantified by a set of error functions and different psychoacoustic concepts were proposed to assess the audibility of them. A simple frequency domain masking model was introduced to quantify the audibility of the aliasing/imaging components introduced in the filter bank. The aliasing/imaging components can be interpreted as frequency shifted and filtered versions of the original signal.

The masking model was applied to the error functions to obtain psychoacoustically weighted errors which could be used in the optimisation method. To apply the masking model to the error functions it was simplified to have constant bandwidth for all frequencies.

The optimisation method without psychoacoustically weighted errors performed better than classical designs like WOLA and window method when evaluated with the proposed error functions and the objective speech quality measure PESQ. The PESQ evaluation was conducted on noise reduced signals by an ideal spectral subtraction in the filter bank. The design with psychoacoustically weighted errors showed poor performance in the PESQ evaluation and in informal subjective listening tests.

The artefacts in the noise reduced speech signal processed by the design with psychoacoustic weighted errors were investigated and it was found that the artefacts were most pronounced at frequencies below the fundamental of the speech signal. This can be explained by three assumptions in the masking model. The simplification to constant bandwidth for all frequencies, the assumption of no off-frequency listening and the assumption that only the generating frequency of the signal masks the aliasing/imaging components.

Finally, the aliasing/imaging artefacts were reinterpreted as time domain modulations instead of frequency shifts. The modulation spectrum of the noise reduced signals were analysed and modulation artefacts were present. According to the psychoacoustic model in [DKK97a, DKK97b] the artefacts should be interpreted as time domain modulation when the carrier and sidebands are located in the same auditory filter.

## 8.1 Further Work

The limitations imposed by the DFT modulated filter bank, such as uniformly spaced equal bandwidth filters, limits the usefulness of psychoacoustic weighting. The psychoacoustic model could benefit from a filter bank structure with centre frequency dependent filters. One efficient structure could be the warped DFT modulated filter bank.

A psychoacoustic model with a better correlation between the error functions and the perception of the artefacts is desired. To obtain such a model the artefacts introduced by a filter bank should be investigated through psychoacoustic experiments. Such an investigation would have to combine many aspects of psychoacoustics.

## APPENDIX A

# Derivation of Matrices for Prototype Filter Design

---

In this appendix the derivation of the error functions are available. This appendix is tightly coupled with chapter 3 where an overview of the minimisation of the error functions is presented.

## A.1 Passband Error

The passband error,  $\epsilon_p$ , is given by

$$\epsilon_p = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} |H_0(e^{j\omega}) - H_d(e^{j\omega})|^2 d\omega \quad (\text{A.1})$$

where  $H_d(e^{j\omega})$  is the desired total response. The desired response is assumed to be  $H_d(e^{j\omega}) = e^{-j\omega\tau_h}$ , i.e. a magnitude response of 1 and a group delay of  $\tau_h$  in the passband.

### A.1.1 Rewrite Passband Error to Matrix Form

To solve the optimisation problem the following structure is desired

$$\epsilon_p = \mathbf{h}^T \mathbf{A} \mathbf{h} - 2\mathbf{h}^T \mathbf{b} + 1 \quad (\text{A.2})$$

By expanding the squared error function the problem can be separated into smaller parts which fits the desired structure

$$\begin{aligned}
\epsilon_p &= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} H_0(e^{j\omega})H_0^*(e^{j\omega}) - 2\Re\{H_d^*(e^{j\omega})H_0(e^{j\omega})\} + 1 \, d\omega \\
&= \underbrace{\frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} H_0(e^{j\omega})H_0^*(e^{j\omega}) \, d\omega}_{\text{Part 1}} - 2 \underbrace{\frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{H_d^*(e^{j\omega})H_0(e^{j\omega})\} \, d\omega}_{\text{Part 2}} + 1
\end{aligned} \tag{A.3}$$

To obtain the desired form  $\mathbf{h}$  has to be isolated.

Substituting the definition of  $H_0(e^{j\omega})$  (3.1) into Part 1

$$\begin{aligned}
\text{Part 1} &= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} H_0(e^{j\omega})H_0^*(e^{j\omega}) \, d\omega \\
&= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \mathbf{h}^T \boldsymbol{\phi}_h(e^{j\omega}) \mathbf{h}^T \boldsymbol{\phi}_h^*(e^{j\omega}) \, d\omega
\end{aligned} \tag{A.4}$$

To isolate  $\mathbf{h}$  the inner product,  $\mathbf{h}^T \boldsymbol{\phi}_h^*(e^{j\omega})$ , can be transposed

$$\begin{aligned}
\text{Part 1} &= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \mathbf{h}^T \boldsymbol{\phi}_h(e^{j\omega}) \boldsymbol{\phi}_h^H(e^{j\omega}) \mathbf{h} \, d\omega \\
&= \mathbf{h}^T \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \boldsymbol{\phi}_h(e^{j\omega}) \boldsymbol{\phi}_h^H(e^{j\omega}) \, d\omega \mathbf{h} \\
&= \mathbf{h}^T \mathbf{A} \mathbf{h}
\end{aligned} \tag{A.5}$$

where  $\mathbf{A}$  is an  $L_h \times L_h$  matrix

$$\mathbf{A} = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \boldsymbol{\phi}_h(e^{j\omega}) \boldsymbol{\phi}_h^H(e^{j\omega}) \, d\omega \tag{A.6}$$

Rearrange Part 2 by substituting the definition of  $H_0(e^{j\omega})$  (3.1)

$$\begin{aligned}
\text{Part 2} &= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{H_d^*(e^{j\omega})H_0(e^{j\omega})\} \, d\omega \\
&= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{e^{j\omega\tau_h} \mathbf{h}^T \boldsymbol{\phi}_h(e^{j\omega})\} \, d\omega \\
&= \mathbf{h}^T \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{e^{j\omega\tau_h} \boldsymbol{\phi}_h(e^{j\omega})\} \, d\omega \\
&= \mathbf{h}^T \mathbf{b}
\end{aligned} \tag{A.7}$$

where  $\mathbf{b}$  is an  $L_h \times 1$  vector

$$\mathbf{b} = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{e^{j\omega\tau_h} \boldsymbol{\phi}_h(e^{j\omega})\} \, d\omega \tag{A.8}$$

### A.1.2 Simplifying the Passband Error

$\mathbf{A}$  is an  $L_h \times L_h$  matrix defined as

$$\mathbf{A} = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \phi_h(e^{j\omega}) \phi_h^H(e^{j\omega}) d\omega \quad (\text{A.9})$$

Each element in  $\mathbf{A}$  can be calculated by

$$\begin{aligned} A_{p,q} &= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} e^{-j\omega p} e^{j\omega q} d\omega \\ &= \text{sinc}\left(\frac{\omega_p}{\pi}(q-p)\right) \end{aligned} \quad (\text{A.10})$$

where  $A_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{A}$  and

$$\text{sinc}(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x}, & x \neq 0 \\ 1, & x = 0 \end{cases} \quad (\text{A.11})$$

$\mathbf{b}$  is an  $L_h \times 1$  vector defined as

$$\mathbf{b} = \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{e^{j\omega\tau_h} \phi_h(e^{j\omega})\} d\omega \quad (\text{A.12})$$

Each element in  $\mathbf{b}$  can be calculated by

$$\begin{aligned} b_p &= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \Re\{e^{j\omega\tau_h} e^{-j\omega p}\} d\omega \\ &= \frac{1}{2\omega_p} \int_{-\omega_p}^{\omega_p} \cos(\omega(\tau_h - p)) d\omega \\ &= \text{sinc}\left(\frac{\omega_p}{\pi}(\tau_h - p)\right) \end{aligned} \quad (\text{A.13})$$

where  $b_p$  is the  $p$ -th row in  $\mathbf{b}$ .

## A.2 Inband Aliasing Error

The inband aliasing error,  $\epsilon_a$ , is given by

$$\epsilon_a = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D-1} \sum_{d=1}^{D-1} \left| H_0(e^{j\omega/D} W_D^d) \right|^2 d\omega \quad (\text{A.14})$$

### A.2.1 Rewrite Inband Aliasing Error to Matrix Form

To solve the optimisation problem the following structure is desired

$$\epsilon_a = \mathbf{h}^T \mathbf{C} \mathbf{h} \quad (\text{A.15})$$

This structure can be achieved by expanding the error function as in section A.1.1

$$\begin{aligned} \epsilon_a &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D-1} \sum_{d=1}^{D-1} H_0(e^{j\omega/D} W_D^d) H_0^*(e^{j\omega/D} W_D^d) d\omega \\ &= \frac{1}{2\pi(D-1)} \sum_{d=1}^{D-1} \int_{-\pi}^{\pi} \mathbf{h}^T \phi_h(e^{j\omega/D} W_D^d) \mathbf{h}^T \phi_h^*(e^{j\omega/D} W_D^d) d\omega \\ &= \mathbf{h}^T \frac{1}{2\pi(D-1)} \sum_{d=1}^{D-1} \int_{-\pi}^{\pi} \phi_h(e^{j\omega/D} W_D^d) \phi_h^H(e^{j\omega/D} W_D^d) d\omega \mathbf{h} \\ &= \mathbf{h}^T \mathbf{C} \mathbf{h} \end{aligned} \quad (\text{A.16})$$

where  $\mathbf{C}$  is an  $L_h \times L_h$  matrix

$$\mathbf{C} = \frac{1}{2\pi(D-1)} \sum_{d=1}^{D-1} \int_{-\pi}^{\pi} \phi_h(e^{j\omega/D} W_D^d) \phi_h^H(e^{j\omega/D} W_D^d) d\omega \quad (\text{A.17})$$

### A.2.2 Simplifying the Inband Aliasing Error

$\mathbf{C}$  is an  $L_h \times L_h$  matrix defined as

$$\mathbf{C} = \frac{1}{2\pi(D-1)} \sum_{d=1}^{D-1} \int_{-\pi}^{\pi} \phi_h(e^{j\omega/D} W_D^d) \phi_h^H(e^{j\omega/D} W_D^d) d\omega \quad (\text{A.18})$$

Each element in  $\mathbf{C}$  can be calculated by

$$\begin{aligned} C_{p,q} &= \frac{1}{2\pi(D-1)} \sum_{d=1}^{D-1} \int_{-\pi}^{\pi} e^{-j\omega p/D} W_D^{-pd} e^{j\omega q/D} W_D^{qd} d\omega \\ &= \frac{1}{2\pi(D-1)} \sum_{d=1}^{D-1} W_D^{(q-p)d} \int_{-\pi}^{\pi} e^{j\omega(q-p)/D} d\omega \\ &= \frac{1}{(D-1)} (D\Delta_D[q-p] - 1) \text{sinc}\left(\frac{1}{D}(q-p)\right) \end{aligned} \quad (\text{A.19})$$

where  $C_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{C}$  and

$$\Delta_D[n] = \sum_{m=-\infty}^{\infty} \delta[n - mD] \quad (\text{A.20})$$

i.e. a Kronecker comb function with period  $D$ .

### A.3 Linear Response Error

The linear response error,  $\epsilon_l$ , is given by

$$\epsilon_l = \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_l(e^{j\omega}) - T_d(e^{j\omega})|^2 d\omega \quad (\text{A.21})$$

where  $T_d(e^{j\omega}) = e^{-j\omega\tau_t}$  and  $T_l(e^{j\omega})$  is given by (2.14).

#### A.3.1 Rewrite Linear Response Error to Matrix Form

To solve the optimisation problem the following structure is desired

$$\epsilon_l = \mathbf{g}^T \mathbf{E} \mathbf{g} - 2\mathbf{g}^T \mathbf{f} + 1 \quad (\text{A.22})$$

By expanding the squared error function the problem can be separated into smaller parts which fits the desired structure

$$\begin{aligned} \epsilon_l &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D} \sum_{k=0}^{K-1} H_k(e^{j\omega}) G_k(e^{j\omega}) \frac{1}{D} \sum_{l=0}^{K-1} H_k^*(e^{j\omega}) G_k^*(e^{j\omega}) \\ &\quad - 2\Re\{e^{j\omega\tau_t} \frac{1}{D} \sum_{m=0}^{K-1} H_k(e^{j\omega}) G_k(e^{j\omega})\} + 1 d\omega \\ &= \underbrace{\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D} \sum_{k=0}^{K-1} H_k(e^{j\omega}) G_k(e^{j\omega}) \frac{1}{D} \sum_{l=0}^{K-1} H_k^*(e^{j\omega}) G_k^*(e^{j\omega}) d\omega}_{\text{Part 1}} \\ &\quad - 2 \underbrace{\frac{1}{2\pi} \int_{-\pi}^{\pi} \Re\{e^{j\omega\tau_t} \frac{1}{D} \sum_{m=0}^{K-1} H_k(e^{j\omega}) G_k(e^{j\omega})\} d\omega + 1}_{\text{Part 2}} \end{aligned} \quad (\text{A.23})$$

Rearranging Part 1 by substituting the definition of  $G_k(e^{j\omega})$  (3.2),  $H_k(e^{j\omega})$  (3.1) and isolating the filter yields

$$\begin{aligned} \text{Part 1} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ &\quad \frac{1}{D} \sum_{k=0}^{K-1} \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \mathbf{g}^T \phi_g(e^{j\omega} W_K^k) \\ &\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l) W_K^{-\tau_t l} \mathbf{g}^T \phi_g^*(e^{j\omega} W_K^l) d\omega \end{aligned}$$

$$\begin{aligned}
&= \mathbf{g}^T \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \\
&\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \mathbf{g} \\
&= \mathbf{g}^T \mathbf{E} \mathbf{g} \tag{A.24}
\end{aligned}$$

where  $\mathbf{E}$  is an  $L_g \times L_g$  matrix

$$\begin{aligned}
\mathbf{E} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \\
&\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \tag{A.25}
\end{aligned}$$

Rearrange Part 2 by substituting the definition of  $G_k(e^{j\omega})$  (3.2) and  $H_k(e^{j\omega})$  (3.1)

$$\begin{aligned}
\text{Part 2} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Re \left\{ e^{j\omega\tau_t} \frac{1}{D} \sum_{k=0}^{K-1} H_k(e^{j\omega}) G_k(e^{j\omega}) \right\} d\omega \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \Re \left\{ e^{j\omega\tau_t} \frac{1}{D} \sum_{k=0}^{K-1} \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \mathbf{g}^T \phi_g(e^{j\omega} W_K^k) \right\} d\omega \\
&= \mathbf{g}^T \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \Re \left\{ e^{j\omega\tau_t} \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \right\} d\omega \\
&= \mathbf{g}^T \mathbf{f} \tag{A.26}
\end{aligned}$$

where  $\mathbf{f}$  is an  $L_g \times 1$  vector

$$\mathbf{f} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Re \left\{ e^{j\omega\tau_t} \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \right\} d\omega \tag{A.27}$$

## A.3.2 Simplifying Linear Response Error

$\mathbf{E}$  is an  $L_g \times L_g$  matrix defined as

$$\begin{aligned} \mathbf{E} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ &\quad \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \\ &\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \end{aligned} \quad (\text{A.28})$$

Each element in  $\mathbf{E}$  can be calculated by

$$\begin{aligned} E_{p,q} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ &\quad \frac{1}{D} \sum_{k=0}^{K-1} e^{-j\omega p} W_K^{-pk} \sum_{r=0}^{L_h-1} h_0[r] e^{-j\omega r} W_K^{-rk} W_K^{\tau_t k} \\ &\quad \frac{1}{D} \sum_{l=0}^{K-1} \sum_{s=0}^{L_h-1} h_0[s] e^{j\omega s} W_K^{sl} W_K^{-\tau_t l} e^{j\omega q} W_K^{ql} d\omega \\ &= \frac{1}{2\pi D^2} \sum_{r=0}^{L_h-1} \sum_{s=0}^{L_h-1} h_0[r] h_0[s] \int_{-\pi}^{\pi} e^{j\omega(q+s-p-r)} d\omega \\ &\quad \sum_{k=0}^{K-1} W_K^{(-p-r+\tau_t)k} \sum_{l=0}^{K-1} W_K^{(q+s-\tau_t)l} \\ &= \frac{1}{D^2} \sum_{r=0}^{L_h-1} \sum_{s=0}^{L_h-1} h_0[r] h_0[s] \text{sinc}(q+s-p-r) \\ &\quad K \Delta_K[p+r-\tau_t] K \Delta_K[q+s-\tau_t] \end{aligned} \quad (\text{A.29})$$

where  $E_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{E}$  and

$$\Delta_K[n] = \sum_{\kappa=-\infty}^{\infty} \delta[n - \kappa K] \quad (\text{A.30})$$

i.e. a Kronecker comb function with period  $K$ . The sinc-function is zero except when  $s = p + r - q$ , so the sum over  $s$  and the sinc-function can be removed by substituting  $s = p + r - q$ .

$$\begin{aligned} E_{p,q} &= \frac{K^2}{D^2} \sum_{r=0}^{L_h-1} h_0[r] h_0[p+r-q] \\ &\quad \Delta_K[p+r-\tau_t] \Delta_K[p+r-\tau_t] \end{aligned} \quad (\text{A.31})$$

The two  $\Delta_K$ -functions are zero except when  $r = cK + \tau_t - p$  where  $c \in \mathbb{Z}$ , so the sum over  $r$  and the two  $\Delta_K$ -functions can be replaced by a sum over  $c$  by substituting  $r = cK + \tau_t - p$

$$E_{p,q} = \frac{K^2}{D^2} \sum_{c=-\infty}^{\infty} h_0[cK + \tau_t - p] h_0[cK + \tau_t - q] \quad (\text{A.32})$$

where  $h_0[n] = 0$  when  $n \neq 0, 1, \dots, L_h - 1$ .

$\mathbf{f}$  is an  $L_g \times 1$  vector defined as

$$\mathbf{f} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Re \left\{ e^{j\omega\tau_t} \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k) W_K^{\tau_t k} \right\} d\omega$$

Each element in  $\mathbf{f}$  can be calculated by

$$\begin{aligned} f_p &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Re \left\{ e^{j\omega\tau_t} \frac{1}{D} \sum_{k=0}^{K-1} e^{-j\omega p} W_K^{-pk} \sum_{r=0}^{L_h-1} h_0[r] e^{-j\omega r} W_K^{-rk} W_K^{\tau_t k} \right\} d\omega \\ &= \frac{1}{2\pi D} \sum_{r=0}^{L_h-1} h_0[r] \sum_{k=0}^{K-1} W_K^{(-p-r+\tau_t)k} \int_{-\pi}^{\pi} \Re \left\{ e^{j\omega(\tau_t-p-r)} \right\} d\omega \\ &= \frac{1}{2\pi D} \sum_{r=0}^{L_h-1} h_0[r] K \Delta_K[p+r-\tau_t] \int_{-\pi}^{\pi} \cos(\omega(\tau_t-p-r)) d\omega \\ &= \frac{K}{D} \sum_{r=0}^{L_h-1} h_0[r] \Delta_K[p+r-\tau_t] \text{sinc}(\tau_t-p-r) \end{aligned} \quad (\text{A.33})$$

where  $f_p$  is the  $p$ -th row in  $\mathbf{f}$ . The sinc-function is zero except when  $r = \tau_t - p$ , so the sum over  $r$ , the sinc-function and the  $\Delta_K$ -function can be removed by substituting  $r = \tau_t - p$

$$f_p = \frac{K}{D} h_0[\tau_t - p] \quad (\text{A.34})$$

where  $h_0[n] = 0$  when  $n \neq 0, 1, \dots, L_h - 1$ .

## A.4 Aliasing/Imaging Cancellation Error

The aliasing/imaging cancellation error,  $\epsilon_c$ , is given by

$$\epsilon_c = \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_c(e^{j\omega})|^2 d\omega \quad (\text{A.35})$$

where  $T_c(e^{j\omega})$  is given by (2.16).

### A.4.1 Rewrite Aliasing/Imaging Cancellation Error to Matrix Form

To solve the optimisation problem the following structure is desired

$$\epsilon_c = \mathbf{g}^T \mathbf{Q} \mathbf{g} \quad (\text{A.36})$$

This structure can be achieved by inserting  $T_c(e^{j\omega})$  and expanding the error function as in section A.1.1

$$\begin{aligned}
\epsilon_c &= \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \frac{1}{D} \sum_{k=0}^{K-1} H_k(e^{j\omega} W_D^d) G_k(e^{j\omega}) \\
&\quad \frac{1}{D} \sum_{l=0}^{K-1} H_l^*(e^{j\omega} W_D^d) G_l^*(e^{j\omega}) d\omega \\
&= \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \frac{1}{D} \sum_{k=0}^{K-1} \mathbf{h}^T \boldsymbol{\phi}_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_k} \mathbf{g}^T \boldsymbol{\phi}_g(e^{j\omega} W_K^k) \\
&\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \boldsymbol{\phi}_h^*(e^{j\omega} W_K^l W_D^d) W_K^{-\tau_l} \mathbf{g}^T \boldsymbol{\phi}_g^*(e^{j\omega} W_K^l) d\omega \\
&= \mathbf{g}^T \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \frac{1}{D} \sum_{k=0}^{K-1} \boldsymbol{\phi}_g(e^{j\omega} W_K^k) \mathbf{h}^T \boldsymbol{\phi}_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_k} \\
&\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \boldsymbol{\phi}_h^*(e^{j\omega} W_K^l W_D^d) W_K^{-\tau_l} \boldsymbol{\phi}_g^H(e^{j\omega} W_K^l) d\omega \mathbf{g} \\
&= \mathbf{g}^T \mathbf{Q} \mathbf{g} \quad (\text{A.37})
\end{aligned}$$

where  $\mathbf{Q}$  is an  $L_g \times L_g$  matrix

$$\begin{aligned} \mathbf{Q} &= \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ &\quad \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\ &\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l W_D^d) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \end{aligned} \quad (\text{A.38})$$

#### A.4.2 Simplifying the Aliasing/Imaging Cancellation Error

$\mathbf{Q}$  is an  $L_g \times L_g$  matrix defined as

$$\begin{aligned} \mathbf{Q} &= \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ &\quad \frac{1}{D} \sum_{k=0}^{K-1} \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\ &\quad \frac{1}{D} \sum_{l=0}^{K-1} \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^l W_D^d) W_K^{-\tau_t l} \phi_g^H(e^{j\omega} W_K^l) d\omega \end{aligned} \quad (\text{A.39})$$

Each element in  $\mathbf{Q}$  can be calculated by

$$\begin{aligned} Q_{p,q} &= \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ &\quad \frac{1}{D} \sum_{k=0}^{K-1} e^{-j\omega p} W_K^{-pk} \sum_{r=0}^{L_h-1} h_0[r] e^{-j\omega r} W_K^{-rk} W_D^{-rd} W_K^{\tau_t k} \\ &\quad \frac{1}{D} \sum_{l=0}^{K-1} \sum_{s=0}^{L_h-1} h_0[s] e^{j\omega s} W_K^{sl} W_D^{sd} W_K^{-\tau_t l} e^{j\omega q} W_K^{ql} d\omega \\ &= \frac{1}{2\pi D^2} \sum_{r=0}^{L_h-1} \sum_{s=0}^{L_h-1} h_0[r] h_0[s] \int_{-\pi}^{\pi} e^{j\omega(q+s-p-r)} d\omega \\ &\quad \sum_{k=0}^{K-1} W_K^{(-p-r+\tau_t)k} \sum_{l=0}^{K-1} W_K^{(q+s-\tau_t)l} \sum_{d=1}^{D-1} W_D^{(s-r)d} \\ &= \frac{K^2}{D^2} \sum_{r=0}^{L_h-1} \sum_{s=0}^{L_h-1} h_0[r] h_0[s] \text{sinc}(q+s-p-r) \\ &\quad \Delta_K[p+r-\tau_t] \Delta_K[q+s-\tau_t] (D\Delta_D[s-r]-1) \end{aligned} \quad (\text{A.40})$$

where  $Q_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{Q}$ . As in  $E_{p,q}$  we substitute  $s = p + r - q$  and  $r = cK + \tau_t - p$  to obtain

$$\begin{aligned} Q_{p,q} &= \frac{K^2}{D^2} \sum_{c=-\infty}^{\infty} h_0[cK + \tau_t - p]h_0[cK + \tau_t - q](D\Delta_D[p - q] - 1) \\ &= E_{p,q}(D\Delta_D[p - q] - 1) \end{aligned} \quad (\text{A.41})$$

where  $h_0[n] = 0$  when  $n \neq 0, 1, \dots, L_h - 1$  and  $E_{p,q}$  is from (A.32).

## A.5 Aliasing/Imaging Error

The aliasing/imaging error,  $\epsilon_r$ , is given by

$$\epsilon_r = \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |T_r(e^{j\omega})|^2 d\omega \quad (\text{A.42})$$

where  $T_r(e^{j\omega})$  is given by (2.17).

### A.5.1 Rewrite Aliasing/Imaging Error to Matrix Form

To solve the optimisation problem the following structure is desired

$$\epsilon_r = \mathbf{g}^T \mathbf{P} \mathbf{g} \quad (\text{A.43})$$

This structure can be achieved by inserting  $T_r(e^{j\omega})$  and expanding the error function as in section A.1.1

$$\begin{aligned} \epsilon_r &= \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{D} \sqrt{\sum_{k=0}^{K-1} |H_k(zW_D^d)G_k(z)|^2} \right|^2 d\omega \\ &= \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{D^2} \sum_{k=0}^{K-1} |H_k(zW_D^d)G_k(z)|^2 d\omega \\ &= \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\ &\quad H_k(e^{j\omega}W_D^d)G_k(e^{j\omega}) \\ &\quad H_k^*(e^{j\omega}W_D^d)G_k^*(e^{j\omega}) d\omega \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \mathbf{g}^T \phi_g(e^{j\omega} W_K^k) \\
&\quad \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^k W_D^d) W_K^{-\tau_t k} \mathbf{g}^T \phi_g^*(e^{j\omega} W_K^k) d\omega \\
&= \mathbf{g}^T \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\
&\quad \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^k W_D^d) W_K^{-\tau_t k} \phi_g^H(e^{j\omega} W_K^l) d\omega \mathbf{g} \\
&= \mathbf{g}^T \mathbf{P} \mathbf{g} \tag{A.44}
\end{aligned}$$

where  $\mathbf{P}$  is an  $L_g \times L_g$  matrix

$$\begin{aligned}
\mathbf{P} &= \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\
&\quad \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^k W_D^d) W_K^{-\tau_t k} \phi_g^H(e^{j\omega} W_K^l) d\omega \tag{A.45}
\end{aligned}$$

### A.5.2 Simplifying the Aliasing/Imaging Error

$\mathbf{P}$  is an  $L_g \times L_g$  matrix defined as

$$\begin{aligned}
\mathbf{P} &= \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad \phi_g(e^{j\omega} W_K^k) \mathbf{h}^T \phi_h(e^{j\omega} W_K^k W_D^d) W_K^{\tau_t k} \\
&\quad \mathbf{h}^T \phi_h^*(e^{j\omega} W_K^k W_D^d) W_K^{-\tau_t k} \phi_g^H(e^{j\omega} W_K^l) d\omega \tag{A.46}
\end{aligned}$$

Each element in  $\mathbf{P}$  can be calculated by

$$\begin{aligned}
P_{p,q} &= \frac{1}{D^2} \sum_{k=0}^{K-1} \sum_{d=1}^{D-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} \\
&\quad e^{-j\omega p} W_K^{-pk} \sum_{r=0}^{L_h-1} h_0[r] e^{-j\omega r} W_K^{-rk} W_D^{-rd} W_K^{\tau_t k} \\
&\quad \sum_{s=0}^{L_h-1} h_0[s] e^{j\omega s} W_K^{sk} W_D^{sd} W_K^{-\tau_t k} e^{j\omega q} W_K^{qk} d\omega
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi D^2} \sum_{r=0}^{L_h-1} \sum_{s=0}^{L_h-1} h_0[r]h_0[s] \int_{-\pi}^{\pi} e^{j\omega(q+s-p-r)} d\omega \\
&\quad \sum_{k=0}^{K-1} W_K^{(q+s-p-r)k} \sum_{d=1}^{D-1} W_D^{(s-r)d} \\
&= \frac{K}{D^2} \sum_{r=0}^{L_h-1} \sum_{s=0}^{L_h-1} h_0[r]h_0[s] \text{sinc}(q+s-p-r) \\
&\quad \Delta_K[q+s-p-r](D\Delta_D[s-r]-1) \tag{A.47}
\end{aligned}$$

where  $P_{p,q}$  is the  $p$ -th row and  $q$ -th column in  $\mathbf{P}$ . The sinc-function is zero except when  $s = r + p - q$ , so the sum over  $s$ , the sinc-function and the  $\Delta_K$ -function can be removed by substituting  $s = r + p - q$

$$\begin{aligned}
P_{p,q} &= \frac{K}{D^2} \sum_{r=0}^{L_h-1} h_0[r]h_0[r+p-q](D\Delta_D[p-q]-1) \\
&= \frac{K}{D^2} \gamma_{h_0}[q-p](D\Delta_D[p-q]-1) \tag{A.48}
\end{aligned}$$

where  $\gamma_{h_0}[m]$  is the raw autocorrelation of  $h_0[n]$  at lag  $m$ , i.e.

$$\gamma_{h_0}[m] = \sum_n h_0[n]h_0[n-m] \tag{A.49}$$



# Bibliography

---

- [AAUA60] American National Standards Institute, American Standards Association, United States of America Standards Institute, and Acoustical Society of America. *Acoustical Terminology (including Mechanical Shock and Vibration): ANSI ASA S1.1-1960, Revision of Z24.1-1951 and Including Z24.1a, UDC 001.4:534 + 681.8*. United States of America Standards Institute, 1960.
- [CR83] R. E. Crochiere and L. R. Rabiner. *Multirate digital signal processing*. Prentice-Hall Signal Processing Series. Prentice Hall, 1983.
- [CRALMBL02] F. Cruz-Roldán, P. Amo-López, S. Maldonado-Bascón, and S. S. Lawson. An efficient and simple method for designing prototype filters for cosine-modulated pseudo-qmf banks. *Signal Processing Letters, IEEE*, 9(1):29–31, 2002.
- [Cro80] R. E. Crochiere. A weighted overlap-add method of short-time fourier analysis/synthesis. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 28(1):99–102, 1980.
- [dH01] J. M. de Haan. *Filter Bank Design for Subband Adaptive Filtering: Methods and Applications*. Licentiate dissertation, Blekinge Institute of Technology, 2001.
- [dHGCN01] J. M. de Haan, N. Grbić, I. Claesson, and S. Nordholm. Design of oversampled uniform dft filter banks with delay specification using quadratic optimization. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001*

- IEEE International Conference on*, volume 6, pages 3633–3636. IEEE, 2001.
- [dHGCN03] J. M. de Haan, N. Grbić, I. Claesson, and S. Nordholm. Filter bank design for subband adaptive microphone arrays. *Speech and Audio Processing, IEEE Transactions on*, 11(1):14–23, 2003.
- [DKK97a] T. Dau, B. Kollmeier, and A. Kohlrausch. Modeling auditory processing of amplitude modulation. i. detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, 102(5):2892–2905, 1997.
- [DKK97b] T. Dau, B. Kollmeier, and A. Kohlrausch. Modeling auditory processing of amplitude modulation. ii. spectral and temporal integration. *The Journal of the Acoustical Society of America*, 102(5):2906–2919, 1997.
- [DNC05] H. H. Dam, S. Nordholm, and A. Cantoni. Uniform fir filter-bank optimization with group delay specifications. *Signal Processing, IEEE Transactions on*, 53(11):4249–4260, 2005.
- [DNCdH04] H. H. Dam, S. Nordholm, A. Cantoni, and J. M. de Haan. Iterative method for the design of dft filter bank. *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 51(11):581–586, 2004.
- [EM01] K. Eneman and M. Moonen. Dft modulated filter bank design for oversampled subband systems. *Signal Processing*, 81(9):1947–1973, 2001.
- [GdHCN01] N. Grbić, J. M. de Haan, I. Claesson, and S. Nordholm. Design of oversampled uniform dft filter banks with reduced inband aliasing and delay constraints. In *Signal Processing and its Applications, Sixth International, Symposium on. 2001*, volume 1, pages 104–107. IEEE, 2001.
- [GL84] D. Griffin and J. Lim. Signal estimation from modified short-time fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(2):236–243, 1984.
- [GM90] B. R. Glasberg and B. C. J. Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1):103–138, 1990.
- [GM00] B. R. Glasberg and B. C. J. Moore. Frequency selectivity as a function of level and frequency measured with uniformly

- exciting notched noise. *The Journal of the Acoustical Society of America*, 108:2318, 2000.
- [HKS<sup>+</sup>00] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi. Frequency-warped signal processing for audio applications. *Journal of the Audio Engineering Society*, 48(11):1011–1031, 2000.
- [HL07] Y. Hu and P. C. Loizou. Subjective comparison and evaluation of speech enhancement algorithms. *Speech communication*, 49(7):588–601, 2007.
- [HL08] Y. Hu and P. C. Loizou. Evaluation of objective quality measures for speech enhancement. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(1):229–238, 2008.
- [HP00] H.-G. Hirsch and D. Pearce. The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)*, 2000.
- [HS08] E. Hänsler and G. Schmidt. *Speech and audio processing in adverse environments*. Springer, 2008.
- [ITU01] ITU-T. *Rec. P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, 2001.
- [ITU03] ITU-T. *Rec. P.862.1: Mapping function for transforming P.862 raw result scores to MOS-LQO*, 2003.
- [ITU06a] ITU-T. *Rec. P.800.1: Mean Opinion Score (MOS) terminology*, 2006.
- [ITU06b] ITU-T. *Rec. P.835: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*, 2006.
- [ITU11] ITU-T. *Rec. P.56: Objective measurement of active speech level*, 2011.
- [Kat04] J. M. Kates. Hearing-aid compression using digital frequency warping. *Proc. 18th Inter. Congr. on Acoustics (Kyoto, Japan, 2004)*, 2004.
- [LGK09] K. A. Lee, W. S. Gan, and S. M. Kuo. *Subband Adaptive Filtering: Theory and Implementation*. Wiley, 2009.

- [Loi13] P. C. Loizou. *Speech enhancement: theory and practice*. CRC press, 2013.
- [Löl11] H. W. Löllmann. *Allpass-based analysis-synthesis filterbanks: design and application*. PhD thesis, RWTH Aachen University, 2011.
- [LPM01] E. A. Lopez-Poveda and R. Meddis. A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, 110:3107, 2001.
- [LV98] Y.-P. Lin and P. P. Vaidyanathan. A kaiser window approach for the design of prototype filters of cosine modulated filterbanks. *Signal Processing Letters, IEEE*, 5(6):132–134, 1998.
- [Lyo11] R. F. Lyon. Cascades of two-pole–two-zero asymmetric resonators are good models of peripheral auditory function. *The Journal of the Acoustical Society of America*, 130(6):3893–3904, 2011.
- [Mer99] A. Mertins. *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*. Ultrasound in Biomedicine Research Series. Wiley, 1999.
- [Moo12] B. C. J. Moore. *An Introduction to the Psychology of Hearing*. Emerald, 2012.
- [OSB99] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck. *Discrete-time signal processing*. Prentice-Hall signal processing series. Prentice Hall, 1999.
- [Pat86] R. D. Patterson. Auditory filters and excitation patterns as representations of frequency resolution. *Frequency selectivity in hearing*, pages 123–177, 1986.
- [PB87] T. W. Parks and C. S. Burrus. *Digital filter design*. Topics in digital signal processing. Wiley, 1987.
- [Pla05] C. J. Plack. *The sense of hearing*. Lawrence Erlbaum Associates Publishers, 2005.
- [PM07] J. G. Proakis and D. G. Manolakis. *Digital Signal Processing*. Pearson Prentice Hall, 2007.
- [PNSHR87] R. D. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice. An efficient auditory filterbank based on the gammatone function. In *A meeting of the IOC Speech Group on Auditory Modelling at RSRE*, volume 2, 1987.

- [PNSWM82] R. D. Patterson, I. Nimmo-Smith, D. L. Weber, and R. Milroy. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, 72:1788, 1982.
- [Smi11] J. O. Smith. *Spectral Audio Signal Processing*. W3K, 2011.
- [Vai93] P. P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice Hall Signal Processing Series. Prentice Hall, 1993.
- [WTRD08] C. Wu, K. L. Teo, V. Rehbock, and H. H. Dam. Global optimum design of uniform fir filter bank with magnitude constraints. *Signal Processing, IEEE Transactions on*, 56(11):5478–5486, 2008.
- [YGNT04] K. F. C. Yiu, N. Grbić, S. Nordholm, and K. L. Teo. Multicriteria design of oversampled uniform dft filter banks. *Signal Processing Letters, IEEE*, 11(6):541–544, 2004.
- [ZF99] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and models*, volume 2. Springer Berlin, 1999.