



KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Electrical Engineering

Digital signal processing algorithms for noise reduction, dynamic range compression, and feedback cancellation in hearing aids

Kim Ngo

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering

July 2011

Digital signal processing algorithms for noise reduction, dynamic range compression, and feedback cancellation in hearing aids

Kim Ngo

Jury:

Prof. em. dr. ir. Y. Willems, chairman

Prof. dr. ir. M. Moonen, promotor

Prof. dr. J. Wouters, co-promotor

Prof. dr. ir. S. H. Jensen, co-promoter
(Aalborg University, Denmark)

Prof. dr. ir. S. Doclo
(University of Oldenburg, Germany)

Prof. dr. ir. W. Verhelst, assessor
(Vrije Universiteit Brussel, Belgium)

Prof. dr. ir. H. Van Hamme, assessor

Prof. dr. ir. J. Vandewalle

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering

July 2011

© Katholieke Universiteit Leuven – Faculty of Engineering
Kasteelpaark Arenberg 1/2200, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 978-94-6018-389-8
D/2011/7515/91

Preface

A five-year journey has come to an end and I am finally ready to write the preface for my PhD thesis. At this moment, I don't have that much to say other than to express my gratitude and countless thanks to all of those who have helped me during my PhD.

I would like to thank Prof. Marc Moonen for giving me the opportunity to join his research group and for the guidance. The support and feedback from my co-promoters Prof. Jan Wouters and Prof. Søren Holdt Jensen has without any doubt been very helpful. This thesis has been build on a number of collaboration with other researchers and I would therefore like to send a special thank to Toon van Waterschoot and Ann Spriet. During the period of my research I visited Aalborg University and I would like to send a special gratitude to Prof. Mads Græsbøll Christensen who introduced me to pitch estimation which has shown to be very fruitful in my research. Another great experience was the opportunity to visit University of Illinois at Urbana-champagne and especially Prof. Douglas L. Jones.

I would also like to thank the jury members: Prof. Simon Doclo, Prof. Werner Verhelst, Prof. Hugo Van Hamme, Prof. Joos Vandewalle, and Prof. Yves Willems (chairman) for their time, effort, and valuable comments and suggestions to improve my thesis.

To the research group at the Katholieke Universiteit Leuven: Geert V.M., Geert R., Vincent, Paschal, Jan, Deepak, Pepe, Geert C., Alexander, Bruno, Beier, Amir, Rodrigo, Javi, Joseph, Sam, Guang, and Sylwester. Thank you all for the wonderful moments and discussions. Bram for being my Dutch translator when needed. To the people in the SIGNAL project: Mikael, Matthias, Pietro, Elena, Many, Nuria, Li Jun, Jean-marc, and Johan thank you all for the many travels and courses that we had together. I would also like to thank David and Eric from UIUC who offered their help when I just arrived and they both made my visit more pleasant. A special thanks to my good friends Prabin, Romain, and Daniele.

I would also like to thank my family and friends in Denmark for supporting me through my PhD.

Kim Ngo
Leuven, July 2011

Abstract

Hearing loss can be caused by many factors, e.g., daily exposure to excessive noise in the work environment and listening to loud music. Another important reason can be age-related, i.e., the slow loss of hearing that occurs as people get older. In general hearing impaired people suffer from a frequency-dependent hearing loss and from a reduced dynamic range between the hearing threshold and the uncomfortable level. This means that the uncomfortable level for normal hearing and hearing impaired people suffering from so called sensorineural hearing loss remains the same but the hearing threshold and the sensitivity to soft sounds are shifted as a result of the hearing loss. To compensate for this kind of hearing loss the hearing aid should include a frequency-dependent and a level-dependent gain. The corresponding digital signal processing (DSP) algorithm is referred to as dynamic range compression (DRC). Background noise (from competing speakers, traffic etc.) is also a significant problem for hearing impaired people who indeed have more difficulty understanding speech in noise and so in general need a higher signal-to-noise-ratio (SNR) than people with normal hearing. Because of this the noise reduction (NR) is also an important algorithmic component in hearing aids. Another issue in hearing aids is the undesired acoustic coupling between the loudspeaker and the microphone which is referred to as the acoustic feedback problem. Acoustic feedback produces an annoying howling sound and limits the maximum amplification that can be used in the hearing aid without making it unstable. To tackle the acoustic feedback problem adaptive feedback cancellation (AFC) algorithms are used. Acoustic feedback is becoming an even more significant problem due to the use of open fittings and the decreasing distance between the microphone and the loudspeaker.

In this thesis several DSP techniques are presented to address the problems introduced above. For the background noise problem, we propose a NR algorithm based on the speech distortion weighted multi-channel Wiener filter (SDW-MWF) that is designed to allow for a trade-off between NR and speech distortion. The first contribution to the SDW-MWF based NR is based on using a weighting factor that is updated for each frequency and for each frame such that speech dominant segments and noise dominant segments can be weighted differently. This can be

done by incorporating the conditional speech presence probability (SPP) in the SDW-MWF. The second contribution is based on an alternative and more robust method to estimate and update the correlation matrices, which is very important since an SDW-MWF based NR is uniquely based on these correlation matrices. The proposed SDW-MWF based NR shows better performance in terms of SNR improvement and signal distortion compared to a traditional SDW-MWF.

For the problem of background noise and reduced dynamic range, we propose a combined algorithm of an SDW-MWF based NR and DRC. First the DRC is extended to a dual-DRC approach that allows for a switchable compression characteristic based on the conditional SPP. Secondly the SDW-MWF incorporating the conditional SPP is combined and analysed together with the dual-DRC. The proposed method shows that the SNR degradation can be partially controlled by using the dual-DRC.

For the acoustic feedback problem, we propose a prediction error method based AFC (PEM-based AFC) exploiting an improved cascaded near-end signal model. The challenge in PEM-based AFC is to accurately estimate the near-end signal model such that the inverse of this model can be used as a decorrelation of the loudspeaker and the microphone signals. Due to the closed signal loop the loudspeaker and the microphone signal are now correlated which causes standard adaptive filtering methods to fail. The proposed PEM-based AFC shows improved performance in terms of maximum stable gain (MSG) and filter misadjustment compared to a PEM-based AFC using a single near-end signal model.

Korte Inhoud

Gehoorverlies kan worden veroorzaakt door vele factoren, voorbeelden zijn dagelijkse blootstelling aan overmatig lawaai in de werkomgeving of luisteren naar luide muziek. Een andere belangrijke reden is gerelateerd aan de leeftijd, met name de langzame achteruitgang van het gehoor die optreedt als mensen ouder worden. In het algemeen lijden slechthorenden aan een frequentieafhankelijk gehoorverlies en aan een verminderd dynamisch bereik tussen de gehoordrempel en het oncomfortabele niveau. Dit betekent dat het oncomfortabele niveau voor normaalhorenden en slechthorenden, die lijden aan een zogenaamd sensorineuraal verlies, hetzelfde blijft, terwijl de gehoordrempel en de gevoeligheid voor zachte geluiden worden verschoven ten gevolge van het gehoorverlies. Ter compensatie voor dit soort van gehoorverlies moet het hoorapparaat een frequentieafhankelijke en niveauafhankelijke versterking toepassen. Het corresponderende digitale signaalverwerkingsalgoritme (DSP) is het zogenaamde Dynamisch Bereik Compressie-algoritme (DRC). Achtergrondgeluiden (van door elkaar pratende personen, verkeer enz.) vormen ook een groot probleem voor slechthorenden, die inderdaad meer moeite hebben met spraakverstaan in ruis en over het algemeen dus behoefte hebben aan een hogere signaal-ruisverhouding (SNR) dan normaalhorenden. Hierdoor kan ruisonderdrukking (NR) ook worden beschouwd als een belangrijke algoritmische component in hoorapparaten. Een ander probleem in hoorapparaten is de ongewenste akoestische koppeling tussen de luidspreker en de microfoon, die wordt aangeduid als het akoestische terugkoppelings- of feedbackprobleem. Akoestische terugkoppeling produceert een irritant fluitend geluid en beperkt de maximale versterking die in het hoortoestel kan worden toegepast zonder dat het onstabiel wordt. Ter bestrijding van het akoestische terugkoppelingsprobleem worden Adaptieve Feedbackonderdrukkingsalgoritmes (AFC) gebruikt. Akoestische terugkoppeling is recentelijk een nog groter probleem geworden door het gebruik van open aanpassingen en de afnemende afstand tussen de microfoon en de luidspreker.

In dit proefschrift worden verschillende DSP-technieken gepresenteerd om de problemen aan te pakken die hierboven werden geïntroduceerd. Voor het achtergrondgeluid probleem, stellen we een NR algoritme voor dat is gebaseerd

op de spraak distortie gewogen meerkanaals Wiener filter (SDW-MWF), die ontworpen is om een afweging tussen NR en spraak distortie mogelijk te maken. De eerste bijdrage aan de SDW-MWF gebaseerde NR is gebaseerd op het gebruik van een wegingsfactor, die wordt bijgewerkt voor elke frequentie en voor elk frame, zodanig dat spraak-dominante segmenten en ruis-dominante segmenten op een verschillende manier kunnen gewogen worden. Dit kan gedaan worden door het opnemen van de voorwaardelijke kans op spraak aanwezigheid (SPP) in de SDW-MWF. De tweede bijdrage is gebaseerd op een alternatieve en robuustere methode om correlatie matrices te schatten en bij te werken, wat heel belangrijk is aangezien de SDW-MWF gebaseerde NR enkel gebruik maakt van deze correlatie matrices. De voorgestelde SDW-MWF gebaseerde NR toont betere prestaties in termen van SNR verbetering en spraak distortie, vergeleken met een traditionele SDW-MWF.

Voor het probleem van achtergrondlawaai en verminderd dynamisch bereik, stellen we een combinatie van SDW-MWF gebaseerde NR en DRC voor. Eerst wordt de DRC uitgebreid met een duale DRC benadering die een omschakeling van de compressie karakteristiek op basis van de voorwaardelijke SPP toelaat. Ten tweede wordt de SDW-MWF met voorwaardelijke SPP samen met de duale DRC gecombineerd en geanalyseerd. De voorgestelde methode toont aan dat de SNR degradatie gedeeltelijk kan worden gecontroleerd met behulp van de duale DRC.

Voor het akoestische terugkoppelingsprobleem, stellen we een Predictie Fout Methode-gebaseerde AFC (PEM-gebaseerde AFC) voor, waarbij een verbeterd gecascadeerd bronsignaalmodel wordt aangewend. De uitdaging in PEM-gebaseerde AFC is een nauwkeurige schatting van het bronsignaalmodel te bekomen zodat de inverse van dit model gebruikt kan worden als decorrelatie van de luidspreker en de microfoonsignalen. Door de gesloten signaallus zijn de luidspreker en de microfoonsignalen nu gecorreleerd waardoor standaard adaptieve filtering methodes mislukken. De voorgestelde PEM-gebaseerde AFC toont verbeterde prestaties in termen van maximale stabiele versterking (MSG) en filter misaanpassing, vergeleken met een PEM-gebaseerde AFC met een enkelvoudig bronsignaalmodel.

Nomenclature

Mathematical Notation

a	scalar a
\mathbf{a}	vector \mathbf{a}
\mathbf{A}	matrix \mathbf{A}
$\mathbf{A}^T, \mathbf{a}^T$	transpose of matrix \mathbf{A} , vector \mathbf{a}
$\mathbf{A}^H, \mathbf{a}^H$	Hermitian transpose of matrix \mathbf{A} , vector \mathbf{a}
$\hat{a}, \hat{\mathbf{a}}, \hat{\mathbf{A}}$	estimate of scalar a , vector \mathbf{a} , matrix \mathbf{A} .
$\varepsilon\{\cdot\}$	expectation operator
$\text{Tr}\{\cdot\}$	trace operator
$ \cdot $	absolute value
$\ \cdot\ $	2-norm
t	discrete time variable
\in	element of
\mathbb{C}	set of complex numbers
ω	radial frequency variable (rad)
\log_{10}	common logarithm
$\max(\cdot)$	maximum
$\min(\cdot)$	minimum
$\exp(\cdot)$	exponential operator

Fixed Symbols

$d(t)$	feedback compensated signal
\mathbf{e}_l	l -th canonical vector
f_s	sampling frequency
\mathbf{f}	feedback path impulse response vector
$\hat{\mathbf{f}}(t)$	estimated feedback path impulse response vector

$F(q, t)$	feedback path model
$H(q, t)$	near-end signal model
k	frequency bin index
l	frame index
M	number of microphones
n_F	feedback path model order
$r(t)$	source excitation signal
T_{60}	reverberation time
$u(t)$	loudspeaker signal
$x(t)$	microphone signal
$X_i^s(k, l)$	speech component in the i -th microphone
$X_i^n(k, l)$	noise component in the i -th microphone
$X_i(k, l)$	i -th microphone signal
$\mathbf{X}^s(k, l)$	stacked speech vector
$\mathbf{X}^n(k, l)$	stacked noise vector
$\mathbf{X}^s(k, l)$	stacked data vector
$v(t)$	near-end signal
$\mathbf{W}(k, l)$	stacked filter vector of multi-channel noise reduction
$y(t)$	microphone signal
$Z(k, l)$	output of the noise reduction algorithm
μ	weighting factor to trade-off between noise reduction and speech distortion
α_n	exponential weighting factor for the noise correlation matrix
α_x	exponential weighting factor for the speech-plus-noise correlation matrix
$\varepsilon(t)$	prediction error

Acronyms and Abbreviations

AFC	Adaptive Feedback Cancellation
AR	autoregressive
BTE	Behind-the-ear
CIC	Completely-in-the-canal
CPZLP	Constrained Pole-Zero Linear Prediction
dB	Decibels
DRC	Dynamic Range Compression
DSP	Digital Signal Processing
e.g.	<i>exempli gratia</i> : for example
etc.	<i>et cetera</i> : and so forth
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GSC	Generalized Sidelobe Canceller

HRTF	Head-Related Transfer Function
Hz	hertz
i.e.	<i>id est</i> : that is
IFFT	Inverse Fast Fourier Transform
IIR	Infinite Impulse Response
ITC	In-the-canal
ITE	In-the-ear
kHz	kilohertz
LMS	Least Mean Squares
LP	Linear Prediction
ms	milliseconds
MMSE	Minimum Mean Square Error
MSG	Maximum Stable Gain
MVDR	Minimum Variance Distortionless Response
MWF	Multi-channel Wiener Filter
NIHL	Noise-induced hearing loss
NR	Noise Reduction
PEM	Prediction Error Method
PEM-AFC	PEM-based AFC
PZLP	Pole-Zero Linear Prediction
RCB	Robust Capon Beamformer
SAP	Speech Absence Probability
SCB	Standard Capon Beamformer
SD	Signal Distortion
SFM	Spectral Flatness Measure
SDW-MWF	Speech Distortion Weighted MWF
SPL	Sound Pressure Level
SNR	Signal-to-Noise-Ratio
SPP	Speech Presence Probability
STFT	Short-Time Fourier Transform
vs.	versus
VAD	Voice Activity Detection

Contents

Contents	xi
1 Introduction	1
1.1 Preliminaries	2
1.1.1 Hearing impairment	2
1.1.2 Some statistics	3
1.1.3 Commercial hearing aids	7
1.1.4 Characterization of signals	8
1.1.5 Acoustic environment	9
1.1.6 Reduced dynamic range	10
1.1.7 Acoustic feedback	11
1.1.8 Signal processing challenges	11
1.2 Noise reduction in hearing aids	13
1.2.1 Single-channel noise reduction	15
1.2.2 Multi-channel noise reduction	17
1.3 Dynamic range compression in hearing aids	21
1.3.1 Design of DRC algorithms	21
1.3.2 Perceptual benefits from DRC	22
1.4 Feedback cancellation in hearing aids	24
1.4.1 Feedforward suppression	25

1.4.2	Feedback cancellation	25
1.4.3	Bias problem and decorrelation	26
1.5	Outline of the thesis and main contributions	28
1.5.1	Main research objectives	28
1.5.2	Chapter by chapter outline and contributions	28
2	Speech distortion weighted multi-channel Wiener filter (SDW-MWF_{μ})	33
2.1	Preliminaries	34
2.1.1	Estimation of correlation matrices	35
2.2	Multi-channel Wiener filter (MWF)	36
2.3	Speech distortion weighted MWF (SDW-MWF _{μ})	38
2.4	Rank-1 SDW-MWF _{μ}	39
2.5	Analysis of the SDW-MWF _{μ}	40
2.5.1	Robustness and tracking	42
2.6	Experimental results	42
2.6.1	Experimental set-up	43
2.6.2	Performance measures	44
2.6.3	Results	45
2.7	Conclusion	46
3	SDW-MWF_{μ} based on speech presence probability (SPP)	49
3.1	Conditional speech presence probability (SPP)	50
3.1.1	Multi-channel a priori and a posteriori SNR estimation	51
3.1.2	A priori speech absence probability (SAP) estimation	52
3.2	SDW-MWF incorporating the conditional SPP (SDW-MWF _{SPP})	53
3.2.1	Derivation of SDW-MWF _{SPP}	54
3.2.2	Combined solution	55
3.3	SDW-MWF incorporating a flexible weighting factor (SDW-MWF _{Flex})	56

3.4 Rank-1 SDW-MWF incorporating the conditional SPP 60

3.5 Experimental results 62

 3.5.1 Experimental set-up 62

 3.5.2 Results 62

3.6 Conclusion 63

4 SDW-MWF_μ based on robust estimation of the correlation matrices 67

4.1 Robust estimation of the correlation matrices 68

 4.1.1 Uncertainty of the correlation matrices 68

 4.1.2 Continuous updating of the correlation matrices 69

 4.1.3 Selection of prior correlation matrices 70

4.2 Analysis of estimation errors 71

4.3 Experimental results 72

 4.3.1 Experimental set-up 72

 4.3.2 Results 72

4.4 Conclusion 74

5 Robust Capon beamforming for small arrays 83

5.1 Introduction 84

5.2 Standard Capon Beamforming (SCB) 85

 5.2.1 Optimization criterion for SCB 85

 5.2.2 Mismatch between presumed and actual steering vector 85

5.3 Previous work on robust Capon beamformers 85

 5.3.1 Linearly constrained minimum variance 86

 5.3.2 Diagonal-loading-based beamformer 86

 5.3.3 Uncertainty-based beamformer 86

 5.3.4 Max-min optimization 87

5.4 Robust Capon beamforming (RCB) 89

5.4.1	Proposed RCB formulation	89
5.4.2	Gradient update of the steering vector	91
5.4.3	Computational complexity	92
5.5	Experimental results	93
5.5.1	Experimental set-up	93
5.5.2	Results	95
5.6	Conclusion	97
6	Dynamic range compression (DRC)	101
6.1	Design of DRC algorithms	102
6.1.1	Multi-band compression	102
6.1.2	DRC parameters	103
6.2	The effect of background noise on DRC	105
6.2.1	Undesired amplification over frequencies	105
6.2.2	Undesired amplification over time	107
6.2.3	Compensation of speech and noise dominant segments	108
6.3	Experimental results	109
6.3.1	Experimental set-up	109
6.3.2	Analysis procedure	110
6.3.3	Results	111
6.4	Conclusion	112
7	SDW-MWF based noise reduction and dynamic range compression	119
7.1	Problem statement and motivation	120
7.2	Combined SDW-MWF _{μ} based NR and DRC	122
7.3	Combined SDW-MWF _{spp} based NR and dual-DRC	124
7.4	Combined SDW-MWF _{flex} based NR and flex dual-DRC	127
7.5	Experimental results	128

- 7.5.1 Experimental set-up 128
- 7.5.2 Results 129
- 7.6 Conclusion 132
- 8 Prediction error method-based adaptive feedback cancellation 135**
 - 8.1 Adaptive feedback cancellation (AFC) 136
 - 8.1.1 Prediction error method 137
 - 8.2 Single near-end signal model 138
 - 8.3 Cascaded near-end signal model 138
 - 8.4 Experimental results 139
 - 8.4.1 Experimental set-up 139
 - 8.4.2 Performance measures 140
 - 8.4.3 Results 140
 - 8.5 Conclusion 141
- 9 PEM-based AFC using a harmonic sinusoidal near-end signal model 143**
 - 9.1 Harmonic sinusoidal near-end signal model 144
 - 9.1.1 Optimal-filtering based pitch estimation 144
 - 9.1.2 Subspace-orthogonality based pitch estimation 145
 - 9.1.3 Subspace-shift-invariance based pitch estimation 146
 - 9.1.4 Amplitude and models order estimation 146
 - 9.2 PZLP using pitch estimation based PEF 147
 - 9.2.1 Incorporating amplitude, order and pitch information . . . 147
 - 9.3 Voiced-unvoiced detection 149
 - 9.3.1 ZCR and energy based voiced-unvoiced detection 150
 - 9.3.2 Spectral flatness of the residual 151
 - 9.4 Experimental results 152
 - 9.4.1 Experimental set-up 153

9.4.2 Results	153
9.5 Conclusion	155
10 Conclusion and further research	161
10.1 Conclusion	161
10.1.1 Noise reduction	161
10.1.2 Combined noise reduction and dynamic range compression	163
10.1.3 Feedback cancellation	164
10.2 Suggestions for further research	165
10.2.1 Noise reduction	165
10.2.2 Combined noise reduction and dynamic range compression	166
10.2.3 Feedback cancellation	167
Bibliography	169
List of publications	191
Curriculum vitae	193

Chapter 1

Introduction

Digital signal processing (DSP) is widely used to manipulate, modify, enhance or filter signals such as speech, audio, image and telecommunication signals [41][69][80][145][176][177][203][239]. These signals can be processed in the analog domain but the digital domain offers high speed, better accuracy, greater flexibility, increased storing capabilities, and simpler implementation. DSP has become a fundamental area of research for many real-world applications, e.g., mobile phones, digital cameras, GPS, video/tele-conference, radar, MP3 players and many more. The work presented here is focused on DSP for hearing aids which is important for hearing impaired people in order to communicate and interact with other people in the daily life. It should be mentioned that some of the algorithms developed here can be applied to, e.g., hands-free telephony, in-vehicle communication, and public address systems. The two types of technology for hearing aids are analog and digital [45][100][109]. The majority of hearing aids sold today are digital namely because of the increased performance and flexibility compared to analog hearing aids. Current state-of-the-art hearing aids are exploiting various aspects of DSP and according to [216][217] 93 percent of all hearing aids sold in 2005 were digital. The core function of traditional hearing aids is mainly based on signal amplification. However, digital hearing aids allow for more advanced signal processing since the purpose of modern hearing aids is not only to amplify sounds.

This dissertation addresses several topics in DSP for hearing aids, namely noise reduction (NR), dynamic range compression (DRC), and adaptive feedback cancellation (AFC) which is only a subset of DSP algorithms that are used to build a digital hearing aid. The design of NR, DRC and AFC is closely related and equally important. The purpose of DRC is to make the speech signal audible by providing proper amplification. However acoustic feedback limits the amplification and therefore AFC is included to make the hearing aid stable. Reducing acoustic

feedback increases the available gain and allows the hearing aid to get closer to the prescribed gain. Making speech audible does not mean that hearing aid users can understand the speech without enhancement of, e.g., spectral or spatial signal information. This of course becomes more crucial when the hearing aid user is listening in the presence of background noise which makes NR an important component as well.

In this introduction, we will briefly motivate and explain the problems related to hearing aids and hearing impairment. An overview of open problems and state-of-the-art DSP algorithms in the areas of NR, DRC and AFC will also be discussed. At the end of the introduction we will explain how this work fits within the current open problems in hearing aids and point out the main contributions of this work together with a chapter-by-chapter outline.

1.1 Preliminaries

1.1.1 Hearing impairment

Hearing impairment is becoming more common and can be caused by many reasons. The most important reason is age-related (high-frequency hearing loss), i.e., the slow loss of hearing that occurs as people get older [70]. Other reasons are daily exposure to excessive noise in the work environment (construction site, factory etc.) [143] and listening to loud music (MP3 players, iPod, concerts, night clubs etc.) [191]. In general two factors can be mentioned as the primary reasons that can cause hearing loss, i.e., the level of the sound and the duration that people are exposed to this sound. This can damage the inner ear or more specifically the inner and outer hair cells (outer hair cells are more susceptible to noise exposure than inner hair cells) which is referred to as noise-induced hearing loss (NIHL) [123]. The function of these hair cells is to convert sound energy into electrical signals that are sent to the brain by the auditory nerve.

In our daily-life we are often exposed to sounds with high intensity without realizing the danger to our hearing abilities. The consequence of NIHL can typically not be reversed by surgical or medical procedures, i.e., once the hair cells are damaged they cannot grow back again. Typically the damage is done when people realize that they have a NIHL [39]. Sound levels are typically measured in decibels (dB) which is not necessarily something that we think about when we are in various environments. On a dB-scale an increase of 10 means that a sound is 10 times more intense and this will sound twice as loud to our ears. To give a perspective on the different sound levels that we can be exposed to some examples are shown in Figure 1.1. Figure 1.2 shows hazardous exposure limits for various sound levels. This shows that the louder the sound is the shorter the time is before

Degree of hearing loss	Hearing loss range (dB HL)	Effect
Normal	-10 to 15	
Slight	16 to 25	Difficulty understanding normal speech
Mild	26 to 40	
Moderate	41 to 55	Difficulty understanding loud speech
Moderately severe	56 to 70	
Severe	71 to 90	Can understand only amplified speech
Profound	91+	Difficulty understanding amplified speech

Table 1.1: Degree of hearing loss.

NIHL occurs. Sounds less than 75dB are unlikely to cause NIHL even after a long exposure time. Another factor that can play a role is of course the distance to the sound source(s).

In general the degree of hearing loss can be classified into the following categories, see Table 1.1. For a perspective the degree of hearing loss can be compared to the level and frequency of average speech which is shown in Figure 1.3. For hearing impaired people with mild to moderate hearing loss a hearing aid is needed in specific situations or at least on a frequent basis. For severe hearing loss a hearing aid is needed for all communications and for profound hearing loss the use of a hearing aid may be combined with speech-reading (lip-reading) or sign language. Furthermore there exist three distinct types of hearing loss, i.e.,

- **Sensorineural hearing loss** results from damages to the hair cells in the cochlea in the inner ear.
- **Conductive hearing loss** occurs when the ability to conduct sound from the external and middle ear into the inner ear is lost.
- **Mixed hearing loss**, i.e., combined sensorineural and conductive hearing loss.

1.1.2 Some statistics

The exact number of hearing impaired people world wide is unknown but here we will provide some statistics in order to give a perspective on the hearing loss problem. According to [195] 71 million adults in 2006 aged 18 to 80 years in Europe have a hearing loss. In the European Union alone the number is 55 million. Table 1.2 shows the hearing loss statistics for specific countries in Europe [195] and in

Painful

150 dB = rock music peak
140 dB = fire arms, air raid siren, jet engine
130 dB = jackhammer, motorcycles
120 dB = jet plane take-off, car stereo, band practice

Extremely Loud

115 dB = headphones
110 dB = rock music, model airplane, and chain saw
106 dB = timpani, bass drum rolls, and concerts
100 dB = snowmobile, chain saw, pneumatic drill
90 dB = lawnmower, shop tools, truck traffic, subway

Very Loud

80 dB = heavy city traffic
80 dB = alarm clock, busy street
70 dB = busy traffic, vacuum cleaner
60 dB = conversation, dishwasher

Moderate

50 dB = moderate rainfall
45 dB = humming of a refrigerator
40 dB = quiet room

Faint

30 dB = whisper, quiet library

Figure 1.1: Hazardous noise levels.

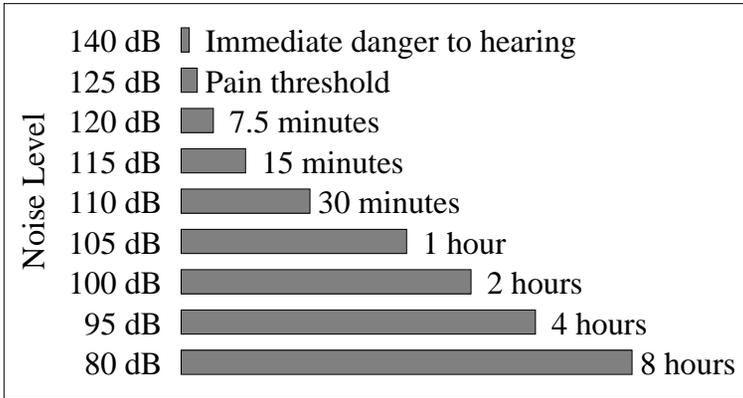


Figure 1.2: Hazardous exposure limits.

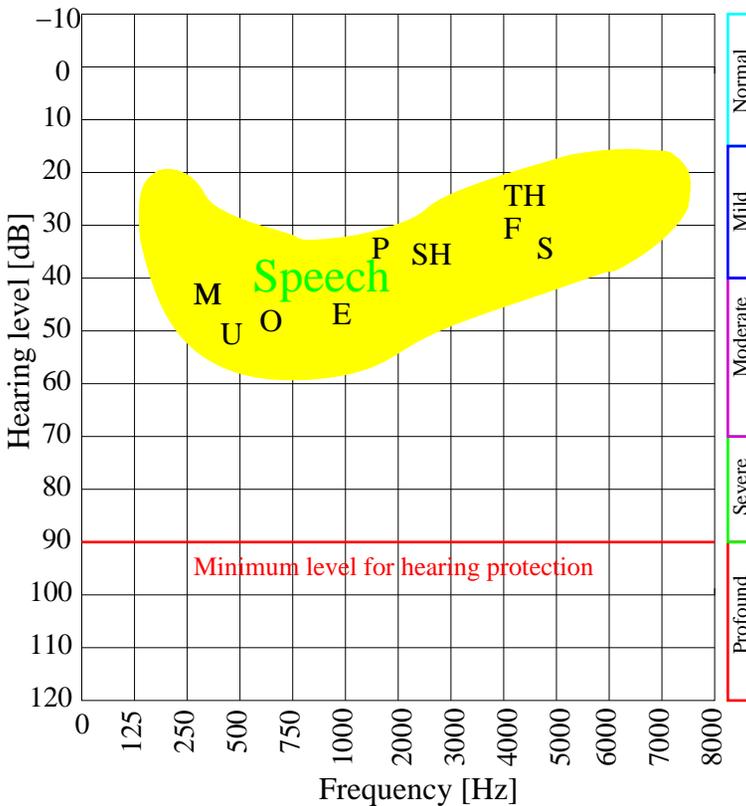


Figure 1.3: Shaded region shows the level and frequency of average speech

Country	Million people
Germany	10.2
France	7.6
United Kingdom	7.5
Italy	7.2
Spain	5.5
Poland	4.7
The Netherlands	2
United States	35

Table 1.2: Hearing loss statistics for different countries [117][195].

United States (2008) [117]. It was further reported in [117] that more than 25 million out the 35 million Americans suffering from hearing loss did not have a hearing aid. There can be many reasons why people with a hearing loss do not wish to use a hearing aid. The work in [115] investigated this issue and some of the reasons are: poor benefit, background noise, negative side effects, price and cost, sound quality, and volume adjustments. The most often heard complaints are [113][115]:

- *"It does not work well in background noise"*
- *"I can't adjust the hearing aids constantly to every noise"*
- *"Volume is OK, but I can't distinguish words"*
- *"Hearing aids amplify other sounds so much that I actually feel pain"*

The work in [116] investigated improvements sought in the United States hearing aid market from a consumer point of view. Basically the consumers were asked to rate different items on a scale between one (not desirable) to five (highly desirable). In Table 1.3 we have extracted the numbers related to benefit and listening experience and sound quality which is related to the DSP part of the hearing aid that is addressed in this dissertation. Other categories like cosmetics, price and cost, batteries, maintenance etc. can be found in [116]. It is clear that speech in noise is the most significant problem for hearing aid users together with the desire for less whistling and buzzing. These problems are directly related to NR and AFC. Making loud sounds less painful and making soft sounds audible is related to DRC. The desire for better sound quality is more objective and depends on the overall output from the different hearing aid algorithms, see Figure 1.8.

	% Not Desirable				% Highly Desirable
Improvement sought	1	2	3	4	5
Speech in noise	0.5	0.5	4.2	13.1	81.7
Less whistling & buzzing	2.4	2.0	10.7	18.6	66.3
Better sound quality	1.0	0.7	9.9	23.0	65.4
Work better - telephone	2.8	2.1	13.5	22.6	59.1
Loud sounds less painfull	2.2	2.8	14.3	22.8	58.0
More soft sounds	2.4	2.1	12.7	25.0	57.8
Speech in quiet	1.2	1.9	15.9	25.7	55.2
Mask tinnitus	5.7	4.6	18.7	18.5	52.5
Work better - cellphone	11.4	5.0	20.2	17.6	45.9
Better sound to music	6.0	4.4	27.3	27.9	34.1

Table 1.3: Improvements sought in the US hearing aid market from a consumer point view [116].

1.1.3 Commercial hearing aids

Commercial hearing aids exists in many different styles and sizes, some of which are listed below [45]:

- **Behind the ear** (BTE) hearing aids fit above and to the rear of the outer ear.
- **Completely in canal** (CIC) hearing aids fit entirely in the ear canal and are nearly invisible
- **In the canal** (ITC) hearing aids fit into the ear canal and fill roughly half of the ear
- **In the ear** (ITE) hearing aids fit completely within the outer ear and fill the entire ear

The choice of hearing aids depends on many factors, e.g., the impaired ear may be too small to be fitted with a CIC or ITC hearing aid and these models are most appropriate for hearing impaired people with a mild to moderate hearing loss. The reason for this is the small size which removes options like volume control or directional microphones. The ITE hearing aid is larger than the CIC and the ITC hearing aid and can be used for hearing impaired people with a mild to severe hearing loss. The large size of the BTE hearing aid makes it suitable for hearing impaired people with a mild to profound hearing loss since BTE hearing aids in

general can provide more amplification compared to smaller hearing aids due to a stronger amplifier and a larger battery [45][109].

Typical design constraints for a commercial hearing aid are the size of the hearing aid, e.g., number of microphones, microphone spacing, battery power, processing complexity, and power restrictions. These constraints can limit the amount of DSP algorithms that can be implemented in a commercial hearing aid. DSP is only part of a larger system which includes many components such as microphone, receiver, earmold, A/D converter, D/A converter, central processing unit, memory etc. This part of the hearing aid is not considered in this dissertation and for a general overview of these components we refer to [45][100][109].

1.1.4 Characterization of signals

Most hearing aids today are designed with different settings depending on whether the input signal is a speech signal or music, which has a great influence on the design of the hearing aids [109]. In this dissertation the input signal is assumed to be a speech signal and therefore the characteristics of a speech signal will be shortly explained. Speech signals have frequency components ranging from 100Hz to 8000Hz and are composed of voiced and unvoiced (noiselike) sounds. Voiced speech is produced by a periodic vibration of the vocal chords and in general contains very little energy above 4kHz. Unvoiced speech is produced by a turbulent airflow and is considered to be broadband. The important frequencies for speech understanding are between 300Hz and 3400Hz, which means that a sampling frequency of 8kHz is sufficient to achieve acceptable speech quality and which is actually the classical telephone bandwidth. Increasing the sampling frequency can increase the speech quality and in this dissertation a sampling frequency of 16kHz is used. Speech signals are considered to be non-stationary both spectrally and temporally and can only be considered stationary over frames of 10-30ms [133]. Besides changes from voiced to unvoiced sounds speech signals also consist of many silence periods. These properties can be exploited using a voice activity detector (VAD) to classify speech and noise-only periods. For a complete description of the speech production model we refer to [41][133].

The knowledge of the speech production model is of course important but the knowledge of the noise sources is crucial. Background noise can be classified as either localized noise (e.g. computer fans) or diffuse noise, i.e., coming from all directions (e.g. wind). Also the background noise can be considered to be stationary (e.g. car noise) or nonstationary (e.g. multi-talker babble). The most difficult scenario arises when the background noise is a speech signal (e.g. competing speakers) since the spectral and temporal structure is similar as the desired speaker. Reverberation (e.g. multipath propagation) and echo/feedback (e.g. acoustic coupling between loudspeaker(s) and microphone(s)) can also be

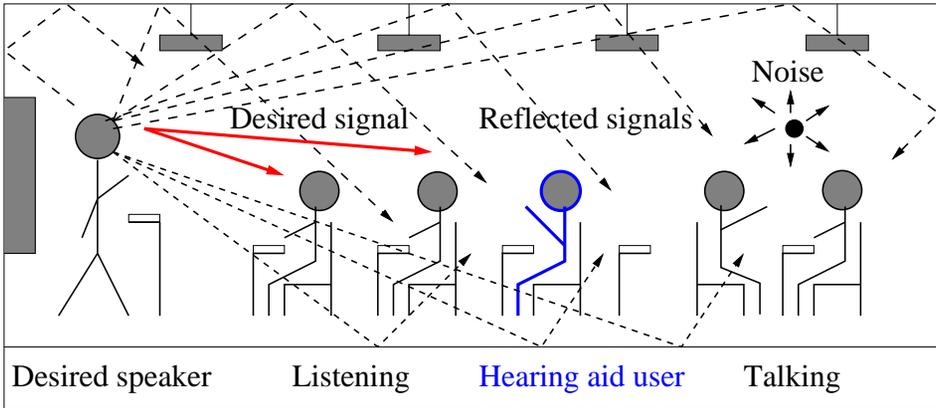


Figure 1.4: Illustration of a hearing aid user in a noisy environment.

considered as noise or at least unwanted signals. Beside the different noise types the noise source(s) can also be classified as either being additive, convolutive, correlated or uncorrelated to the clean speech signal [45][100][109].

1.1.5 Acoustic environment

To communicate effectively in a noisy environment it is important to extract the relevant information from the background noise which is a significant problem for hearing impaired people. In environments with background noise the hearing aid will amplify the noise as well as the desired signal. Due to room reverberation the hearing aid will also amplify signals that are reflected against the walls, ceiling, floor, and other objects in the room. An example with a desired speaker in a classroom is shown in Figure 1.4 where the hearing aid user is facing the desired speaker and the noise is coming from other directions. Two main problems are shown with respect to the hearing aid user, i.e., the desired signal, the reflected signals, and the people talking behind the hearing aid user. The distance between the desired speaker and the hearing aid user and the fact that the desired speaker can move around also plays a crucial role on the speech intelligibility. Furthermore the acoustic environment for hearing aid users can change rapidly, i.e., from being outdoor (e.g. car passing by), in a car (e.g. engine noise) to entering an office (e.g. fan noise, telephone ringing), restaurants (e.g. people talking), home (e.g. television, radio, household appliances), church or a concert hall. All these effects can seriously reduce the speech intelligibility especially for hearing aid users which indeed require a higher SNR, which is also the most desirable improvement sought by hearing aid users, see Table 1.3. Hence there is a strong need for DSP algorithms for hearing aids that can compensate for all these effects.

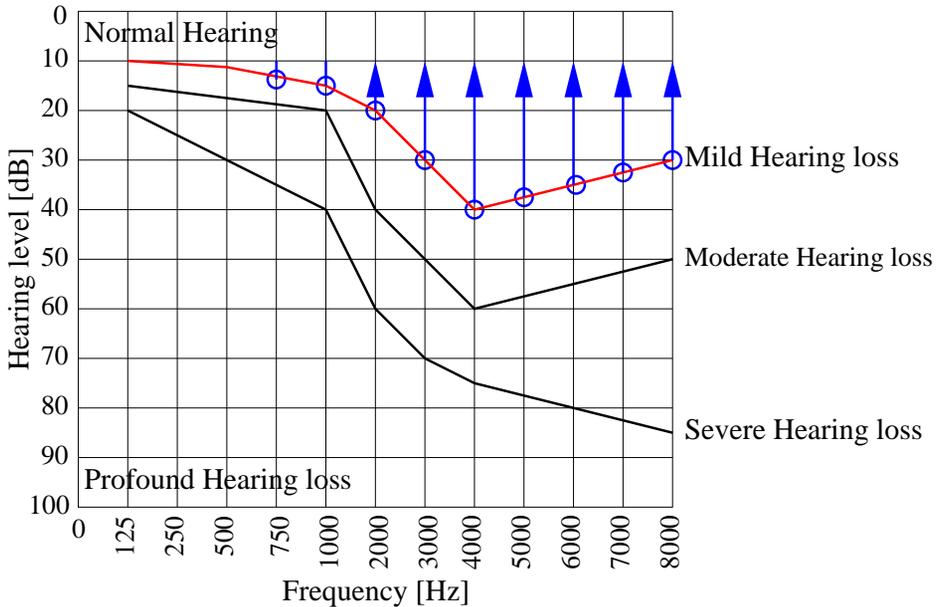


Figure 1.5: Example of an audiogram for mild, moderate, and severe hearing loss

1.1.6 Reduced dynamic range

In general hearing impaired people suffer from a frequency-dependent hearing loss as shown in Figure 1.5. To compensate for this kind of hearing loss the hearing aid should include a frequency-dependent gain such that the high frequencies receive a higher amplification compared to the low frequencies. Typically hearing impaired people also suffer from a reduced dynamic range between the hearing threshold and the uncomfortable level as shown in Figure 1.6. This means that the uncomfortable level for normal hearing and hearing impaired people remains the same but the hearing threshold and the sensitivity to soft sounds are shifted as a result of the hearing loss. It is also clear that a linear amplification in this case will make the soft sounds audible but at the same time loud sounds can become too loud. Therefore the wide dynamic range of speech needs to be reduced by amplifying soft sounds more compared to loud sounds. This problem is also on the list of improvements sought by hearing aid users, see Table 1.3. The rationale behind the DRC is therefore to compensate for the reduced dynamic range in the impaired ear by not only applying a frequency-dependent gain but a level-dependent gain as well.

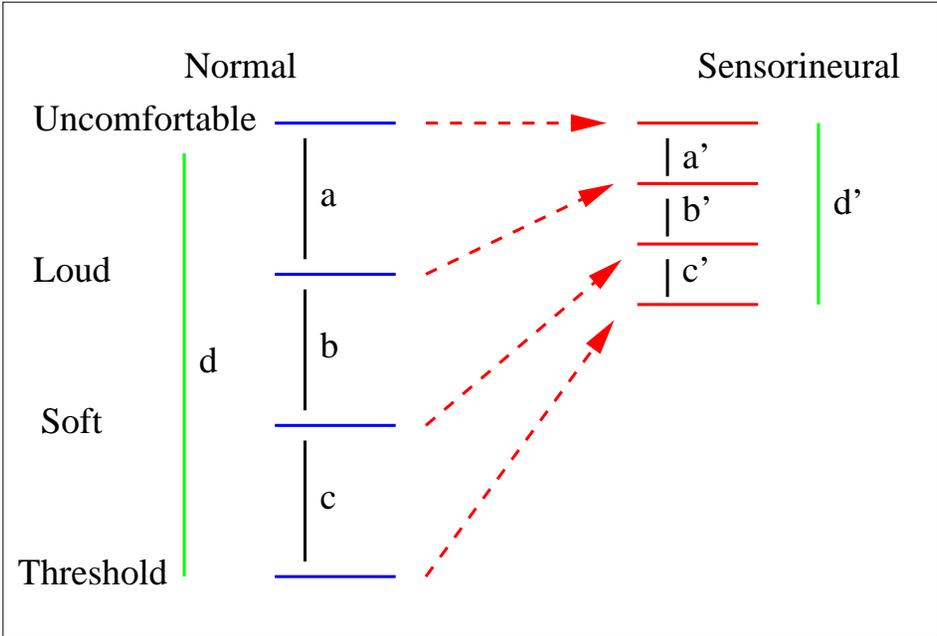


Figure 1.6: Reduced dynamic range compared to dynamic range of normal hearing

1.1.7 Acoustic feedback

Acoustic feedback is a well-known problem in hearing aids, which is caused by the undesired acoustic coupling between the loudspeaker and the microphone as shown in Figure 1.7. This is especially a problem with the use of open fittings and the small distance between the loudspeaker and the microphone. Acoustic feedback produces an annoying howling sound and limits the maximum amplification that can be used in a hearing aid if howling, due to instability, is to be avoided. In many cases this maximum amplification is too small to compensate for the hearing loss, which makes feedback cancellation algorithms an important component in hearing aids [45][100][109]. Actually acoustic feedback is the second highest improvement sought by hearing aid users, see Table 1.3. However acoustic feedback also affects items such as better sound quality and the desire for more soft sounds.

1.1.8 Signal processing challenges

Hearing aid technology is constantly evolving and becoming increasingly more advanced. This is partly due to the ongoing miniaturization of electronics such that more microphones, processing power, and battery power are available in future

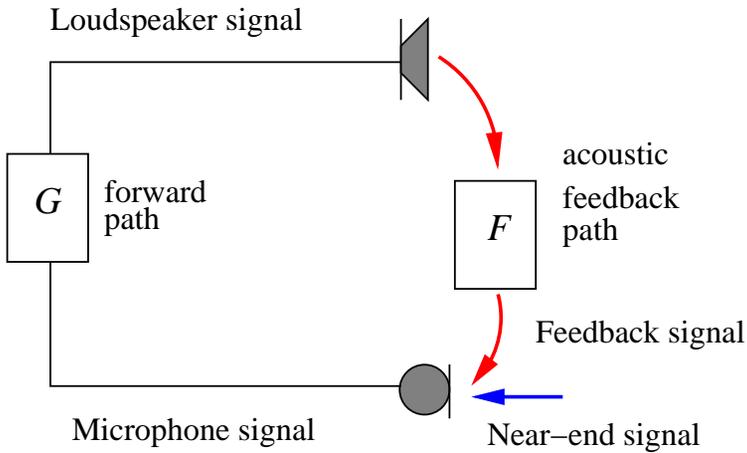


Figure 1.7: Illustration of the acoustic coupling between the loudspeaker and the microphone resulting in acoustic feedback.

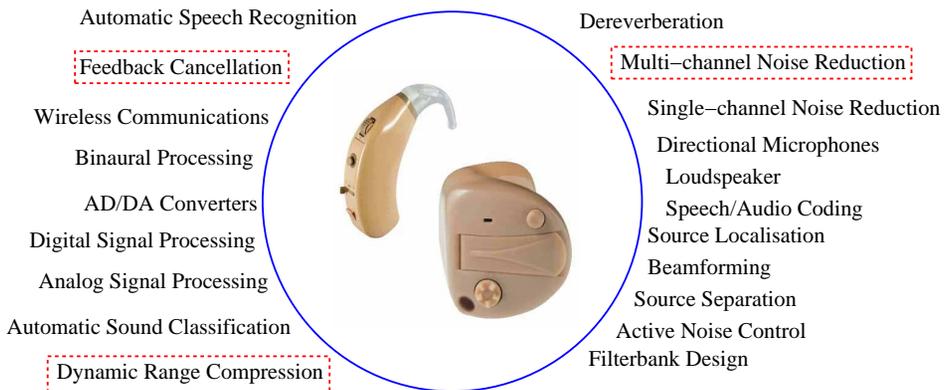


Figure 1.8: Illustration of typical DSP algorithms in hearing aids.

hearing aids. This means that future DSP algorithms can be designed with greater flexibility and potential. Hearing aids include many DSP algorithms that assist hearing impaired people to hear and understand speech. A hearing aid today typically contains several DSP algorithms, some of which are listed in Figure 1.8. Besides the reduced dynamic range (hearing threshold to discomfort) and increased hearing threshold (loss of sensitivity to weak sounds) hearing impaired people also suffer from reduced frequency resolution (separating sounds of different frequencies), reduced temporal resolution (intense sounds may mask following weaker sounds), and reduced spatial cues (spatially separating a desired signal

from noise). These are the effects that make hearing impaired people susceptible to masking produced by background noise which consequently degrades the speech intelligibility. For this purpose we will describe some state-of-the-art NR, DRC, and AFC algorithms that have been proposed in hearing aids which is also the topic of this dissertation. For an extensive overview of the different topics shown in Figure 1.8 we refer the reader to [45][109][100].

1.2 Noise reduction in hearing aids

Background noise tends to decrease the speech intelligibility especially for people suffering from hearing loss [124][51]. The topic of NR in hearing aids is therefore of great importance and many different DSP strategies have been addressed in the past [29]. The goal of NR algorithms is to reduce the background noise and enhance the desired speech signal in complex acoustical environments in order to improve speech intelligibility and/or listening comfort by increasing the SNR without introducing any signal distortion.

NR algorithms can be classified as either fixed filters or adaptive filters [72][237]. The design of fixed filters relies on prior knowledge of the signal, the noise, and the acoustic environment. Adaptive filters on the other hand are more flexible and can adapt the filter characteristics automatically depending on the input signals. The general trade-off in NR is the amount of noise that can be removed versus speech distortion. NR algorithms can also be categorized into single-channel NR and multi-channel NR and here we will provide a broad overview of different NR algorithms.

Voice activity detection

The fundamental component of any NR algorithm is the voice activity detector (VAD). Typically a stationarity assumption is made such that the noise characteristics can be estimated and updated during noise-only periods. The purpose of a VAD is to distinguish between speech dominant segments and noise dominant segments (silence) which can be a challenging problem especially at low input SNR and with non-stationary noise sources. In the past various VAD algorithms have been proposed which all are aimed at improving the robustness, accuracy and reliability. A VAD algorithm can be divided into two separate blocks, i.e., feature extraction and a decision module. The objective of feature extraction is to find discriminative speech features that can be used in the decision module. In this section we will give a brief overview of existing VAD algorithms.

Features used in VAD algorithms have been based on: energy levels, pitch, and zero crossing rates [99][121][179][221], the LPC distance measure [178], cepstral features

[77], adaptive noise modeling of voiced signals [243], the periodicity measure [223], and high order statistics (HOS) [153][183]. The problem with these approaches is the robustness at low input SNR and with non-stationary noise sources since the VAD is typically based on a fixed threshold.

Recent approaches to improve the VAD performance are based on using statistical models with the decision rule derived from the statistical likelihood ratio test (LRT) applied to a set of hypotheses [50][97][98][184]. Other decision rules have been based on: Euclidean distance [71], Itakura-Saito and Kullback-Leibler divergence [185], fuzzy logic [10], and support vector machines (SVM) [186]. Various statistical models have been proposed to improve the VAD performance such as Gaussian, Laplacian, and Gamma models [98]. VADs can also be distinguished based on whether a hard decision (binary) or a soft decision (value between 0 and 1) is used. In [67] a soft VAD has been proposed where the distribution of the clean speech and the noise are assumed to be Laplacian and Gaussian, respectively. The probability of speech being active is then calculated using a maximum likelihood (ML) approach and a hidden Markov model (HMM). In [219] a soft VAD is proposed based on a generalized autoregressive conditional heteroscedasticity (GARCH) filter and a variance gamma distribution (VGD).

It is obvious that an accurate estimate of the noise spectrum is the key to an improved estimate of the original speech. A common noise estimation technique is based on a recursive averaging procedure during periods where the speech is absent and then keeping the noise estimate fixed during periods where speech is present. This approach requires a VAD which in itself suffers from reliability at low input SNR. An interesting approach has therefore been proposed in [35] called improved minima-controlled recursive averaging (IMCRA). Basically the smoothing parameter is now adapted over time and frequency based on the conditional speech presence probability (SPP). The advantage of IMCRA is the continuous update of the noise spectrum and the fact that a binary VAD is not required.

Another common technique to estimate the noise characteristics is known as the minimum statistics algorithm. This approach differs from the traditional VAD methods since the minimum statistics algorithm does not need to distinguish between speech activity and speech pauses. Instead the minimum statistics algorithm is based on the fact that during speech pauses the speech energy is close to zero which means that by tracking the minimum power the noise floor can be estimated [140][141]. In [86] an approach for noise tracking is proposed where the noise PSD can be updated in the presence of both speech and noise. This method is based on the eigenvalue decomposition such that the noisy speech can be decomposed into a signal-plus-noise subspace and a noise-only subspace. This means that the noise statistics can be updated based on the noise-only subspace even when speech is present. Other techniques that can be mentioned are histogram [188] and quantile based noise estimation techniques [212].

1.2.1 Single-channel noise reduction

Single-channel NR algorithms have been widely studied in the past [133] and these approaches can be broadly categorized as parametric or non-parametric techniques. These algorithms are designed to enhance noisy speech signals using a single microphone, i.e., relying only on temporal and spectral differences between the speech signal and the background noise. Single-channel NR is a difficult problem, especially with non-stationary noise sources and low input SNR, since there is no reference microphone available to estimate the noise or to estimate spatial signal information. Since the speech and the noise typically occupy the same frequency bands single-channel NR usually has problems reducing the noise without introducing artifacts and distortion.

Non-parametric noise reduction

Non-parametric NR relies on an estimate of the noise characteristic estimated during noise-only periods, e.g. using a VAD or a minimum statistics algorithm, which then is applied during speech-plus-noise periods to extract the clean speech signal. Many single-channel NR algorithms have been developed during the past years starting from simple spectral subtraction [15][75] which is the most basic and commonly used technique. The idea behind spectral subtraction is to estimate the noise magnitude spectrum from the noisy speech magnitude spectrum, and then subtract it from the noisy speech assuming that the noise is uncorrelated and additive to the speech signal. The analysis and synthesis part of the different NR algorithms is commonly performed using the short-time Fourier transform (STFT) with overlap-add or overlap-save procedure. Spectral subtraction depends on a VAD such that the noise power can be kept fixed during speech segments and updated during noise-only segments which requires a stationarity assumption on the background noise. Over the years many variations of the spectral subtraction algorithm have been proposed, e.g., generalized spectral subtraction [11][40][144], spectral subtraction using over-subtraction and spectral floor [11], nonlinear spectral subtraction [131][132], spectral subtraction with a minimum mean square error (MMSE) short-time spectral amplitude (STSA) estimator [55][56], spectral subtraction based on perceptual properties [25][172][232][240], and subspace based spectral subtraction [59][187]. The aim of all these methods is to compensate for the drawbacks of the traditional spectral subtraction [15], e.g., speech distortion, musical noise and other artifacts.

Another classical single-channel NR algorithm is the Wiener filter [128] which is based on estimating an optimal filter from the noisy speech signal based on minimizing the mean square error (MSE) between the desired signal and the estimated signal. The Wiener filter requires separate estimates of the clean speech and the noise power which can be estimated using a VAD assuming that the speech

and the noise is short-term stationary. One of the drawbacks of the Wiener filter is the requirement of the clean speech power which makes the Wiener filter highly dependent on the VAD.

Single-channel NR based on subspace estimation has also been proposed [59][60][90]. These techniques are based on decomposing the vector space of the noisy speech into a speech-plus-noise subspace and a noise subspace. The noise subspace can then be removed before processing the speech-plus-noise subspace in order to estimate the clean speech signal. The decomposition can be performed applying either the singular value decomposition (SVD) or the Karkunen-Loeve transform (KLT) on the noisy speech signal [146][187]. A perceptually motivated signal subspace based NR has been proposed in [94] such that the masking properties of the human auditory system are taken into account during the NR process.

Other single-channel NR techniques that can be mentioned are based on cost functions such as MMSE estimators [54][56][79], log-MMSE estimators [57], maximum likelihood (ML) [144], and maximum a posteriori (MAP) estimators [135]. In [33][36][152] an optimally modified log-spectral amplitude estimator is proposed based on incorporating the conditional SPP which is estimated for each frequency bin and for each frame by a soft decision approach.

Parametric noise reduction

In parametric NR the noisy speech is modelled as an autoregressive (AR) process embedded in coloured Gaussian noise, which then can be represented in the state-space domain [95][171]. These techniques are performed in two steps, first the speech AR parameters and the noise variances have to be estimated and then the speech signal is estimated by applying either a Wiener filter [78] or a Kalman filter [65][68] using the estimated parameters. In general parametric based NR differs in the choice of model used to parametrize the speech signal and the method used to estimate the model parameters. Commonly used methods to estimate the model parameters are the estimation maximization (EM) algorithm [42] and the Yule-Walker equations [81]. Harmonic models and HMM have also been used in parametric based NR [54][58][128]. Kalman filtering has shown to have certain advantages compared to Wiener filtering since the Kalman filters can take the quasi-stationarity of speech signals into account. This is mainly due to the fact that Kalman filters can be continuously updated [171]. Various modifications and improvements of Kalman filtering based NR can be found in [64][65][112][246].

Although increased SNR and listening comfort have been reported for single-channel NR, limited benefits in terms of speech intelligibility have been reported [150]. Recently an environment specific NR was proposed [91] where the NR algorithm is adjusted based on the listening situation which could be done manually or automatically using sound classification methods. Using the

environment specific NR substantial improvement in terms of speech intelligibility was reported for CI users. Perceptual evaluation with normal hearing and hearing impaired subjects was performed in [136] for various NR algorithms. Here it was shown that single-channel NR algorithms did not significantly affect the speech reception threshold (SRT). However single-channel NR algorithms were significantly preferred over the unprocessed conditions. Since the speech and the noise overlap in time and frequency it is difficult for single-channel NR to perform NR without introducing speech distortion and musical noise. Other artifacts can also be introduced due to the non-linear filtering approach related to single-channel NR. Another factor is low input SNR and highly non-stationary noise sources which typically result in an inaccurate estimation of the noise characteristic.

1.2.2 Multi-channel noise reduction

In the past, hearing aids were typically designed using a single omni-directional microphone but the limited benefit in terms of speech intelligibility for single-channel NR algorithms has motivated the use of multi-microphones [114][150]. Due to the miniaturization of microphones [170] hearing aids can be equipped with two and even three microphones. Typically, the desired speaker and the noise sources are located at different positions and the spatial separation can then be used, i.e., the spectral and spatial differences between the desired signal and the background noise are exploited. This section will summarize some of the well-known multi-channel NR techniques.

Omni-directional and directional microphones

Directional microphones are used to preserve a desired signal coming from a certain direction while reducing noise and interferences from other directions. Directional microphones are preferred when the background noise is present to the side or the rear, or when the desired signal is near to the listener, and the reverberation is low. Omni-directional microphones are equally sensitive to sounds coming from all directions. Omni-directional microphones are preferred when the signal is far from the listener or when the reverberation is high [45][100][109]. In hearing aids the spacing between the microphones is typically small compared to the wavelength of the sound. This can be a problem since the directional response pattern is defined by the microphone spacing and the time delay. The directional response pattern can also be affected by microphone mismatch and the head-shadow effect. A realistic assumption in hearing aids is that the desired signal is located in front of the hearing aid user and the interferences can come from any direction, which can be caused by room reverberation or from noisy environments.

Results obtained in the laboratory often favour directional microphones but in real-world situations results show equal support for directional and omni-directional microphones [9][37][119]. In [8] it was shown that directional microphones provide better speech perception compared to omni-directional microphones in a stationary noise environment. In a moving noise environment an adaptive two microphone mode was preferred compared to a fixed two microphone and an omni-directional microphone mode. In [31] three different noise scenarios were evaluated and the results favoured the adaptive directional microphones compared to omni-directional and fixed directional microphones.

Fixed beamformers

In fixed beamforming, the filter coefficients are fixed for a predefined target position and are hence data-independent. The success of fixed beamformers relies on correct assumptions on the microphone characteristics, array geometry, target position etc. [230]. A fixed beamformer is a spatial filter that focuses on a desired speaker, i.e., speech coming from a predefined target position is passed through without distortion, while reducing the effect of background noise coming from all other directions. The overall output from a beamformer is then formed by summing the output from each filter. Typical fixed beamformers include delay-and-sum beamformers, filter-and-sum beamformers, differential microphone arrays, and superdirective microphone arrays.

Delay-and-sum beamformers basically delay the microphone signals and the single output is then formed by summing the microphone signals. Delays between the microphones are used to steer the mainlobe of the beamformer in a particular direction. The beamwidth of a beamformer is the width of the main lobe which indicates the ability of the beamformer to suppress interferences that are close in azimuth to the desired signal. This beamwidth depends on the array length such that a longer array length results in a narrower main lobe. Filter-and-sum beamformers are based on filter coefficients that are optimized for a certain spatial direction based on a given cost function [168][242]. Filter-and-sum beamformers differ from delay-and-sum beamformers since an independent weight is applied to each microphone signal before the signals are summed together to form the overall output. In differential microphone arrays one of the microphone signals is delayed and the output from the two microphones are subtracted from each other [52]. Superdirective beamformers [106] are designed to maximize the directivity pattern in a desired direction, while suppressing noise coming from all other directions.

Fixed beamformers are typically very sensitive to microphone mismatch, missteer, array geometry, and speaker position especially when applied in small-sized arrays such as in hearing aids [211]. Robustness against these errors can be achieved by calibrating each hearing aid which unfortunately can be time consuming and

expensive [211][218].

Adaptive beamformers

Adaptive beamformers make use of data-dependent filter coefficients and adapt to non-stationary signals and various acoustic environments [12][82][173][238]. In general adaptive beamformers have a better NR performance compared to fixed beamformers especially when the number of interferers is smaller than the number of microphones, when no diffuse noise and little reverberation is present. The idea is to steer the main lobe towards the desired signal and adapt the nulls in the direction of the interferences. In most hearing aid scenarios there are more sources than microphones available which means that there are not enough nulls to be steered toward the interferences. Instead the overall power of the interferences can be minimized by reducing the sidelobes without taking the nulls direction into account. The performance of adaptive beamformers can be improved by increasing the size of the array or the number of microphones. The challenge of adaptive beamformers is robustness against reverberation and scenarios where there are more sources than microphones. In theory, an array of M -microphones can generate $M-1$ nulls in the directional response pattern, i.e., $M-1$ interferences can be suppressed. The length of the filter also has an influence on the performance and therefore needs to be carefully chosen as well. A long filter can model the direct sound and the reflections and should therefore perform better than a short filter. However long filters adapt slowly and may not respond fast enough to changes in the acoustic environment.

Linearly constrained minimum variance (LCMV) beamforming is a well-known technique that is designed to minimize the energy of the output signal with a constraint on a certain direction such that the target speech is preserved [93]. An alternative implementation of the LCMV beamformer is known as the generalized sidelobe canceler (GSC) [74] where the constrained optimization problem is reformulated as an unconstrained optimization problem. The GSC is based on a fixed spatial pre-processor consisting of a fixed beamformer and a blocking matrix together with an adaptive noise canceller (ANC). The ANC then minimizes the output power while the blocking matrix is designed to avoid the speech leaking into the noise references [17][46][81][180][224]. Obviously the performance of the GSC will be affected if the speech signal is leaked into the noise references causing speech distortion, which can happen due to reverberation, microphone mismatch, misteering etc. Many variations of the GSC have been proposed in order to make the GSC robust against these errors [18][32][34][66][88][235]. Adaptive beamformers generally provide better NR performance compared to fixed beamformers since the filter coefficients can be adapted to the changing acoustic environments. However adaptive beamformers are very sensitive to modelling or adaptation errors which can cause speech distortion. For this reason the fixed beamformer is sometimes

desired especially if robustness and low complexity is preferred compared to NR performance. Minimum variance distortionless response (MVDR) is a special case of the LCMV beamformer [182], where the filter weights are designed to minimize the variance of the output signal subject to a unity constraint in the target direction.

In [204] it was shown that a two microphone adaptive beamformer significantly improved the SRT compared to a standard hardware directional microphone. The evaluation was carried out on a Nucleus Freedom cochlear implant (CI) system based on five adult CI users. Perceptual evaluation for a dual-microphone hearing aid also favoured the adaptive beamformer compared to a fixed software directional microphone which is state-of-the-art in most modern commercial hearing aids [137].

Multi-channel Wiener filter

Another multi-channel NR algorithm is the multi-channel Wiener filter (MWF) [62][139][169]. MWF based NR provides an MMSE estimate of the speech component in one of the received microphone signals. The MWF is uniquely based on second order-statistics of the speech and the noise signals and makes no a priori assumptions about the signal model. This is particularly beneficial in terms of robustness, especially when working with small-sized arrays such as in hearing aids. MWF algorithms exploit both spectral and spatial differences between the speech and the noise sources. The MWF has been extended to the speech distortion weighted MWF (SDW-MWF) such that the MMSE optimization criterion now allows for a tradeoff between speech distortion and noise reduction [207][49]. The performance of the MWF has been evaluated in [48][205] which showed that the MWF outperformed the GSC in adverse listening environments. Perceptual evaluation with normal hearing and hearing impaired subjects was performed in [136] for various NR algorithms. Here it was shown that the MWF was the only algorithm that provided a significant SRT improvement compared to four other NR algorithms. The problem is the high computational complexity of the MWF which has limited the usage of the MWF in hearing aids. However, work has been done to reduce the complexity of the MWF by exploiting low cost subband and stochastic gradient implementations of the MWF [206][207].

In this thesis, we will focus on the MWF with unknown reference which means that no a priori information and calibration is needed [48][189]. This property is highly desirable from a robustness point of view.

1.3 Dynamic range compression in hearing aids

DRC is a basic component in digital hearing aids and the use of DRC has increased over the years [150][201]. The role of the DRC is to estimate a desirable gain to map the wide dynamic range of an input audio (e.g. speech) signal into the reduced dynamic range of a hearing impaired listener. DRC is a DSP strategy that makes speech audible over a wide range of sound levels and reduces the dynamic range of speech signals. Basically, a DRC is an automatic gain control, where the gain is automatically adjusted based on the intensity level of the input signal. Segments with a high intensity level (loud sounds) are amplified less compared to segments with a low intensity level (soft sounds), since a comfortable listening level for loud sounds makes soft sounds inaudible. In this way it is also guaranteed that loud sounds are not becoming uncomfortably loud, so beside audibility a DRC is also designed to avoid discomfort, distortion, and damage. A hearing aid with DRC has a gain that changes over time and frequency depending on the intensity level and is therefore considered to be a non-linear DSP algorithm. Changing the gain in each frequency band will modify the speech spectrum and a rapid change of the gain could also lead to audible processing artifacts. The design and perceptual benefit of DRC will be shortly reviewed in the following sections.

1.3.1 Design of DRC algorithms

A DRC algorithm is typically performed in the following operations. First the input signal is divided into a number of frequency bands and then an envelope detector estimates the level of each frequency band. In the last step the level of each frequency band is inserted in the compression characteristic which estimates the desired gain which is applied to the input signal.

Most digital hearing aids today use a multi-band DRC approach which can be implemented using either filter banks or a fast Fourier transform (FFT). The problem with an FFT approach is the constant frequency resolution whereas typically it is desirable to have a frequency resolution that can match the resolution of the human auditory system. Another problem is the optimal number of frequency bands that should be used in a DRC which is still unclear. General trade-offs in the design of the DRC involves complexity, frequency resolution, time delay, and quantization noise [108]. For any given application, increased frequency resolution comes at the price of increased delay. In [109][110] a DRC using digital frequency warping has been proposed with two main features such as a frequency analysis that is better matched to the human auditory system, i.e., close to the auditory Bark scale and a reduced group delay compared to traditional designs. In [87] a multi-band DRC has been proposed based on instantaneous compression performed on each frequency band using a gammatone filterbank.

In the envelope detection the setting of the attack and the release time have been widely discussed, i.e., whether a fast or a slow time constant should be used. Using a fast time constant can make all segments of the speech audible [231] whereas a slow time constant can preserve the speech envelope [175]. The attack and release time specify how fast the gain is changed according to changes in the input signal. The attack time is defined as the time taken for the compressor to react to an increase in input signal level. The release time is the time taken for the compressor to react to a decrease in input level. Typically a fast attack time is used such that an increase in signal level can be quickly detected in order to avoid overamplification of loud sounds. The release time is usually slower to avoid audible fluctuations and distortion and it is also assumed that insufficient amplification is not as damaging compared to overamplification. However a problem could arise in terms of audibility, if the release time is too slow, such that the soft sounds are amplified with a gain that was previously appropriate for loud sounds.

The attack and release time only specify how fast the DRC algorithm should react to an increase or a decrease in signal level but it does not define how much the gain should be increased or decreased which is defined by a predefined compression characteristic. A compression characteristic of a DRC algorithm is typically defined by the compression ratio (CR), the compression threshold (CT), and a desired gain which leads to a input-output characteristic that shows the output sound pressure level (SPL) as a function of the input SPL. The CT is defined in dB and corresponds to the point where the DRC becomes active, i.e., where the gain is reduced. The CR determines the degree of compression. A CR of 2 (i.e. 2:1) means that for every 2dB increase in the input signal, the output signal increases by 1dB.

In the past extensive works have analyzed the challenges and difficulties in the settings of the DRC parameters such as the CR, CT, attack and release time, and the number of compression bands [21][53][120][149][154][175][215][231]. Another design criterion of the DRC has also been developed where the goal is to match the estimated loudness in the impaired ear to that of a normal ear [103][118]. The general design of different DRC algorithms can be found in [13][87][110][193][127].

1.3.2 Perceptual benefits from DRC

The benefits of using DRC in hearing aids have been reported with different results and here we will review some of the conclusions made from perceptual experiments with DRC. In [154] it was shown that listeners preferred a CR less than 2:1 for all conditions. However in the presence of multi-talker babble noise, subjects with a residual dynamic range greater than 30dB preferred a CR of 1:1 which corresponds to a linear amplification. Subjects with a residual dynamic range of

less than 30dB preferred a CR of 1.5:1 or 2:1. The same trend was shown in [92] where an improved speech intelligibility was reported when using DRC with speech in quiet but in the presence of noise a linear amplification was preferred. This shows that the performance of the DRC in the presence of background noise is different. In [202] it was reported that the DRC degrades the SNR and the effective CR is lower for speech in noise.

The perceptual benefits with regard to the number of frequency bands needed in DRC has also been widely discussed [149][236]. In [111] it was shown that increasing the number of frequency bands from 1 to 4 did not have any significant improvement on speech intelligibility and most subjects did not have any preference for the number of frequency bands. In [151][215] the perceptual benefits of fast-acting multi-band DRC was evaluated and compared to a linear amplification. The first conclusion made was that the number of frequency bands did not have a significant effect on the SRT. However it was shown that a fast-acting multi-band DRC had a significant benefit compared to a linear amplification. In [225] a DRC algorithm with 1, 4, and 16 frequency bands was evaluated, which showed no significant difference, however the pleasantness for speech and music were both highest with a single frequency band. In [236] it was concluded that multi-band DRC unnecessarily attenuates important information regarding the shape of the short-time speech spectrum, i.e., the height of spectral peaks are reduced and the floor of the spectral valleys are increased which flattens the short-time speech spectrum, resulting in poor speech perception. In [19] it was suggested that multi-band DRC increased the overall gain compared to linear amplification. This resulted in greater audibility and intelligibility for low level sounds. However at high input levels the multi-band DRC degraded the intelligibility compared to linear amplification. The analysis suggested that the multi-band DRC caused spectral distortion due to independent compression in each frequency band which was the reason for the degraded intelligibility. However the work in [87] investigated the benefit of having slow-acting versus fast-acting multi-band DRC and it was concluded that fast-acting DRC indeed introduces some distortion but this did not have any negative impact on speech intelligibility. A different DRC approach that does not use a compression characteristic with a CR and a CT has been proposed in [13][14] called adaptive dynamic range optimization (ADRO) amplification. ADRO is designed with 64 frequency bands and uses statistical analysis of the signal to optimize the output dynamic range in each frequency band independently. Here it was reported that, speech perception scores showed improved audibility for sounds in many narrow frequency bands while still maintaining improved comfort and sound quality.

In general it can be concluded that DRC is preferred over linear amplification. However the optimal design of a DRC is still an open question. In [45][245] it is suggested that multi-band DRC would mostly benefit people with a steeply sloping hearing loss and that the full advantage of multi-band DRC may not be achieved

until the hearing aid user has sufficient experience using a multi-band DRC hearing aid. It is also mentioned that a multi-band DRC would probably work best at very low and at very high input levels whereas most work has evaluated the multi-band DRC at mid-range levels.

Despite the disagreement of how to design DRC algorithms and the perceptual benefits the target of DRC is clear, i.e., make soft sounds audible by increasing the gain while making loud sounds comfortable by reducing the gain. For sound levels between soft and loud typically a linear amplification is applied. In this dissertation we will focus on the scenario where the DRC operates in the presence of background noise. Furthermore the aim is to analyze the undesired interaction effect between NR and DRC.

1.4 Feedback cancellation in hearing aids

Modern hearing aids are becoming smaller and smaller these days which is desired from an aesthetic point of view. Unfortunately this poses other problems such as acoustic feedback, i.e., the acoustic coupling between the loudspeaker and the microphone(s). The acoustic feedback problem is caused when amplified sound leaks out and gets picked up by the microphone in the hearing aid creating a closed signal loop. The acoustic feedback problem stems from the vent in the earmold of the hearing aid which is becoming larger with the desire for open fittings. The vent is used to reduce the occlusion effect which refers to the hearing aid users' unnatural perception of their own voice [45][100][102][109]. Increasing the size of the vent makes the acoustic feedback problem worse. Acoustic feedback limits the maximum gain that can be used in the hearing aid. Furthermore, acoustic feedback is audible as a continuous high-frequency tone emanating from the hearing aid.

There exist two distinct techniques to tackle the acoustic feedback problem, i.e., feedforward suppression and feedback cancellation both with a common goal, i.e., to maximize the hearing gain while minimizing any processing artifacts such as distortion, ringing, howling etc. The acoustic feedback problem can be solved by either completely removing the acoustic coupling or partially by removing the howling artifacts from the loudspeaker signal. The most widely studied techniques are phase-modulating feedback control (PFC), notch-filter based howling suppression (NHS), and adaptive feedback cancellation (AFC). The PFC attempts to smooth the system loop gain whereas NHS aims to actually suppress the howling. The target of AFC is to completely remove the acoustic coupling. Here we will give a short overview of the two concepts and describe the advantages and disadvantages of these techniques for reducing acoustic feedback.

1.4.1 Feedforward suppression

In feedforward suppression techniques, the acoustic feedback is reduced in the forward path (i.e., in the closed signal loop) such that the hearing aid is stable in conjunction with the acoustic feedback path. A simple method for reducing feedback is with the use of notch filters such that the gain is reduced in a narrow frequency region (around the critical frequencies) whenever acoustic feedback is expected to occur [105][142]. However there exist some disadvantages, e.g., reducing the gain at certain frequencies could potentially reduce the speech quality or even speech intelligibility if the notch filters are wrongly placed. In general the benefits of notch filters have been limited since these techniques operate in the forward path which can compromise the frequency response and the speech quality of the hearing aid. Furthermore, these techniques usually have a reactive nature which means that howling or ringing can usually be perceived before the detection algorithm is activated. Other feedforward suppression techniques have been proposed based on equalizing the phase of the open-loop response [234] or by using techniques such as frequency shifting and phase modulation [20][101][166][194].

1.4.2 Feedback cancellation

Recently attention has been focused on feedback cancellation algorithms where the target is to model the acoustic feedback path which is used to predict the feedback signal in the microphone signal, i.e., part of the loudspeaker signal that is leaked into the microphone signal. This predicted feedback signal is then subtracted from the microphone signal and the feedback compensated signal should correspond to the desired signal if the estimated acoustic feedback path is accurately estimated. Ideally this kind of approach will preserve the desired signal at the input to the forward path. The acoustic feedback path is typically modelled with an adaptive finite impulse response (FIR) filter, since the acoustic feedback path can change rapidly depending on the acoustic environment. The acoustic feedback path includes a slowly varying part, i.e., the microphone, the amplifier, and the receiver and a rapidly varying part, i.e., the vent acoustic, the earmold (acoustics leaks), and external acoustics all of which can have a large effect on the acoustic feedback path [84][181]. Examples of external acoustic factors can be scenarios where a telephone is used [213] or if the hearing aid user is putting a hat on. Room reverberation can also affect the performance of feedback cancellation in hearing aids [104]. It is therefore preferable to use adaptive feedback cancellation (AFC) algorithms. The accuracy of the estimated acoustic feedback path is of course crucial since mismatch between the estimated acoustic feedback path and the true acoustic feedback path can lead to instability and distortion.

AFC algorithms exist in two different setups, i.e., with continuous adaptation or with non-continuous adaptation. The latter is considered to be a reactive approach since the filter coefficients are only adapted when instability is detected. This kind of systems actually allows the hearing aid to become unstable before the AFC algorithm is activated which in general is not desirable. Continuous adaptation AFC is considered to be proactive, i.e., the critical frequencies are identified before howling and ringing occurs. The acoustic feedback path can be estimated and updated continuously by using standard adaptive filtering techniques such as least mean squares (LMS) or recursive least squares (RLS) to adapt the filter coefficients. However the continuous adaptation poses a problem that is highly non-trivial due to the presence of the closed signal loop which introduces a correlation between the loudspeaker and the near-end signal which is a major problem especially if the near-end signal is spectrally colored such as for speech and music signals.

Therefore, applying standard adaptive filtering to the AFC problem in hearing aids results in a biased estimate of the acoustic feedback path [85][197]. This can lead to part of the desired signal being partially cancelled or at least distorted and of course the risk of ringing and howling since the acoustic feedback path is not accurately estimated. For this reason decorrelation procedures are usually included in the AFC algorithm. The correlation problem is worst when the input signal is tonal, e.g., speech and music signals which causes large changes in the estimated acoustic feedback path. The problem is that standard adaptive filters adapt to cancel the tonal components rather than modelling the acoustic feedback path. It is clear that the bias and the correlation problem are crucial to achieve good AFC performance which is the main topic of the next section.

1.4.3 Bias problem and decorrelation

In the past, many different solutions have been proposed to reduce the bias or the correlation between the loudspeaker and the near-end signal. Decorrelation can either be performed in the closed signal loop or in the adaptive filtering circuit.

Decorrelation in the closed signal loop can be achieved by injecting a white noise signal in the forward path of the hearing aid. The problem with injecting a noise signal is that it affects the overall sound quality. For this reason attempts have been made to shape the noise based on psychoacoustic models. However in this case the shaped noise had to be amplified to a level that was found to be more disturbing than the white noise injection [229]. Decorrelation can also be achieved by including a nonlinear or a time-varying operation in the forward path of the hearing aid. Inserting a processing delay can also reduce the correlation between the loudspeaker signal and the near-end signal which is assumed to be short-term correlated. Although decorrelation in the closed signal loop can be effective, the sound quality can be significantly affected. Therefore the attention

has been directed towards decorrelation in the adaptive filtering circuit, such that the closed loop signals remain unaffected. In the remainder of this section we will focus on decorrelation in the adaptive filtering circuit.

Filtered-X LMS algorithms can be used to reduce the sensitivity of the system towards tonal inputs. In [89] an AFC algorithm is proposed based on a filtered-X LMS approach which was compared to a traditional LMS algorithm. The filtered-X LMS approach showed better convergence behavior and improved AFC performance. Another filtered-X LMS approach is proposed in [83] where the adaptation is based on closed loop identification [63][196]. A bias reduction approach has been proposed in [16] based on inserting all-pass filters in the forward path of the hearing aid. These filters are time-varying with constant magnitude and varying phase. This idea is motivated by the fact that the human ear is not sensitive to moderate phase perturbations. Incorporating prior knowledge of the acoustic feedback path has also been used in AFC. This leads to a constrained adaptation approach that guarantees that the estimated filter coefficients do not deviate too much from a reference [107]. Bandlimited adaptation has also been used such that feedback cancellation is restricted to frequency bands that are unstable [89].

A well-known approach is to use decorrelation prefilters [83][85] that are designed to whiten the desired signal component in the microphone and the loudspeaker signal. The challenge of this approach is the joint identification of the decorrelation prefilters and the acoustic feedback path. For this purpose it has been proposed to use a prediction-error-method (PEM)-based approach [190][209][210]. The idea behind PEM-based AFC is based on using a model for the near-end signal, e.g., a linear prediction (LP) model, where the inverse of the LP model is used as a time-varying FIR decorrelation prefilter. A LP model is widely used in speech applications since the speech is well modelled with an LP model. Once the near-end signal model is estimated the inverse of this model is applied to the loudspeaker and the microphone signals before feeding these signals to the adaptive filtering algorithm. This method has been shown to improve the filter convergence which is of great importance since AFC algorithms need to track changes in the acoustic feedback path fast and accurately. In [208] four commercial hearing aids were evaluated and compared to the PEM-based AFC [209][210]. It was shown that the PEM-based AFC together with the Starkey Destiny 400 hearing aids offered a high added stable gain (ASG) compared to the Phonak Savia Art 411 dSZ and the Siemens Centra HP hearing aids. However the PEM-based AFC and the Starkey Destiny 400 hearing aid were more sensitive towards tonal input signals. This is mainly due to the near-end signal model used which in this case was based on a LP model.

In this dissertation we will only focus on the PEM-based AFC with emphasis on improving the near-end signal model and the estimation of the parameters used in the near-end signal model since this method has the greatest potential to reduce

acoustic feedback without introducing distortion.

1.5 Outline of the thesis and main contributions

This section starts by summarizing the main research objectives of this dissertation. In addition, an overview of each chapter is given followed by the references to the publications that have been produced in the frame of this research work. References that directly motivate the research behind this dissertation are also given.

1.5.1 Main research objectives

The main research objectives of this dissertation are divided into three parts:

1. The first part addresses the NR problem in hearing aids, i.e., extensions of the SDW-MWF based NR algorithm. First the target is to improve the trade-off between NR and speech distortion by introducing different weighting factors. The second target is to improve the robustness and the tracking related to how the correlation matrices are estimated and updated. Another contribution to the NR problem is based on a Capon beamformer designed for small arrays. The aim here is to design a low computational complexity beamformer in which the steering vectors are estimated in a robust way.
2. The second part addresses the problem of having NR and DRC combined in hearing aids. The work is focussed on a combined SDW-MWF based NR and DRC. First the aim is to analyze the undesired interaction effects when background noise is present in the DRC. Secondly the goal is to compensate for any negative effects that can counteract the NR performance.
3. The third part addresses the acoustic feedback problem in hearing aids, i.e., the developments on the PEM-based AFC algorithm. The aim here is to improve decorrelation prefilters by introducing a novel near-end signal model based on exploiting a harmonic sinusoidal near-end signal model that incorporates various features extracted from a typical speech signal.

1.5.2 Chapter by chapter outline and contributions

In **Chapter 2** the MWF based NR is reviewed together with the problem statement and the motivation for further research. First two formulations of the SDW-MWF are given, namely the standard SDW-MWF and the rank-1

SDW-MWF. Secondly problems related to the estimation and the update of the correlation matrices are discussed such as the robustness, when estimating the clean speech correlation matrix, and the problems of tracking spatial and especially spectral signal characteristics. A third problem addressed here is the use of a weighting factor that is fixed for each frequency and for each frame to make the trade-off between NR and speech distortion. Through simulations it is shown how the reverberation, low input SNR, different spatial angles, and the fixed weighting factor negatively affects the SDW-MWF and the rank-1 SDW-MWF performance. The evaluation is based on objective quality measures such as intelligibility weighted SNR and signal distortion measures.

The references for the MWF based NR that serves as a baseline for the research and development on the MWF are given in [49][38][207].

Chapter 3 focuses on the problem of having a fixed weighting factor to make the trade-off between NR and speech distortion as explained in **Chapter 2**. For this particular reason it is proposed to incorporate the conditional SPP, which is estimated for each frequency and for each frame, in the SDW-MWF such that the speech dominant segments and the noise dominant segments are weighted differently. In the same process another solution is proposed that offers a flexible weighting based on the conditional SPP combined with a per frame decision. Making the weighting factor change for each frequency and for each frame improves the spectral tracking of the speech which is very important since speech signals are typically highly nonstationary. Through simulations it is shown how the proposed weighting factor can improve the SNR improvement while keeping the signal distortion low.

The main findings of this chapter are published in [158][160][161].

Chapter 4 focuses on the problem of estimating and updating the correlation matrices in a robust manner. For this purpose a novel method to estimate and update the correlation matrices is presented. The first step of the proposed method is based on using prior knowledge of the correlation matrices such that the estimated correlation matrices have a certain structure which guarantees that the corresponding filter is valid. The second step, is based on combining the use of the prior knowledge of the correlation matrices with a continuous update approach based on the conditional SPP to improve the spectral as well as the spatial tracking. Using the conditional SPP in the update of the correlation matrices certain frequencies can be weighted higher such that frequencies dominated by speech and those dominated by noise are weighted differently. By combining the proposed weighting factor in **Chapter 3** with the proposed method to estimate and update the correlation matrices, a novel MWF based NR is proposed compared to the MWF based NR introduced in **Chapter 2**. Through simulations it is shown how the proposed MWF based NR outperforms the standard MWF based NR both in terms of SNR improvement and signal distortion.

A publication related to this chapter is in preparation [159].

Chapter 5 presents a different multi-channel NR technique based on the well-known Capon beamformer. The challenge here is not necessarily to estimate the correlation matrices as in **Chapter 4**. The target here is to estimate the steering vector in a robust way. Therefore a robust Capon beamformer is proposed based on using an uncertainty principle such that the steering vector can be properly constrained and estimated. The proposed approach is focussed on low computational complexity and small arrays which are important factors in hearing aids. The proposed robust Capon beamformer is compared to the standard Capon beamformer both in terms of SNR improvement, signal distortion and computational complexity. Through simulations it is shown how the robust Capon beamformer is able to outperform the standard Capon beamformer with a very low increase in computational complexity.

The main findings of this chapter can be found in [156].

Chapter 6 explains the design of the multi-band DRC used in this work together with the various parameters included in a typical DRC algorithm. However the focus here is not the design of DRC algorithms but rather the problem of having the DRC operate in the presence of background noise. First the DRC is analyzed when background noise is present and then it is shown how the DRC gain should be chosen if the SNR should be preserved. It is also shown that incorporating knowledge of speech dominant segments and noise dominant segments can help avoiding further SNR degradation. Beside the problem of SNR degradation amplifying noise dominant segments could also lead to masking of speech dominant segments especially if the speech dominant segments due to a high intensity level receives less amplification compared to dominant segments. Through simulations it is shown how the presence of background noise can negatively affect the DRC and the SNR.

A publication related to this chapter is in preparation [157].

Chapter 7 presents a combined SDW-MWF based NR and DRC. Based on the analysis in **Chapter 6** a dual-DRC approach is proposed that incorporates the conditional SPP, introduced in **Chapter 3**, in the traditional DRC such that a different gain can be applied to speech dominant segments and to noise dominant segments in order to avoid any SNR degradation. The purpose of the DRC is to estimate a gain based solely on the intensity level without considering speech dominant segments or noise dominant segments. A problem arises if a larger gain is applied to the noise dominant segments which has previously been reduced by the NR compared to the speech dominant segments which should be preserved by the NR. First a standard SDW-MWF based NR as introduced, in Chapter 2, and the DRC is concatenated and the undesired interaction effects are discussed. Then the modified SDW-MWF based NR as introduced in, Chapter 3, and dual-DRC is combined together with the conditional SPP. Through experimental simulations

it shown that the DRC indeed counteracts the MWF based NR and that the proposed combined MWF based NR and dual-DRC is able to compensate for a certain amount of the SNR degradation.

The main findings of this chapter are in [155][162].

Chapter 8 reviews the acoustic feedback problem which is related to the need for high amplification in hearing aids as explained in Chapter 6. It is shown that in PEM-based AFC the near-end signal can be modelled using LP. However the decorrelation is not sufficient with a single near-end signal model based on LP which is the reason that a cascaded near-end signal model has been proposed using LP combined with a pole-zero LP (PZLP) model. In this way, the LP models the noise component and the PZLP models the tonal components. Through simulations it is shown that the PEM-based AFC using a single near-end signal model fails to provide sufficient stability compared to the PEM-based AFC using a cascaded near-end signal model. The evaluation is based on objective measures such as the maximum stable gain (MSG) and the filter misadjustment.

The references for the PEM-based AFC that serves as a baseline for the research and development on the AFC are given in [209][210][228].

Chapter 9 presents a novel PEM-based AFC using a harmonic sinusoidal near-end signal model cascaded with an LP model. To improve the near-end signal model, compared to the models presented in **Chapter 8**, information such as pitch, amplitude, and the number of harmonics are built into the near-end signal model. The detection of voiced-unvoiced frames is also included in the cascaded near-end signal model such that in unvoiced frames only a single near-end signal model is used due to the high correlation of voiced frames. Furthermore, a single near-end signal model using LP also has a lower computational complexity compared to using a cascaded near-end signal model. Through simulations it is shown that an accurate modelling of the near-end signal results in improved PEM-based AFC performance in terms of MSG and filter misadjustment.

The main findings of this chapter are in [163][164][165].

Finally **Chapter 10** summarizes the overall conclusions of the research presented in this dissertation. Suggestions for further research are given here as well.

Chapter 2

Speech distortion weighted multi-channel Wiener filter (SDW-MWF _{μ})

In this chapter we will introduce the multi-channel Wiener filter (MWF) and establish a baseline for the research and development related to the MWF based NR algorithm. Over the years many modifications and formulations have been introduced in the MWF either to improve the performance or the robustness. The benefit of using an MWF based NR compared to, e.g. an GSC or an LCMV beamformer, is the reduced sensitivity against signal model errors such as microphone mismatch. For multi-channel NR algorithms like the LCMV or the GSC beamformer a priori assumptions regarding the desired signal model are required, e.g., location of the desired speaker, calibrated microphones, low reverberation etc. The performance of the MWF has been evaluated in [48][205] which showed that the MWF outperformed the GSC in adverse listening environments. Furthermore perceptual evaluation with normal hearing and hearing impaired subjects performed in [136] indicated a significant SRT improvement for the MWF algorithm.

In this work we will focus on two main extensions of the MWF which have shown to have certain interesting features but at the same time we will show that there is still room for improvements. The first extension of the MWF is based on introducing a weighting factor that allows for a trade-off between NR and speech distortion referred to as the speech distortion weighted MWF (SDW-MWF). This is an interesting approach however it is not clear how to actually select this weighting factor and therefore in the past this weighting factor has simply been a scalar value

that is fixed for all frequencies and for all frames. The second extension is based on a rank-1 formulation of the SDW-MWF which has been shown to be more robust against estimation errors in the correlation matrices [38]. The interesting feature of the rank-1 formulation is that the SDW-MWF is now decomposed into a spatial filter and a single-channel postfilter. However it still remains an open problem if the single-channel postfilter is optimal in terms of spectral tracking since it is based on correlation matrices that are adapted slowly over time.

Section 2.1 introduces the multi-channel NR problem with the motivation of using an MWF based NR compared to other multi-channel NR techniques. Furthermore the notation and the relevant definitions are given as well.

Section 2.2 defines the MMSE criteria of the MWF together with the derivation leading to the MWF solution. The estimation and update procedure of the correlation matrices are also defined.

Section 2.3 generalizes the MMSE criterion of the MWF to allow for a trade-off between NR and speech distortion by introducing a weighting factor μ .

Section 2.4 extends the SDW-MWF to a rank-1 formulation referred to as rank-1 SDW-MWF. It is shown how the derivation leads to the SDW-MWF being decomposed into a spatial filter and a single-channel postfilter.

Section 2.5 explains the open problems and challenges remaining in the SDW-MWF and the motivation for further research is given.

Section 2.6 presents the experimental set-up together with the objective quality measures. Then the SDW-MWF and the rank-1 SDW-MWF are compared through a series of experiments where parameters such as the input SNR, spatial angles, number of noise sources, and the reverberation times are changed.

2.1 Preliminaries

A general set-up of a multi-channel noise reduction is shown in Figure 2.1, with M microphones in an environment with one or more noise sources and a desired speaker. First the microphone(s) capture the input signals and then all of the input signals are filtered and finally the outputs from the filters are summed together to produce the output signal. Let $X_i(k, l)$, $i = 1, \dots, M$ denote the frequency-domain microphone signals

$$X_i(k, l) = X_i^s(k, l) + X_i^n(k, l) \quad (2.1)$$

where k is the frequency bin index, and l the frame index of a short-time Fourier transform (STFT), and the superscripts s and n are used to refer to the speech and the noise contribution in a signal, respectively. Let $\mathbf{X}(k, l) \in \mathbb{C}^{M \times 1}$ be defined

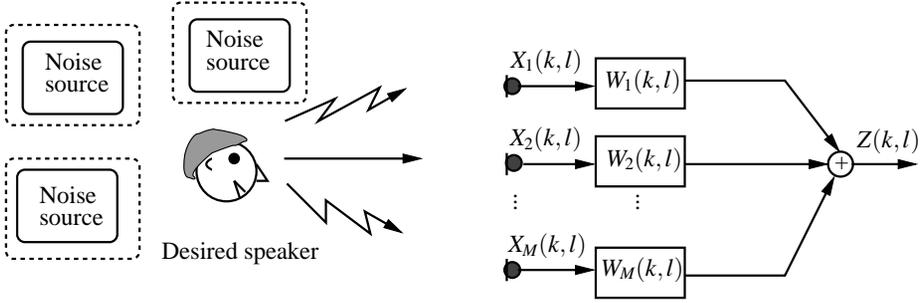


Figure 2.1: Multi-channel noise reduction set-up in an environment with one or more noise sources and a desired speaker.

as the stacked vector

$$\mathbf{X}(k, l) = [X_1(k, l) \ X_2(k, l) \ \dots \ X_M(k, l)]^T \quad (2.2)$$

$$= \mathbf{X}^s(k, l) + \mathbf{X}^n(k, l) \quad (2.3)$$

where the superscript T denotes the transpose. In addition, we define the speech-plus-noise, the speech and the noise correlation matrices as

$$\mathbf{R}_x(k, l) = \varepsilon\{\mathbf{X}(k, l)\mathbf{X}^H(k, l)\} \quad (2.4)$$

$$\mathbf{R}_s(k, l) = \varepsilon\{\mathbf{X}^s(k, l)\mathbf{X}^{s,H}(k, l)\} \quad (2.5)$$

$$\mathbf{R}_n(k, l) = \varepsilon\{\mathbf{X}^n(k, l)\mathbf{X}^{n,H}(k, l)\} \quad (2.6)$$

where $\varepsilon\{\}$ denotes the expectation operator, H denotes Hermitian transpose.

2.1.1 Estimation of correlation matrices

For the estimation of the correlation matrices we will define a two-state model for speech events which can be expressed given two hypotheses $H_0(k, l)$ and $H_1(k, l)$ which represent speech absence and speech presence in frequency bin k of frame l , respectively, i.e.,

$$H_0(k, l) : \mathbf{X}(k, l) = \mathbf{X}^n(k, l)$$

$$H_1(k, l) : \mathbf{X}(k, l) = \mathbf{X}^s(k, l) + \mathbf{X}^n(k, l). \quad (2.7)$$

When a binary VAD is used to distinguish between the $H_0(k, l)$ and the $H_1(k, l)$ state, the correlation matrix estimation can be written as

$$H_0(k, l) : \begin{cases} \hat{\mathbf{R}}_n(k, l+1) = \alpha_n \hat{\mathbf{R}}_n(k, l) + (1 - \alpha_n) \mathbf{X}(k, l) \mathbf{X}^H(k, l) \\ \hat{\mathbf{R}}_x(k, l+1) = \hat{\mathbf{R}}_x(k, l) \end{cases} \quad (2.8)$$

and

$$H_1(k, l) : \begin{cases} \hat{\mathbf{R}}_x(k, l+1) = \alpha_x \hat{\mathbf{R}}_x(k, l) + (1 - \alpha_x) \mathbf{X}(k, l) \mathbf{X}^H(k, l) \\ \hat{\mathbf{R}}_n(k, l+1) = \hat{\mathbf{R}}_n(k, l) \end{cases} \quad (2.9)$$

where α_n and α_x are the forgetting factors for the noise-only and the speech-plus-noise correlation matrix, respectively. The second-order statistics of the noise are assumed to be (short-term) stationary which means that $\mathbf{R}_s(k, l)$ can be estimated as $\hat{\mathbf{R}}_s(k, l) = \hat{\mathbf{R}}_x(k, l) - \hat{\mathbf{R}}_n(k, l)$ where $\hat{\mathbf{R}}_x(k, l)$ and $\hat{\mathbf{R}}_n(k, l)$ are estimated (i.e. adapted) during periods of speech-plus-noise and periods of noise-only, respectively (and "frozen" otherwise). The concept of using (2.8) and (2.9) to update the correlation matrices is shown in Figure 2.2 where the binary VAD is normally referred to as a perfect VAD. In this work we will focus on the MWF since this approach does not require any a priori information regarding the desired signal model. The MWF is uniquely based on the correlation matrices defined (2.8) and (2.9). This of course makes the MWF sensitive to errors in the estimated correlation matrices which will be the main part of this work regarding the MWF.

2.2 Multi-channel Wiener filter (MWF)

The MWF optimally estimates the speech signal, based on a Minimum Mean Squared Error (MMSE) criterion, i.e.,

$$\mathbf{W}_{\text{MMSE}}(k, l) = \arg \min_{\mathbf{W}(k, l)} \varepsilon\{ |X_1^s(k, l) - \mathbf{W}^H(k, l) \mathbf{X}(k, l)|^2 \} \quad (2.10)$$

where the desired signal in this case is the (unknown) speech component $X_1^s(k, l)$ in the first microphone signal (could be any other microphone signal). Notice that the delay in the speech component $X_1^s(k, l)$ is assumed to be zero. Typically a delay is included to allow for non-causal taps in the filter $\mathbf{W}(k, l)$. The equation in (2.10) can be written as

$$\begin{aligned} J_{\text{MMSE}}(\mathbf{W}(k, l)) &= \varepsilon\{ |X_1^s(k, l) - \mathbf{W}^H(k, l) \mathbf{X}(k, l)|^2 \} \\ &= \varepsilon\{ X_1^s(k, l) X_1^{s,H}(k, l) \} + \varepsilon\{ \mathbf{W}^H(k, l) \mathbf{X}(k, l) \mathbf{X}^H(k, l) \mathbf{W}(k, l) \} \\ &\quad - \varepsilon\{ X_1^s(k, l) \mathbf{X}^H(k, l) \mathbf{W}(k, l) \} - \varepsilon\{ \mathbf{W}^H(k, l) \mathbf{X}(k, l) X_1^{s,H}(k, l) \} \end{aligned} \quad (2.11)$$

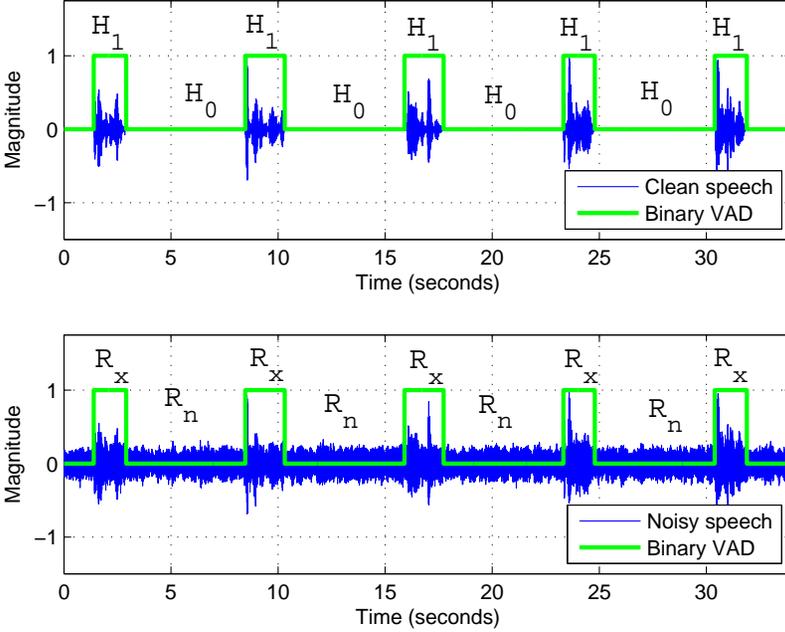


Figure 2.2: Illustration of updating speech-plus-noise and noise-only correlation matrices using a binary VAD.

which is minimized by setting the derivative of (2.11) to zero, i.e.,

$$\frac{\partial J_{\text{MMSE}}(\mathbf{W}(k, l))}{\partial \mathbf{W}(k, l)} = -2\varepsilon\{\mathbf{X}(k, l)X_1^{s, H}(k, l)\} + 2\varepsilon\{\mathbf{X}(k, l)\mathbf{X}^H(k, l)\}\mathbf{W}(k, l). \quad (2.12)$$

We assume that the speech and the noise signals are uncorrelated, i.e.,

$$\varepsilon\{\mathbf{X}^n(k, l)X_1^{s, H}(k, l)\} = \mathbf{0} \quad (2.13)$$

such that the estimated speech correlation vector can be written as

$$\varepsilon\{\mathbf{X}^s(k, l)X_1^{s, H}(k, l)\} = \varepsilon\{\mathbf{X}(k, l)X_1^H(k, l)\} - \varepsilon\{\mathbf{X}^n(k, l)X_1^{n, H}(k, l)\}. \quad (2.14)$$

Finally by solving (2.12) the MWF is given by

$$\mathbf{W}_{\text{MMSE}}(k, l) = \left[\mathbf{R}_s(k, l) + \mathbf{R}_n(k, l) \right]^{-1} \mathbf{R}_s(k, l) \mathbf{e}_1 \quad (2.15)$$

where the $M \times 1$ vector \mathbf{e}_1 equals the first canonical vector defined as $\mathbf{e}_1 = [1 \ 0 \ \dots \ 0]^T$. The estimated speech component $Z^s(k, l)$ in one of the microphone signals can then be written as

$$Z^s(k, l) = \mathbf{W}_{\text{MMSE}}^H(k, l)\mathbf{X}(k, l). \quad (2.16)$$

Notice that $Z^s(k, l)$ is the complete output from the MWF based NR which may contain some residual noise depending on the estimated filter $\mathbf{W}_{\text{MMSE}}^H(k, l)$. The superscript s is used here to avoid confusion since the (unknown) noise component in one of the microphone signals can be estimated by modifying the MMSE criterion in (2.10) to

$$\mathbf{V}_{\text{MMSE}}(k, l) = \arg \min_{\mathbf{V}} \varepsilon\{|X_1^n(k, l) - \mathbf{V}^H(k, l)\mathbf{X}(k, l)|^2\} \quad (2.17)$$

and the solution is given by

$$\mathbf{V}_{\text{MMSE}}(k, l) = \left[\mathbf{R}_n(k, l) + \mathbf{R}_s(k, l) \right]^{-1} \mathbf{R}_n(k, l)\mathbf{e}_1. \quad (2.18)$$

The estimated noise component $Z^n(k, l)$ in one of the microphone signals can then be written as

$$Z^n(k, l) = \mathbf{V}_{\text{MMSE}}^H(k, l)\mathbf{X}(k, l). \quad (2.19)$$

Again depending on the estimated filter $\mathbf{V}_{\text{MMSE}}^H(k, l)$ the estimated noise component $Z^n(k, l)$ may contain some residual speech.

2.3 Speech distortion weighted MWF (SDW-MWF_μ)

The MWF has been extended to the SDW-MWF_μ that allows for a trade-off between noise reduction and speech distortion [49][207]. If the speech and the noise signals are uncorrelated the residual error energy of the MWF in (2.10)

$$\varepsilon\{|X_1^s(k, l) - \mathbf{W}^H(k, l)\mathbf{X}(k, l)|^2\} \quad (2.20)$$

can be further decomposed into

$$\varepsilon\{|X_1^s(k, l) - \mathbf{W}^H(k, l)\mathbf{X}^s(k, l)|^2\} + \varepsilon\{|\mathbf{W}^H(k, l)\mathbf{X}^n(k, l)|^2\} \quad (2.21)$$

where the first term corresponds to the speech distortion energy and the last term corresponds to the residual noise energy. Using (2.21) the MWF can be extended to allow for a trade-off between NR and speech distortion by incorporating a weighting factor μ . The design criterion of the SDW-MWF_μ is given by

$$\begin{aligned} \mathbf{W}_{\text{MWF}_\mu}(k, l) = \arg \min_{\mathbf{W}(k, l)} \varepsilon\{|X_1^s(k, l) - \mathbf{W}^H(k, l)\mathbf{X}^s(k, l)|^2\} + \\ \mu\varepsilon\{|\mathbf{W}^H(k, l)\mathbf{X}^n(k, l)|^2\} \end{aligned} \quad (2.22)$$

and the solution of SDW-MWF $_{\mu}$ is then given by

$$\mathbf{W}_{\text{MWF}_{\mu}}(k, l) = \left[\mathbf{R}_s(k, l) + \mu \mathbf{R}_n(k, l) \right]^{-1} \mathbf{R}_s(k, l) \mathbf{e}_1. \quad (2.23)$$

For $\mu = 1$ the SDW-MWF $_{\mu}$ reduces to the MWF solution in (2.15), while for $\mu > 1$ the residual noise level will be reduced at the cost of a higher speech distortion. By setting $\mu = \infty$ all emphasis is put on the NR and the speech distortion is completely ignored and if $\mu = 0$ no NR will be performed.

2.4 Rank-1 SDW-MWF $_{\mu}$

The rank-1 MWF is based on a single target speech source assumption such that the speech signal can be written as

$$\mathbf{X}^s(k, l) = \mathbf{a}(k, l)S(k, l) \quad (2.24)$$

where $S(k, l)$ is the clean speech signal modelled with the M -dimensional steering vector \mathbf{a} that contains the acoustic transfer functions from the speech source to the microphones including room acoustics, microphone characteristics and head shadow effect. Furthermore, the rank-one speech correlation matrix can be written as

$$\mathbf{R}_s(k, l) = P_s \mathbf{a}(k, l) \mathbf{a}^H(k, l) \quad (2.25)$$

with $P_s = \varepsilon\{|S(k, l)|^2\}$ the power of the clean speech signal. Using the definitions in (2.24) and (2.25) it has been shown in [38][66][76][198][206] that the SDW-MWF $_{\mu}$ formulation in (2.23) can be decomposed into a spatial filter and a single-channel postfilter which can be written as

$$\mathbf{W}_{\text{R1-MWF}_{\mu}}(k, l) = \underbrace{\mathbf{R}_n^{-1}(k, l) \mathbf{R}_s(k, l) \mathbf{e}_1}_{\text{spatial filter}} \cdot \underbrace{\frac{1}{\mu + \text{Tr}\{\mathbf{R}_n^{-1}(k, l) \mathbf{R}_s(k, l)\}}}_{\text{single-channel postfilter}} \quad (2.26)$$

where $\text{Tr}\{\cdot\}$ is the trace operator. It is worth noting that the speech distortion weighting factor only appears in the single-channel postfilter and not in the spatial filter. In this case the single-channel postfilter has a similar role as a single-channel constrained Wiener filter [60][133]. If the noise-level is under-estimated or further attenuation of the residual noise is desired the single-channel postfilter can also be viewed as a spectral oversubtraction [11][133]. In this case increasing μ allows for further attenuation of the residual noise which possibly comes at the cost of higher speech distortion.

The overall blocks in the traditional SDW-MWF based NR are shown in Figure 2.3. First an STFT is performed on each frame using an overlap-add or an overlap-save procedure on the noisy speech. Then a binary VAD decides whether the speech-plus-noise or the noise-only correlation matrix should be updated. Based on these correlation matrices the SDW-MWF filter is formed and the filtering is performed followed by an inverse STFT (ISTFT) and a reconstruction procedure of the output signal. It is also clear that the binary VAD algorithm is a crucial part of the algorithm here.

2.5 Analysis of the SDW-MWF_μ

This section addresses two main problems related to estimating and updating the correlation matrices which is of great importance since the MWF based NR is uniquely based on these correlation matrices. Some properties and drawbacks of the SDW-MWF_μ are:

- SDW-MWF_μ is uniquely based on the second-order statistics and in the estimation of the speech-plus-noise and the noise-only correlation matrices (containing spectral as well as spatial information) an averaging time window of 2-3 seconds is typically required to achieve a reliable estimate.
- However speech is a spectrally non-stationary signal and can be considered stationary only in a short time window, much shorter than the 2-3 seconds time window used for the estimation of $\hat{\mathbf{R}}_s(k, l)$, and so the spectral non-stationarity is not captured in $\hat{\mathbf{R}}_s(k, l)$. Also to obtain the noise-only correlation matrix the MWF assumes that the noise statistics are sufficiently stationary such that $\hat{\mathbf{R}}_n(k, l)$ can be updated during noise-only periods.
- Therefore the MWF is highly dependent on the long-time average of the spectral and the spatial signal characteristics. Even though the spatial characteristic can be assumed to vary slowly the assumption is not valid for noise sources such as multi-talker babble. This implies that the MWF can suppress spectrally non-stationary noise, provided that the long-term spectral characteristic of the speech and the noise vary slowly over time. because the filter is only slowly varying. This of course reduces effects such as musical noise.
- In the SDW-MWF_μ the weighting factor μ is a fixed value for each frame and for each frequency. As a consequence, any improvement in NR (by setting μ to a larger value) comes at the cost of a higher speech distortion since speech dominant segments and noise dominant segments are weighted equally.
- Furthermore typical speech signals contain many pauses while the noise is assumed to be continuously present. This means that the weighting factor

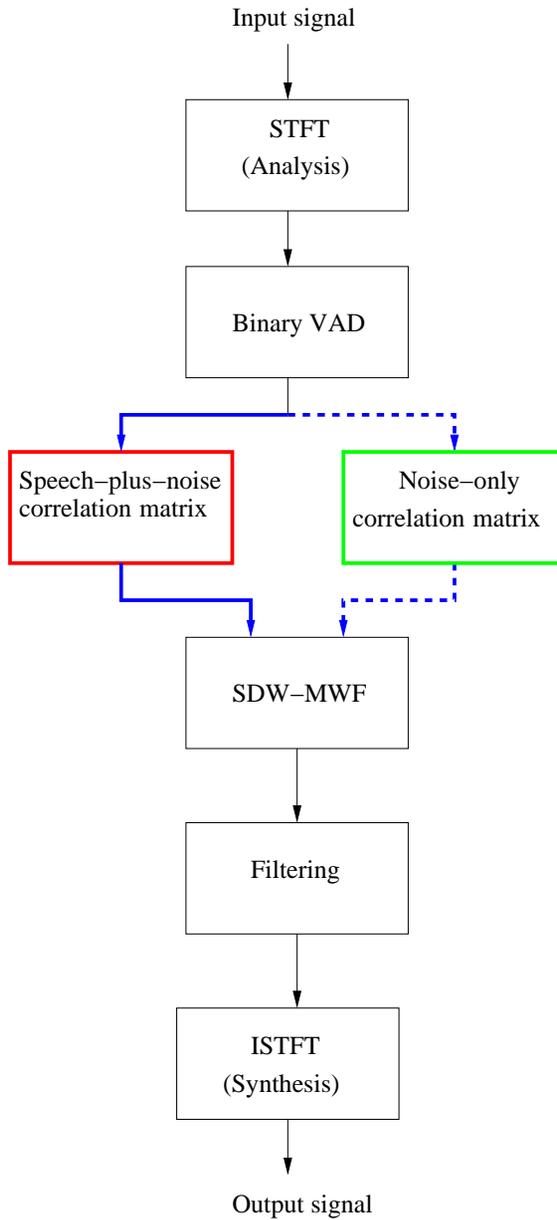


Figure 2.3: Block diagram of the SDW-MWF_μ with a binary VAD for estimating the correlation matrices.

could indeed apply a different weighting dependent on whether speech is present or absent.

2.5.1 Robustness and tracking

In the MWF set-up, the tracking performance depends on the forgetting factors α_x and α_n whereas the robustness is related to the correct detection of the $H_0(k, l)$ and the $H_1(k, l)$ state. On the other hand, since the noise-only and the speech-plus-noise correlation matrices are updated at different time instants this would consequently also limit the tracking and the accuracy of the estimated correlation matrices especially when dealing with non-stationary noise sources and low input SNR. This can become a major problem since the speech correlation matrix is estimated by subtracting the noise-only correlation matrix from the speech-plus-noise correlation matrix, i.e.,

$$\hat{\mathbf{R}}_s(k, l) = \hat{\mathbf{R}}_x(k, l) - \hat{\mathbf{R}}_n(k, l). \quad (2.27)$$

First of all there is no guarantee that the spectral and the spatial signal characteristic remain fixed from a $H_0(k, l)$ to a $H_1(k, l)$ state especially since speech and noise are considered to be non-stationary. Consequently, the subtraction performed in (2.27) can lead to a poor estimate especially if $\hat{\mathbf{R}}_n(k, l)$ is larger or equal to $\hat{\mathbf{R}}_x(k, l)$ which in theory should not happen. A scenario could be that the noise level at a $H_0(k, l)$ state is higher than in the subsequent $H_1(k, l)$ state and since the $\hat{\mathbf{R}}_n(k, l)$ is not updated in the $H_1(k, l)$ state the NR performance can be severely compromised. This makes the estimation of $\hat{\mathbf{R}}_n(k, l)$ the key factor and especially its spectral characteristic rather than the spatial characteristic.

The aim of this work is to improve the robustness and the tracking capabilities when estimating the correlation matrices. First we will address the issue of having a fixed weighting factor μ for each frequency and for each frame in Chapter 3. Secondly, we will show that the update concept using (2.8) and (2.9) is not optimal when the speech and the noise sources are non-stationary and when dealing with a low input SNR scenario. Furthermore we will introduce a method to avoid the subtraction in (2.27) leading to a poor estimate by introducing the usage of prior correlation matrices combined with a continuous update approach which, will be presented in Chapter 4.

2.6 Experimental results

In this section, experimental results for the SDW-MWF_μ and for the rank-1 SDW-MWF_μ are presented and compared against each other. The estimation and the

Notation	Spatial angle of source(s)
S0N0	Speech at 0°, noise source(s) at 0°
S0N30	Speech at 0°, noise source(s) at 30°
S0N60	Speech at 0°, noise source(s) at 60°
S0N90	Speech at 0°, noise source(s) at 90°
S0N120	Speech at 0°, noise source(s) at 120°
S0N2a	Speech at 0°, noise source(s) at 90°, 180°
S0N3a	Speech at 0°, noise source(s) at 90°, 180°, 270°
S0N2b	Speech at 0°, noise source(s) at 45°, 90°
S0N3b	Speech at 0°, noise source(s) at 45°, 90°, 180°
S0N2c	Speech at 0°, noise source(s) at 30°, 60°
S0N3c	Speech at 0°, noise source(s) at 30°, 60°, 90°

Table 2.1: Spatial scenarios for the experimental evaluation

update of the correlation matrices are based on a perfect VAD since this is the traditional way to estimate and update the correlation matrices. The experimental results here then serve as the baseline for the proposed SDW-MWF based NR algorithms in the next chapters.

2.6.1 Experimental set-up

Simulations have been performed with a 2-microphone (with an intermicrophone distance of approximately 1cm) BTE hearing aid mounted on a CORTEX MK2 manikin such that the head-shadow effect is included. The loudspeakers (FOSTEX 6301B) are positioned at 1 meter from the center of the head. Two reverberation times are used, i.e., $T_{60} \approx 0.61s$ which is considered to be a high reverberation scenario and $T_{60} \approx 0.21s$ which is considered to be a low reverberation scenario [43][44]. The microphone signals are generated by convolving the speech and the noise signals with the HRTFs corresponding to the predefined angles of arrival and finally the signals are mixed together at a specific SNR. The speech signal consists of male sentences from the Hearing in Noise Test (HINT) for the measurement of SRTs in quiet and in noise [167] and the noise signal consists of multi-talker babble from Auditory Tests (Revised), Compact Disc, Auditec [5]. The signals are sampled at 16kHz. For the analysis step an STFT length of 128 with 50% overlap with a Hanning window is used to extract the frames. For the correlation matrices the forgetting factors are set to $\alpha_n = \alpha_x = 0.9980$ corresponding to an averaging time of 2 seconds. The algorithms are evaluated at -7.5dB, -5dB, and 0dB input SNR. Besides the various input SNRs the spatial angles of the noise sources are also varied and Table 2.1 shows the 11 different spatial scenarios used in the evaluation. Since the recordings of the clean speech and the noise-only

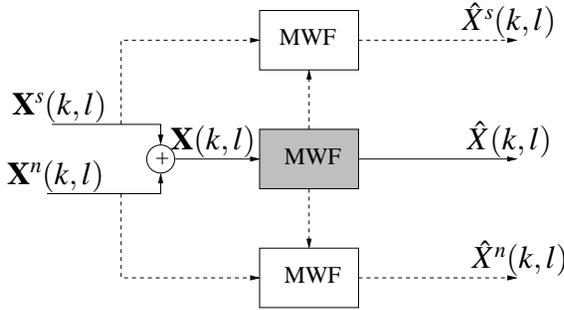


Figure 2.4: Block diagram of the signals used for the objective quality measures.

signals are available the signals used in the objective measures can be estimated as

$$\hat{X}^s(k, l) = \mathbf{W}^H(k, l)\mathbf{X}^s(k, l). \quad (2.28)$$

and

$$\hat{X}^n(k, l) = \mathbf{W}^H(k, l)\mathbf{X}^n(k, l) \quad (2.29)$$

where $\hat{X}^s(k, l)$ and $\hat{X}^n(k, l)$ are the clean speech and the noise-only signal filtered, respectively, as shown in Figure 2.4. The NR performance is then quantified by comparing the ratio between $\hat{X}^s(k, l)$ and $\hat{X}^n(k, l)$ with the ratio between the clean speech $X_1^s(k, l)$ and the noise-only signal $X_1^n(k, l)$ in the reference microphone. The signal distortion is estimated by comparing $\hat{X}^s(k, l)$ with $X_1^s(k, l)$.

2.6.2 Performance measures

To assess the noise reduction performance the intelligibility-weighted signal-to-noise ratio (SNR) [73] is used which is defined as

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i (\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}) \quad (2.30)$$

where I_i is the band importance function defined in ANSI S3.5-1997 [1] and where $\text{SNR}_{i,\text{out}}$ and $\text{SNR}_{i,\text{in}}$ represent the output SNR and the input SNR (in dB) of the i -th band, respectively. For measuring the signal distortion a frequency-weighted log-spectral signal distortion (SD) is used defined as

$$\text{SD} = \frac{1}{K} \sum_{k=1}^K \sqrt{\int_{f_l}^{f_u} w_{\text{ERB}}(f) \left(10 \log_{10} \frac{P_{\text{out},k}^s(f)}{P_{\text{in},k}^s(f)} \right)^2 df} \quad (2.31)$$

where K is the number of frames, $P_{out,k}^s(f)$ is the output power spectrum of the k th frame, $P_{in,k}^s(f)$ is the input power spectrum of k th frame and f is the frequency index. The SD measure is calculated with a frequency-weighting factor $w_{ERB}(f)$ giving equal weight for each auditory critical band, as defined by the equivalent rectangular bandwidth (ERB) of the auditory filter [147]. Notice that the intelligibility-weighted SNR and the signal distortion are only computed during frames of speech-plus-noise.

2.6.3 Results

Simulation results for the SDW-MWF $_{\mu}$ compared to the rank-1 SDW-MWF $_{\mu}$ for a low reverberation scenario are shown in Figure 2.5. From the results it is clear that increasing μ improves the SNR improvement for the SDW-MWF $_{\mu}$ however this comes at the cost of increased signal distortion. For the rank-1 SDW-MWF $_{\mu}$ the SNR improvement seems to be equal to SDW-MWF $_{\mu}$ with $\mu = 5$ but the signal distortion is comparable with SDW-MWF $_{\mu}$ with $\mu = 1$. It should also be mentioned that increasing μ in the rank-1 SDW-MWF $_{\mu}$ does not affect the SNR improvement. It is also clear that the NR performance is better for certain spatial scenarios such as S0N60, S0N90, and S0N120. It is also worth noting that for these particular spatial angles the signal distortion does not increase when $\mu= 3$ and 5 as opposed to the other spatial angles. This suggests, that if the filter is working properly, increasing μ will not increase the signal distortion. Spatial scenarios with one noise source also show less signal distortion compared to scenarios with multiple noise sources which is especially clear at low input SNR.

The same simulation for a high reverberation scenario is shown in Figure 2.6. This shows that the reverberation indeed has a negative impact on the SNR improvement especially for spatial scenarios such as S0N60, S0N90, and S0N120 which showed a great improvement for the low reverberation scenario. At these particular spatial scenarios the signal distortion is also increased compared to the scenario with the low reverberation. However a greater concern is that for other spatial scenarios the high reverberation results in a very limited NR performance even for the rank-1 SDW-MWF $_{\mu}$. For example spatial scenarios such as S0N0, S0N30, S0N2c, and S0N3c seem to cause problems on the NR performance, i.e., when the noise sources are close to the desired speaker or when multiple noise sources are used. This implies that the spatial filtering in these scenarios may be rather poor whereas the spectral filtering is limited by the correlation matrices. However the benefit of the rank-1 SDW-MWF $_{\mu}$ is still that the signal distortion is kept low even at high reverberation.

2.7 Conclusion

In this chapter we have introduced the MWF based NR which provides an MMSE estimate of the speech component in one of the microphone signals. We have also shown that the MWF can be formulated in a way that allows for a trade-off between NR and speech distortion which is referred to as SDW-MWF_μ. Another interesting approach that has been proposed is the rank-1 SDW-MWF_μ where the standard SDW-MWF_μ is decomposed into a spatial filter and a single-channel postfilter. A property of the rank-1 SDW-MWF_μ is that the formulation is now less sensitive to errors in the estimated correlation matrices and therefore a better NR performance is achieved.

Experimental results shows that the SDW-MWF_μ is able to improve the SNR when μ is increased but this comes at the cost of a higher signal distortion. However the rank-1 SDW-MWF_μ is able to provide the same SNR improvement without increasing the signal distortion. Furthermore it has been shown that a high reverberation negatively affects the NR performance compared to a low reverberation. For this reason the development on the SDW-MWF based NR is primarily based on high reverberation and low input SNR scenarios.

Several open problems have been discussed related to the way the correlation matrices are estimated and updated. Especially the estimation of the clean speech correlation matrix can be a problem since it is based on subtracting the noise-only correlation matrix from the speech-plus-noise correlation matrix which makes the estimation of the noise-only correlation matrix very important. Furthermore, it has been pointed out that the spectral tracking can be limited due to the fact that the speech-plus-noise correlation matrix is kept fixed during noise-only periods and the noise correlation matrix is kept fixed during speech-plus-noise periods. Other factors that can limit the spectral and spatial tracking is the long time averaging of the correlation matrices. Questions have also been raised as to whether the weighting factor μ is optimal since it is a fixed value for each frequency and for each frame considering the non-stationarity of the speech and the noise.

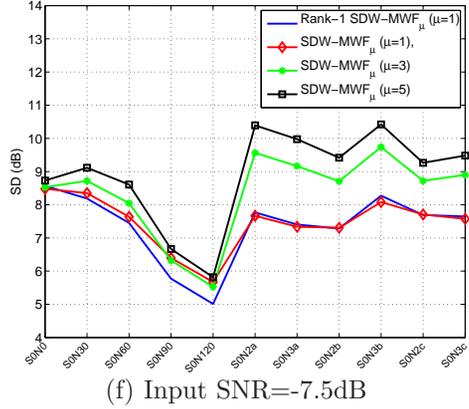
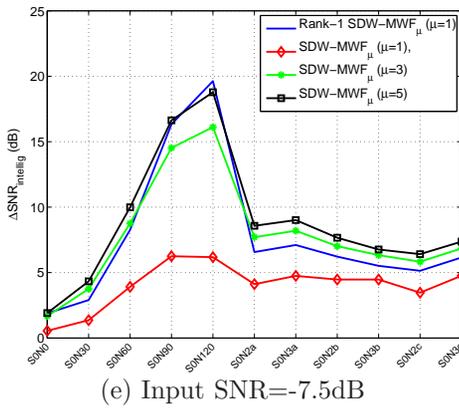
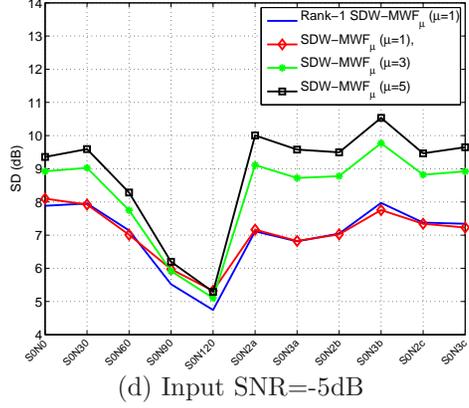
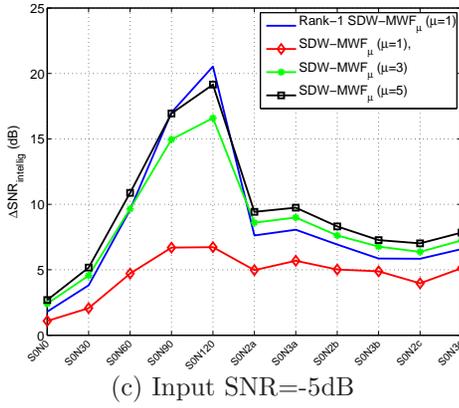
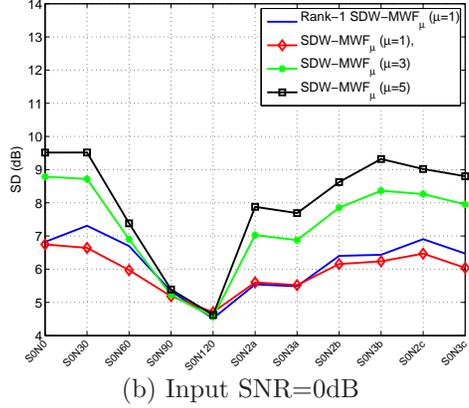
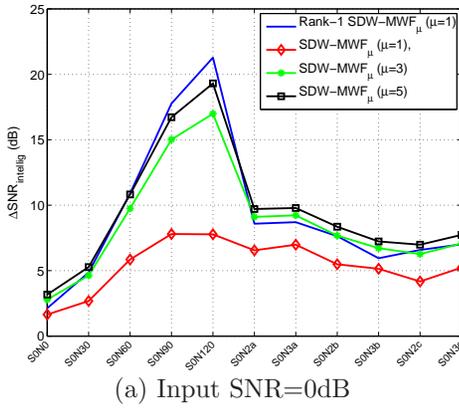


Figure 2.5: Comparison of the SDW-MWF_μ and the rank-1 SDW-MWF_μ for a low reverberation scenario using objective measures.

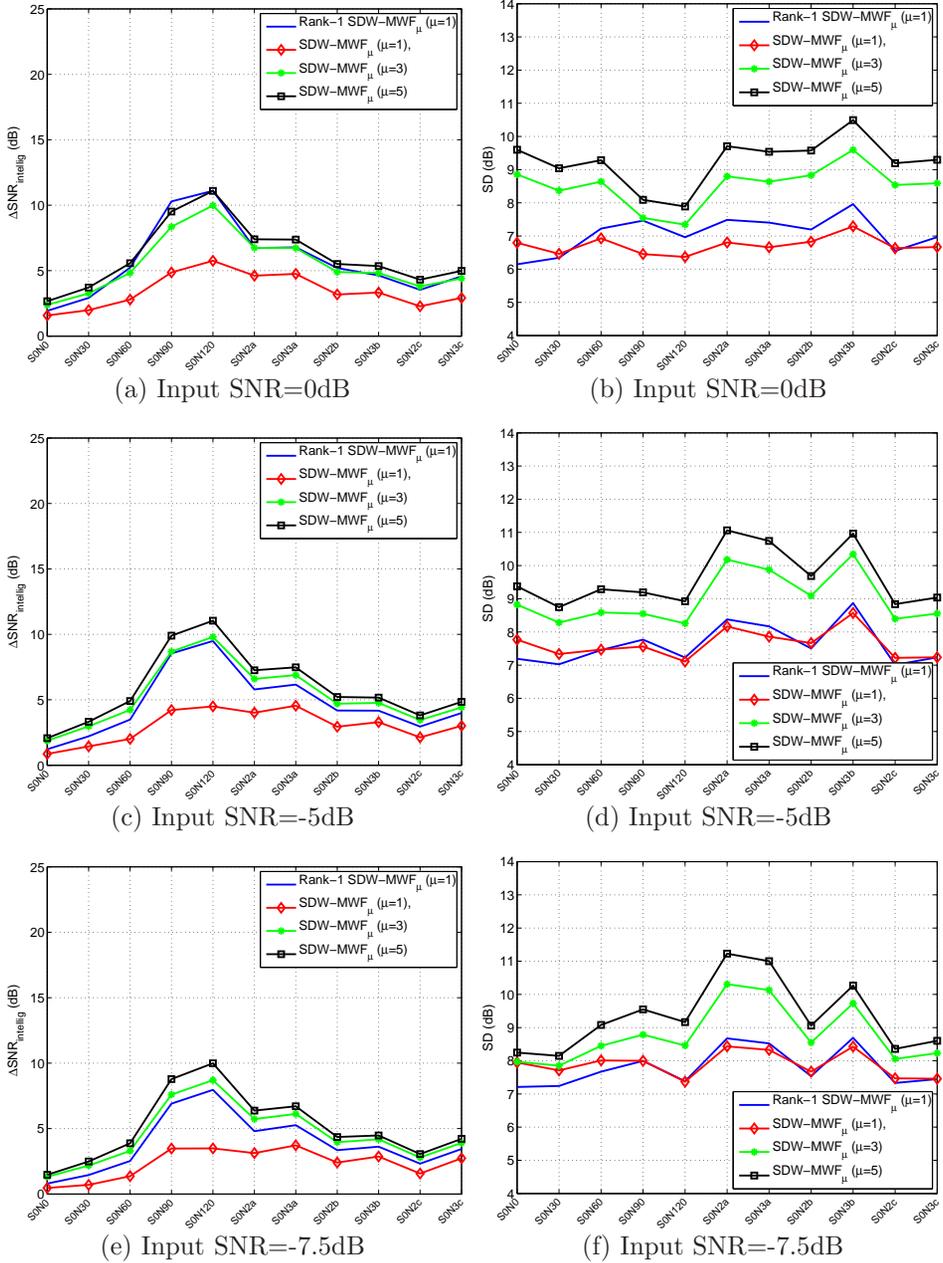


Figure 2.6: Comparison of the SDW-MWF_μ and the rank-1 SDW-MWF_μ for a high reverberation scenario using objective measures.

Chapter 3

SDW-MWF _{μ} based on speech presence probability (SPP)

This chapter address the issue of using a weighting factor to trade-off between NR and speech distortion that is kept fixed for each frequency and for each frame which potentially can limit the spectral tracking of the non-stationarity of the speech and the noise. Combined with the fact that the correlation matrices are kept fixed at different time instants all together with a long averaging time the fixed weighting factor certainly does not help the spectral tracking.

To tackle this problem we have been inspired by an interesting technique which has primarily been used in single-channel NR algorithms, referred to as the conditional speech presence probability (SPP). This technique is based on a two-state speech model, i.e, a $H_0(k, l)$ state represents speech absence and a $H_1(k, l)$ state represents speech presence which is defined for each frequency and for each frame. This model is based on the fact that the noise can be assumed to be continuously present whereas speech typically contains many pauses. For this reason the conditional SPP is estimated and updated for each frequency and for each frame. Since single-channel NR algorithms are not able to exploit spatial signal information extensive research has been conducted to obtain a spectral distinction between the speech and the noise which is something that has not received a lot of attention in multi-channel NR algorithms. In multi-channel NR the concern has primarily been to improve the spatial filtering. For this reason we propose that the conditional SPP is incorporated in the SDW-MWF based NR such that the speech dominant segments and the noise dominant segments can be weighted differently.

Section 3.1 introduces the two-state speech model and the estimation of the conditional SPP which is based on parameters such as the a priori SNR, a posteriori

SNR, and a priori speech absence probability (SAP).

Section 3.2 extends the MMSE criterion of the MWF to incorporate the conditional SPP leading to the SDW-MWF_{SPP}. The derivation of the SDW-MWF_{SPP} shows that the conditional SPP replaces the fixed weighting factor such that the speech dominant segments and the noise dominant segments can be weighted according to the conditional SPP. A minor extension is also proposed referred to as SDW-MWF_{combined} such that in one extreme case the solution corresponds to the SDW-MWF _{μ} and in the other extreme case to the SDW-MWF_{SPP}.

Section 3.3 introduces an extension of the SDW-MWF_{SPP} which is referred to as SDW-MWF_{flex}. The idea here is to use the conditional SPP combined with a per frame decision such that a flexible weighting factor is introduced. The flexible weighting factor is based on the fact that the noise in the $H_0(k, l)$ state can be weighted differently compared to the noise in the $H_1(k, l)$ state. Since in the $H_0(k, l)$ state the speech is absent and hence signal distortion or speech intelligibility does not need to be taken into consideration, therefore more noise can be suppressed in the $H_0(k, l)$ state compared to the $H_1(k, l)$ state.

Section 3.4 extends the SDW-MWF_{SPP} to a rank-1 SDW-MWF_{SPP} which allows to show how the conditional SPP affects the SDW-MWF based NR. It is shown here that the conditional SPP only affects the single-channel postfilter which is interesting since the spectral weighting now only affects the single-channel postfilter and not the spatial filter. The same extension is also shown for the flexible weighting factor referred to as rank-1 SDW-MWF_{flex}.

Section 3.5 presents the experimental results in order to confirm the influence of the proposed weighting factor that is now updated for each frequency and for each frame.

3.1 Conditional speech presence probability (SPP)

The key component to tackle the problems introduced above is the conditional SPP. In this work, we will use a multi-channel approach to estimate the conditional SPP, i.e., all available microphone signals are used [200]. Previous work has been based on a single-channel approach where the reference microphone was used to estimate the conditional SPP [35][161].

The conditional SPP is estimated for each frequency and for each frame by a soft-decision approach [33][36][133], which exploits the strong correlation of speech presence in neighboring frequency bins of consecutive frames. The two-state model

for speech events introduced in Chapter 2 is modified here, i.e.,

$$\begin{aligned} H_0(k, l) : X_i(k, l) &= X_i^n(k, l) + 0 \cdot X_i^s(k, l) \\ H_1(k, l) : X_i(k, l) &= X_i^n(k, l) + 1 \cdot X_i^s(k, l), \end{aligned} \quad (3.1)$$

The inclusion of the second term in the definition of $H_0(k, l)$ will be explained in Section 3.3. Assuming a complex Gaussian distribution of the STFT coefficients for both the speech and the noise, the conditional probability density functions (PDF) of the observed signals are given by

$$p(X_i(k, l)|H_0(k, l)) = \frac{1}{\pi\lambda_i^n(k, l)} \exp\left\{-\frac{|X_i(k, l)|^2}{\lambda_i^n(k, l)}\right\} \quad (3.2)$$

$$p(X_i(k, l)|H_1(k, l)) = \frac{1}{\pi(\lambda_i^s(k, l) + \lambda_i^n(k, l))} \exp\left\{-\frac{|X_i(k, l)|^2}{\lambda_i^s(k, l) + \lambda_i^n(k, l)}\right\} \quad (3.3)$$

where $\lambda_i^s(k, l) \triangleq \varepsilon\{|X_i^s(k, l)|^2\}$ and $\lambda_i^n \triangleq \varepsilon\{|X_i^n(k, l)|^2\}$ denote the power spectrum of the speech and the noise, respectively. Applying Bayes rule, the conditional SPP $p(k, l) \triangleq P(H_1(k, l)|X_i(k, l))$ can be written as [35][56][200]

$$p(k, l) = \left\{ 1 + \frac{q(k, l)}{1 - q(k, l)} (1 + \xi(k, l)) \exp(-v(k, l)) \right\}^{-1} \quad (3.4)$$

where $q(k, l) \triangleq P(H_0(k, l))$ is the a priori speech absence probability (SAP), $\xi(k, l)$ is the a priori SNR, and $v(k, l)$ is defined as

$$v(k, l) \triangleq \frac{\gamma(k, l)\xi(k, l)}{(1 + \xi(k, l))}. \quad (3.5)$$

where $\gamma(k, l)$ is the a posteriori SNR. It should be mentioned that (3.4) remains the same whether a single-channel or a multi-channel approach is used, i.e., only the signals used to estimate the SAP, the a priori SNR, and the posteriori SNR are changed, which will be explained next [33][36][133][161].

For the sake of conciseness the frequency bin index k and frame index l are omitted from now on in $\mathbf{X}(k, l)$, $\mathbf{X}^s(k, l)$, $\mathbf{X}^n(k, l)$ and $X_1^s(k, l)$.

3.1.1 Multi-channel a priori and a posteriori SNR estimation

Recently it has been proposed to use all microphone signals to estimate the a priori SNR and the a posteriori SNR based on the estimated correlation matrices [200]. The multi-channel a posteriori SNR estimate can then be written as

$$\hat{\gamma}(k, l) = \text{tr} \left[\hat{\mathbf{R}}_n^{-1}(k, l) \hat{\mathbf{R}}_x(k, l) \right] \quad (3.6)$$

and in a similar manner the multi-channel a priori SNR estimate can be written as

$$\hat{\xi}(k, l) = \text{tr} \left[\hat{\mathbf{R}}_n^{-1}(k, l) \hat{\mathbf{R}}_s(k, l) \right] \quad (3.7)$$

and $\hat{\mathbf{R}}_s(k, l) = \hat{\mathbf{R}}_x(k, l) - \hat{\mathbf{R}}_n(k, l)$. The advantage here is that these correlation matrices are already estimated in the MWF and can therefore be reused.

3.1.2 A priori speech absence probability (SAP) estimation

Reliable estimation of the a priori SNR is important since it is used in the estimation for the a priori SAP. In [33][36] an a priori SAP estimator is proposed based on the time-frequency distribution of the estimated a priori SNR $\hat{\xi}(k, l)$. The estimation is based on three parameters that each exploit the strong correlation of speech presence in neighboring frequency bins of consecutive frames.

First a global and local averaging is applied to $\hat{\xi}(k, l)$ in the frequency domain. Local means that the a priori SNR is averaged over a small number of frequency bins (small bandwidth) and global means that the a priori SNR is averaged over a larger number of frequency bins (larger bandwidth). The local and global averaging of the a priori SNR is given by

$$\zeta_\eta(k, l) = \sum_{i=-\omega_\eta}^{i=\omega_\eta} h_\eta(i) \hat{\xi}(k-i, l) \quad (3.8)$$

where the subscript η represents either local or global averaging and h_η is a normalized Hanning window of size $2\omega_\eta + 1$. The local and global averaging of the a priori SNR is then normalized to values between 0 and 1 before it is mapped into the following threshold function,

$$P_\eta(k, l) = \begin{cases} 0, & \text{if } \zeta_\eta(k, l) \leq \zeta_{min} \\ 1, & \text{if } \zeta_\eta(k, l) \geq \zeta_{max} \\ \frac{\log(\zeta_\eta(k, l)/\zeta_{min})}{\log(\zeta_{max}/\zeta_{min})}, & \text{otherwise} \end{cases} \quad (3.9)$$

where $P_{local}(k, l)$ is the likelihood of speech presence when the a priori SNR is averaged over a small number of frequency bins, and $P_{global}(k, l)$ is the likelihood of speech presence when the a priori SNR is averaged over a larger number of frequency bins. ζ_{min} and ζ_{max} are empirical constants that decide the threshold for speech or noise. The last term $P_{frame}(l)$ represents the likelihood of speech presence in a given frame based on the a priori SNR averaged over all frequency bins, i.e.,

$$\zeta_{frame}(l) = \text{mean}_{1 \leq k \leq N/2+1} \{\zeta(k, l)\} \quad (3.10)$$

Algorithm 1 Estimation of $P_{frame}(l)$

```

1: if  $\zeta_{frame}(l) > \zeta_{min}$  then
2:   if  $\zeta_{frame}(l) > \zeta_{frame}(l-1)$  then
3:      $P_{frame}(l) = 1$ 
4:      $\zeta_{peak}(l) = \min\{\max[\zeta_{frame}(l), \zeta_{p\ min}], \zeta_{p\ max}\}$ 
5:   else
6:      $P_{frame}(l) = \delta(l)$ 
7:   else
8:      $P_{frame}(l) = 0$ 
9:   end if
10: end if
    where

```

$$\delta(l) = \begin{cases} 0, & \text{if } \zeta_{frame}(l) \leq \zeta_{peak}(l) \cdot \zeta_{min} \\ 1, & \text{if } \zeta_{frame}(l) \geq \zeta_{peak}(l) \cdot \zeta_{max} \\ \frac{\log(\zeta_{frame}(l)/\zeta_{peak}(l)/\zeta_{min})}{\log(\zeta_{max}/\zeta_{min})}, & \\ \text{otherwise} & \end{cases} \quad (3.11)$$

where N is the STFT-size. A pseudocode for the computation of $P_{frame}(l)$ is given by Algorithm 1, where $\delta(l)$ represents a soft transition from speech noise, ζ_{peak} is a confined peak value of ζ_{frame} , and $\zeta_{p\ min}$ and $\zeta_{p\ max}$ are empirical constants that determine the delay of the transition. The a priori SAP estimation is then obtained by [33][36]

$$\hat{q}(k, l) = 1 - P_{local}(k, l) \cdot P_{global}(k, l) \cdot P_{frame}(l). \quad (3.12)$$

This means that if either of the previous frames or recent frequency bins does not contain speech, i.e., if the three likelihood terms are small, then $\hat{q}(k, l)$ becomes larger and the conditional SPP $p(\hat{k}, l)$ in (3.4) becomes smaller. The conditional SPP is then estimated by inserting each of these contributions $\hat{q}(k, l)$, $\hat{\xi}(k, l)$, and $\hat{\gamma}(k, l)$ in (3.4).

3.2 SDW-MWF incorporating the conditional SPP (SDW-MWF_{SPP})

The SDW-MWF_{SPP} derived in this section incorporates the conditional SPP in the SDW-MWF μ to allow for a faster tracking of the spectral non-stationarity of the speech, as well as for exploiting the fact that speech may not be present at all time.

3.2.1 Derivation of SDW-MWF $_{\text{SPP}}$

The conditional SPP in (3.4) and the two-state model in (3.1) for speech events can be incorporated into the optimization criterion of the SDW-MWF $_{\mu}$, leading to a weighted average where the first term corresponds to $H_1(k, l)$ and is weighted by the probability that speech is present, while the second term corresponds to $H_0(k, l)$ and is weighted by the probability that speech is absent, i.e.,

$$\begin{aligned} \mathbf{W}_{\text{MWF}_{\text{SPP}}}(k, l) = \arg \min_{\mathbf{W}(k, l)} & p(k, l)\varepsilon\{|\mathbf{X}_1^s - \mathbf{W}^H(k, l)\mathbf{X}|^2|H_1(k, l)\} \\ & + (1 - p(k, l))\varepsilon\{|\mathbf{W}^H(k, l)\mathbf{X}|^2|H_0(k, l)\} \end{aligned} \quad (3.13)$$

where $p(k, l)$ is the conditional probability that speech is present and $(1 - p(k, l))$ is the conditional probability that speech is absent. The solution is then given by

$$\begin{aligned} \mathbf{W}_{\text{MWF}_{\text{SPP}}}(k, l) &= \left[p(k, l)\varepsilon\{\mathbf{X}\mathbf{X}^H|H_1(k, l)\} + (1 - p(k, l))\varepsilon\{\mathbf{X}\mathbf{X}^H|H_0(k, l)\} \right]^{-1} \\ & p(k, l)\varepsilon\{\mathbf{X}\mathbf{X}_1^{s,H}|H_1(k, l)\}, \\ &= \left[p(k, l)\varepsilon\{\mathbf{X}^s\mathbf{X}^{s,H}\} + p(k, l)\varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\} + \right. \\ & \left. (1 - p(k, l))\varepsilon\{\mathbf{X}^n\mathbf{X}^{n,H}\} \right]^{-1} p(k, l)\varepsilon\{\mathbf{X}\mathbf{X}_1^{s,H}\}, \end{aligned} \quad (3.14)$$

which can be written as

$$\mathbf{W}_{\text{MWF}_{\text{SPP}}}(k, l) = \left[\mathbf{R}_s(k, l) + \left(\frac{1}{p(k, l)} \right) \mathbf{R}_n(k, l) \right]^{-1} \mathbf{R}_s(k, l)\mathbf{e}_1. \quad (3.15)$$

Compared to (2.23) the fixed weighting factor μ is replaced by $\frac{1}{p(k, l)}$, which is now adjusted for each frequency bin k and for each frame l , making the SDW-MWF $_{\text{SPP}}$ change with a faster dynamic. The SDW-MWF $_{\text{SPP}}$ offers more noise reduction when $p(k, l)$ is small, i.e., for noise dominant segments, and less noise reduction when $p(k, l)$ is large, i.e., for speech dominant segments.

Figure 3.1 presents a block diagram of the proposed SDW-MWF $_{\text{SPP}}$. First an STFT is performed on each frame of the noisy speech. Then on the left hand side the conditional SPP is estimated, which includes the estimation of the a posteriori SNR, the a priori SNR and the a priori SAP. On the right hand side the frequency domain correlation matrices are estimated, which are used to estimate the filter coefficients after weighting with the conditional SPP. Notice that the updates of the frequency domain correlation matrices are still based on a longer time window. The difference is now that the weights applied in the filter estimation are now changing for each frequency bin and each frame based on the conditional SPP. The last steps include the filtering operation and the ISTFT.

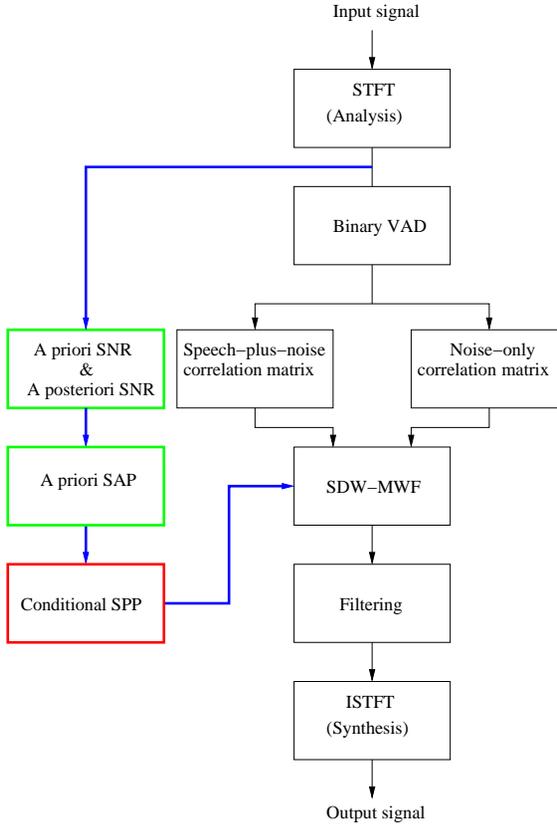


Figure 3.1: Block diagram of the proposed SDW-MWF_{SPP} incorporating the conditional SPP.

3.2.2 Combined solution

A problem with the SDW-MWF_{SPP} derived in (3.15) is that the conditional SPP tends to show significant variation over different frequency bins which then also causes the noise reduction to vary significantly over different frequency bins. We therefore propose a combined solution referred to as SDW-MWF_{combined}, which in one extreme case corresponds to the SDW-MWF_{SPP} and in the other extreme case corresponds to the SDW-MWF _{μ} . The combined solution can be written as

$$\mathbf{W}_{\text{MWF}_{\text{combined}}}(k, l) = \left[\mathbf{R}_s(k, l) + \left(\frac{1}{\alpha(\frac{1}{\mu}) + (1-\alpha)p(k, l)} \right) \mathbf{R}_n(k, l) \right]^{-1} \mathbf{R}_s(k, l) \mathbf{e}_1 \quad (3.16)$$

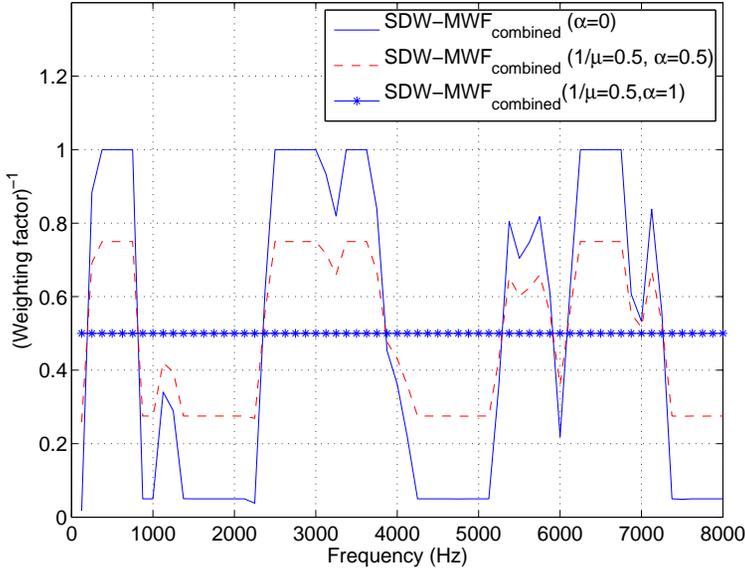


Figure 3.2: Illustration of the different configuration of $(\text{weighting factor})^{-1}$ used in (3.16).

where α is a trade-off factor between SDW-MWF_{μ} and $\text{SDW-MWF}_{\text{SPP}}$. The $(\text{weighting factor})^{-1}$ i.e. $\alpha(\frac{1}{\mu}) + (1 - \alpha)p(k, l)$ used in (3.16) is shown in Figure 3.2 for different configurations, i.e., $\text{SDW-MWF}_{\text{combined}} (\alpha=1)$ and $\text{SDW-MWF}_{\text{combined}} (\alpha=0)$ and $\text{SDW-MWF}_{\text{combined}} (\alpha=0.5)$. The dashed line shows the $(\text{weighting factor})^{-1}$ when $1/\mu=0.5$ and $\alpha=0.5$. This clearly shows that the combined solution corresponds to a scaling of the conditional SPP. Since the variations between the speech dominant segments and the noise dominant segments are reduced, the distortion is also reduced. When $\alpha=1$ the solution corresponds to a fixed μ shown with the (*) marker line. In Section 3.3 a flexible trade-off between noise reduction and speech distortion is introduced.

3.3 SDW-MWF incorporating a flexible weighting factor ($\text{SDW-MWF}_{\text{Flex}}$)

Based on the observations made above the $\text{SDW-MWF}_{\text{SPP}}$ is modified to further exploit the properties of the conditional SPP and to fully exploit the definition of the two-state speech model. The two-state speech model given in (3.1) describes the different states of the noisy speech which leads to the following observations.

First, it is clear that the noise reduction in the $H_0(k, l)$ state and the $H_1(k, l)$ state have a different interpretation, i.e.,

- Reducing the noise in the $H_0(k, l)$ state can be related to increasing listening comfort, since speech is not present in the $H_0(k, l)$ state, which means that a greater attenuation can be applied.
- Reducing the noise in the $H_1(k, l)$ state is a more challenging task since this relates to speech intelligibility and hence the speech distortion weighted concept truly only makes sense in the $H_1(k, l)$ state.

Secondly, as described in Chapter 2, the speech correlation matrix $\mathbf{R}_s(k, l)$ and the noise correlation matrix $\mathbf{R}_n(k, \tilde{l})$ are estimated during periods of speech-plus-noise (l) and periods of noise-only (\tilde{l}), respectively. This means that,

- In theory the SDW-MWF could be an all zero vector during noise-only periods since then $\mathbf{R}_s(k, l) = 0$.
- In practice $\mathbf{R}_s(k, l)$ is "frozen" during noise-only periods where $\mathbf{R}_n(k, \tilde{l})$ is updated. In fact this is in line with the definition of $H_0(k, l)$ in (3.1), where the "0" indicates, that the speech X_i^s can have a non-zero $\mathbf{R}_s(k, l)$ in $H_0(k, l)$, but is not transmitted into X_i .

We then suggest, that if the $H_0(k, l)$ state and the $H_1(k, l)$ state can be properly detected a more flexible trade-off between noise reduction and speech distortion can be achieved.

To this aim, the parameter $P(l)$ is introduced, which is a binary decision, obtained by averaging the conditional SPP $p(k, l)$ over all frequency bins k

$$P(l) = \begin{cases} 1 & \text{if } \frac{1}{K} \sum_{k=1}^K p(k, l) \geq \alpha_{\text{frame}} \\ 0 & \text{otherwise} \end{cases} \quad (3.17)$$

where $P(l) = 1$ means the $H_1(k, l)$ state is detected and $P(l) = 0$ means the $H_0(k, l)$ state is detected, and α_{frame} is a detection threshold. This $P(l)$ will be used in the operation of SDW-MWF_{FLEX}. In Figure 3.3 $P(l)$ is plotted for a given speech segment which shows that even in the $H_1(k, l)$ state there are some frames/samples where the conditional SPP is low. Notice that in this case the noise correlation matrix is kept fixed whereas $p(k, l)$ and $P(l)$ are continuously updated. The two key ingredients of the proposed SDW-MWF_{FLEX} are now as follows:

- A weighting factor μ_{H_1} is introduced, which is a function of $p(k, l)$, and which defines the amount of noise reduction that can be applied in the $H_1(k, l)$ state.

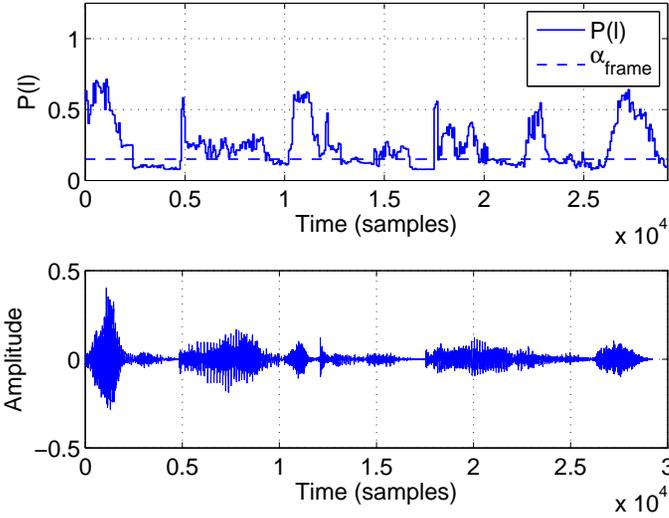


Figure 3.3: Illustration of $P(l)$ for a given speech segment.

- A weighting factor μ_{H_0} is introduced, which is a constant weighting factor, and which defines the amount of noise reduction that can be applied in the $H_0(k, l)$ state.

The SDW-MWF_{Flex} weighting strategy is illustrated in Figure 3.4 which shows the weighting factor as a function of $p(k, l)$. Notice that μ_{H_1} is defined here as $\min(\frac{1}{p(k, l)}, \alpha_{H_1})$, i.e., a function of the conditional SPP $\frac{1}{p(k, l)}$ and a lower threshold α_{H_1} , which is introduced since speech may not be present in all frequency bins even in state $H_1(k, l)$. The optimization criterion for SDW-MWF_{Flex} is given by

$$\begin{aligned}
 \mathbf{W}_{\text{MWF}_{\text{Flex}}}(k, l) &= \arg \min_{\mathbf{W}(k, l)} \\
 &P(l) \left[\max(p(k, l), \frac{1}{\alpha_{H_1}}) \varepsilon\{|X_1^s - \mathbf{W}^H(k, l)\mathbf{X}|^2 | H_1(k, l)\} \right. \\
 &\quad \left. + (1 - \max(p(k, l), \frac{1}{\alpha_{H_1}})) \varepsilon\{|\mathbf{W}^H(k, l)\mathbf{X}|^2 | H_0(k, l)\} \right] + \\
 &(1 - P(l)) \left[\frac{1}{\mu_{H_0}} \varepsilon\{|X_1^s - \mathbf{W}^H(k, l)\mathbf{X}^s|^2 | H_0(k, l)\} + \right. \\
 &\quad \left. \varepsilon\{|\mathbf{W}^H(k, l)\mathbf{X}^n|^2 | H_0(k, l)\} \right] \tag{3.18}
 \end{aligned}$$

where the first term ($P(l) = 1$) is equivalent to (3.13) where $p(k, l)$ is replaced with $\max(p(k, l), \frac{1}{\alpha_{H_1}})$ and the second term $(1 - P(l)) = 0$ is equivalent to (2.22)

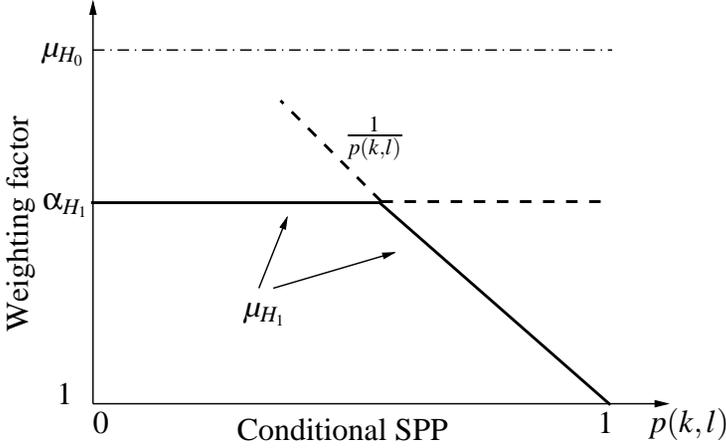


Figure 3.4: The weighting factor used in SDW-MWF_{Flex} as a function of the conditional SPP.

where μ is replaced with μ_{H_0} . The expression in (3.18) can be further simplified to

$$\begin{aligned}
 \mathbf{W}_{\text{MWF}_{\text{Flex}}}(k, l) &= \arg \min_{\mathbf{W}(k, l)} \\
 &\quad \left[P(l) \max(p(k, l), \frac{1}{\alpha_{H_1}}) + (1 - P(l)) \frac{1}{\mu_{H_0}} \right] \\
 &\quad \varepsilon \{ |X_1^s - \mathbf{W}^H(k, l) \mathbf{X}^s|^2 \} + \left[P(l) + (1 - P(l)) \right] \varepsilon \{ |\mathbf{W}^H(k, l) \mathbf{X}^n|^2 \} \\
 &= \arg \min_{\mathbf{W}(k, l)} \\
 &\quad \left[P(l) \max(p(k, l), \frac{1}{\alpha_{H_1}}) + (1 - P(l)) \frac{1}{\mu_{H_0}} \right] \\
 &\quad \varepsilon \{ |X_1^s - \mathbf{W}^H(k, l) \mathbf{X}^s|^2 \} + \varepsilon \{ |\mathbf{W}^H(k, l) \mathbf{X}^n|^2 \}. \tag{3.19}
 \end{aligned}$$

The SDW-MWF_{Flex} can then be written as

$$\mathbf{W}_{\text{MWF}_{\text{Flex}}}(k, l) = \left[\mathbf{R}_s(k, l) + \gamma(k, l) \mathbf{R}_n(k, l) \right]^{-1} \mathbf{R}_s(k, l) \mathbf{e}_1 \tag{3.20}$$

with the weighting factor defined as

$$\begin{aligned}\gamma(k, l) &= \left[P(l) \max\left(p(k, l), \frac{1}{\alpha_{H_1}}\right) + (1 - P(l)) \frac{1}{\mu_{H_0}} \right]^{-1} \\ &= \left[P(l) \min\left(\frac{1}{p(k, l)}, \alpha_{H_1}\right) + (1 - P(l)) \mu_{H_0} \right].\end{aligned}\quad (3.21)$$

The SDW-MWF_{Flex} is summarized in algorithm 2:

Algorithm 2 SDW-MWF_{Flex}

```

1: for each frame  $l$  and each frequency  $k$  do
2:   Estimate noise correlation matrix  $\mathbf{R}_n(k, l)$  (2.4)
3:   Estimate speech correlation matrix  $\mathbf{R}_s(k, l)$  (2.6)
4:   Estimate speech presence probability  $p(k, l)$  (3.4)
5:    $H_1(k, l)$  and  $H_0(k, l)$  detection  $P(l)$  (3.17)
6:   if  $P(l) = 1$  (per-frame decision) then
7:      $H_1$  state detected
8:     Estimate SDW-MWFFlex (3.20)
9:     with  $\gamma(k, l) = \min\left(\frac{1}{p(k, l)}, \alpha_{H_1}\right)$  (3.21)
10:  else
11:     $H_0$  state detected
12:    Estimate SDW-MWFFlex (3.20)
13:    with  $\gamma(k, l) = \mu_{H_0}$  (3.21)
14:  end if
15: end for

```

3.4 Rank-1 SDW-MWF incorporating the conditional SPP

In a similar manner to (2.26) the rank-1 SDW-MWF incorporating the conditional SPP can be written as

$$\mathbf{W}_{\text{R1-MWF}_{\text{SPP}}}(k, l) = \mathbf{R}_n^{-1}(k, l) \mathbf{R}_s(k, l) \mathbf{e}_1 \cdot \frac{1}{\frac{1}{p(k, l)} + \text{Tr}\{\mathbf{R}_n^{-1}(k, l) \mathbf{R}_s(k, l)\}} \quad (3.22)$$

which shows that the conditional SPP only has an influence on the single-channel postfilter. In order to show the characteristic of the single-channel postfilter (3.22) can be rewritten as [38]

$$\mathbf{W}_{\text{R1-MWF}_{\text{SPP}}}(k, l) = \frac{\mathbf{R}_n^{-1}(k, l) \mathbf{R}_s(k, l) \mathbf{e}_1}{1 + \text{SNR}_{\text{out}}} \cdot \frac{1 + \text{SNR}_{\text{out}}}{\frac{1}{p(k, l)} + \text{SNR}_{\text{out}}} \quad (3.23)$$

where $\text{SNR}_{\text{out}} = \text{Tr}\{\mathbf{R}_n^{-1}(k, l)\mathbf{R}_s(k, l)\}$ and the single-channel postfilter can then be written as,

$$G_{\text{post}}(k, l) = \frac{1 + \text{SNR}_{\text{out}}}{\frac{1}{p(k, l)} + \text{SNR}_{\text{out}}} \quad (3.24)$$

and by using the combined weighting factor in (3.16)

$$G_{\text{post}}(k, l) = \frac{1 + \text{SNR}_{\text{out}}}{\frac{1}{\alpha(1/\mu) + (1-\alpha)p(k, l)} + \text{SNR}_{\text{out}}} \quad (3.25)$$

where μ in this case is the constant attenuation factor, and α is a trade-off factor between the rank-1 SDW-MWF $_{\mu}$ and the rank-1 SDW-MWF $_{\text{SPP}}$. The characteristic of the single-channel postfilter in (3.25) is shown in Figure 3.5(a) and 3.5(b) for $\alpha = 0$ and $\alpha = 0.85$, respectively. It is clear that α defines how

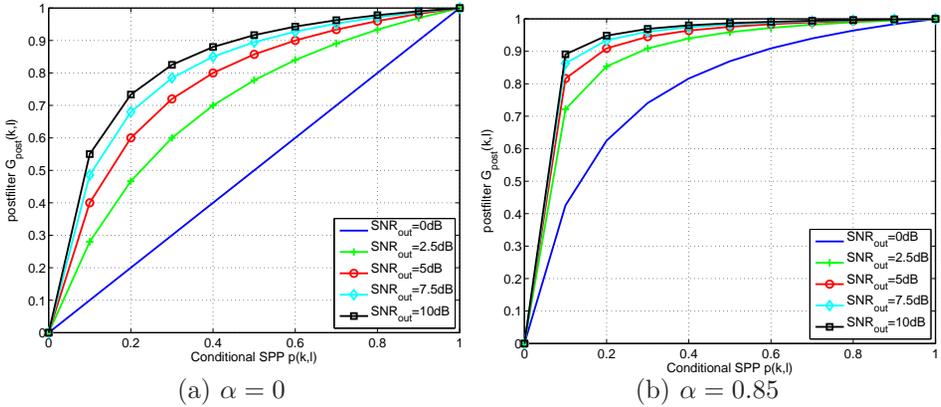


Figure 3.5: Characteristic of the single-channel postfilter incorporating the conditional SPP and α for various SNR_{out} .

aggressively the single-channel postfilter is allowed to behave especially at low output SNR and at low conditional SPP. This means that the trade-off between NR and signal distortion lies within the estimation of the conditional SPP and the value chosen for α .

In this work the focus is on the rank-1 SDW-MWF since this formulation has shown to be less sensitive to estimation errors in the correlation matrices compared to the standard SDW-MWF [38]. This makes the rank-1 SDW-MWF an interesting approach and therefore the target now is to combine the properties of the rank-1 SDW-MWF with a robust and a possibly more accurate method to estimate the correlation matrices, which will be addressed in Chapter 4.

3.5 Experimental results

3.5.1 Experimental set-up

The experimental set-up is similar as explained in Chapter 2. However experimental results are only presented for a high reverberation scenario. For the correlation matrices used to estimate the conditional SPP the forgetting factors are set to $\bar{\alpha}_n = \bar{\alpha}_x = 0.96$ in order to track the spectral non-stationarities. The forgetting factors used to estimate the correlation matrices for the MWF were found to be varying too slowly to track the spectral information.

Table 3.1 shows the parameters used in the estimation of the conditional SPP.

$\omega_{local} = 1$	$\zeta_{min} = -10\text{dB} (0.1)$	$\zeta_{p\ min} = 4\text{dB}$
$\omega_{global} = 10$	$\zeta_{max} = -5\text{dB} (0.3162)$	$\zeta_{p\ max} = 10\text{dB}$

Table 3.1: Parameters used in the estimation of the conditional SPP

3.5.2 Results

Simulation results for the SDW-MWF_{SPP} and the SDW-MWF_{Flex} compared to the SDW-MWF _{μ} for a high reverberation scenario are shown in Figure 3.6. Overall the results show that the SDW-MWF_{SPP} and the SDW-MWF_{Flex} are able to outperform the SDW-MWF _{μ} for $\mu=3$ and 5 both in terms of SNR improvement and signal distortion. The SDW-MWF_{Flex} shows the best SNR improvement but the signal distortion is also slightly higher compared to the SDW-MWF_{SPP}. The performance for SDW-MWF_{SPP} with $\alpha=0.75$ consistently shows a greater SNR improvement compared to SDW-MWF _{μ} with $\mu=1$ and at the same time the signal distortion is almost similar. Even though the signal distortion is lower for SDW-MWF_{SPP} and the SDW-MWF_{Flex} the problem remains the same, i.e., any improvement in the SNR comes at the cost of a higher signal distortion which is highly undesirable.

Simulation results for the rank-1 SDW-MWF_{SPP} and the rank-1 SDW-MWF_{Flex} compared to the rank-1 SDW-MWF _{μ} for a high reverberation scenario are shown in Figure 3.7. In this case the rank-1 SDW-MWF_{SPP} and the rank-1 SDW-MWF_{Flex} can increase the SNR improvement compared to the rank-1 SDW-MWF _{μ} but this does not come without an increase in signal distortion. This suggests, that with the rank-1 SDW-MWF _{μ} the optimal performance has been reached and applying the proposed weighting factor, that is updated for each frequency and for each frame, in the single-channel postfilter only leads to increased signal distortion. A possible explanation could be that the weighting factor $\frac{1}{p(k,l)}$ and the $\text{SNR}_{\text{out}} =$

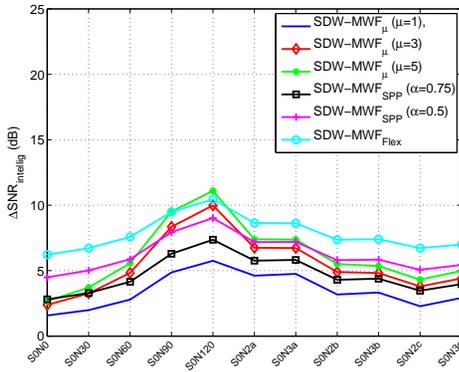
$\text{Tr}\{\mathbf{R}_n^{-1}(k, l)\mathbf{R}_s(k, l)\}$ varies with different dynamic which negatively affects the single-channel postfilter.

3.6 Conclusion

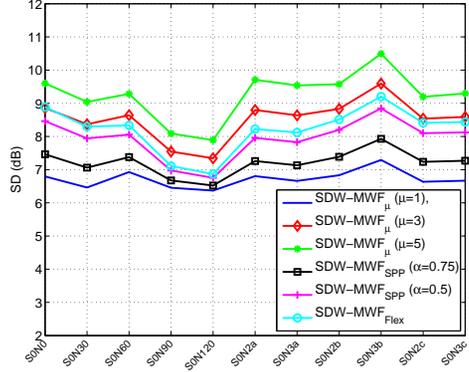
In this chapter we have introduced an SDW-MWF-based NR that incorporates the conditional SPP, referred to as SDW-MWF_{SPP}, such that the weighting factor is now updated for each frequency and for each frame. In addition to SDW-MWF_{SPP} we have also presented an SDW-MWF based NR that incorporates a flexible weighting factor, referred to as SDW-MWF_{Flex} which is based on combining the conditional SPP with a per frame based $H_0(k, l)$ and $H_1(k, l)$ state detection. This is based on the observation that the noise in the $H_0(k, l)$ and the $H_1(k, l)$ can be reduced with different objectives.

Experimental results show that the SDW-MWF_{SPP} and the SDW-MWF_{Flex} outperform the SDW-MWF _{μ} both in terms of SNR improvement and signal distortion. However, for the rank-1 SDW-MWF_{SPP} and the rank-1 SDW-MWF_{Flex} the SNR improvement comes at the cost of a higher signal distortion compared to the rank-1 SDW-MWF _{μ} . This could indicate that the rank-1 SDW-MWF _{μ} already has reached the optimal performance or that the weighting factor that is updated for each frequency and for each frame is limited by the slow time-variation of the correlation matrices.

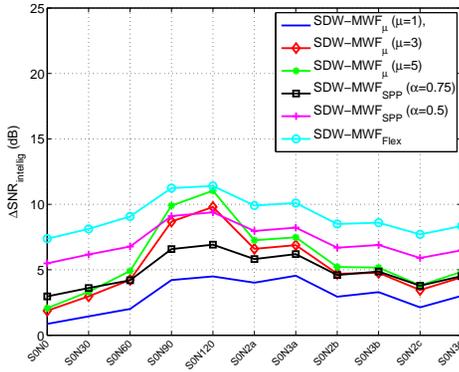
Allowing the weighting factor to be updated for each frequency and for each frame has shown to improve the NR performance for certain scenarios. However the estimation and the update of the correlation matrices are still an open problem even when the weighting factor is updated more frequently, the correlation matrices are still updated less frequently. This could be the reason that the rank-1 SDW-MWF_{SPP} and the rank-1 SDW-MWF_{Flex} perform worse than the rank-1 SDW-MWF _{μ} , due to the mismatch in the dynamics of the single-channel postfilter.



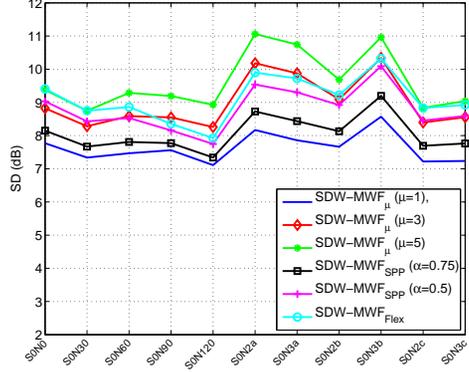
(a) Input SNR=0dB



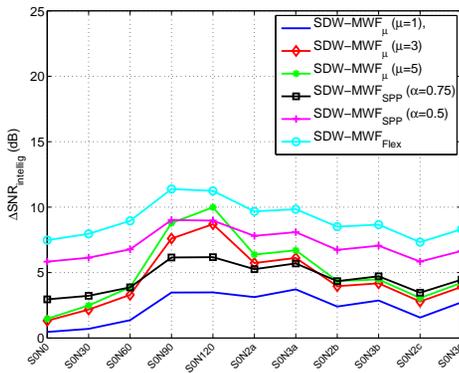
(b) Input SNR=0dB



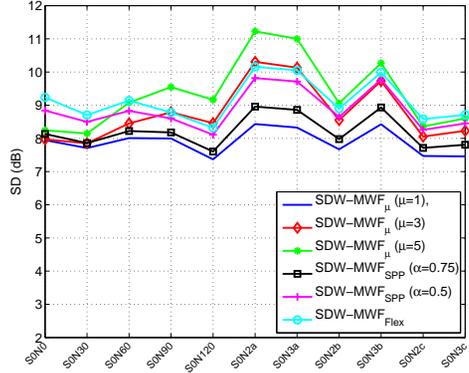
(c) Input SNR=-5dB



(d) Input SNR=-5dB

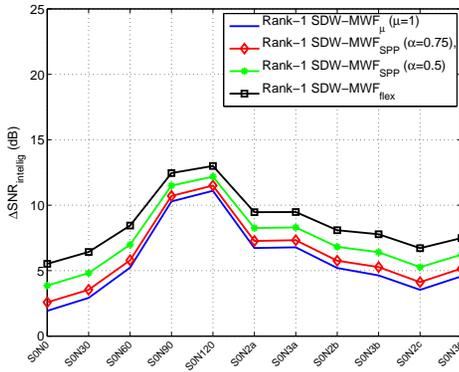


(e) Input SNR=-7.5dB

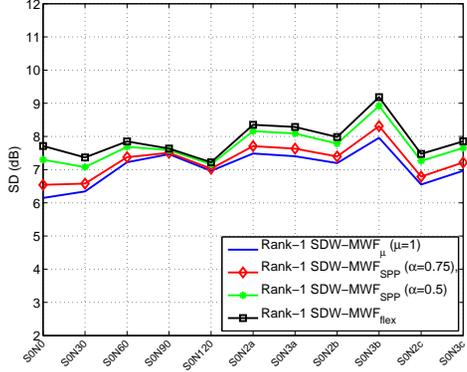


(f) Input SNR=-7.5dB

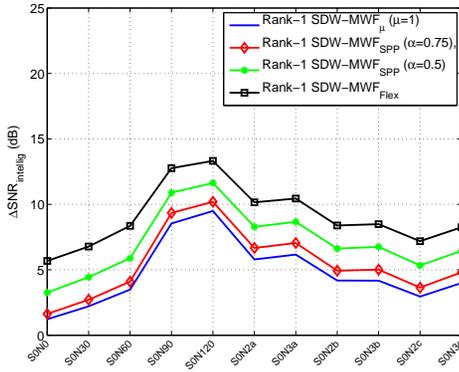
Figure 3.6: Comparison of SDW-MWF_μ to SDW-MWF_{SPP} and SDW-MWF_{Flex} a high reverberation scenario using objective measures.



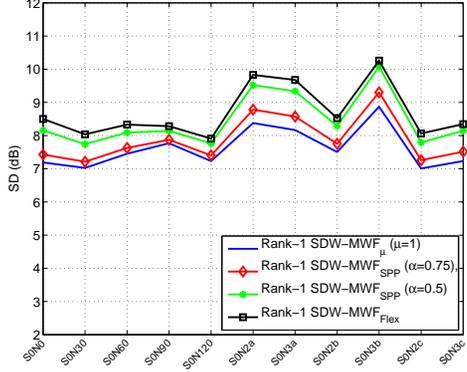
(a) Input SNR=0dB



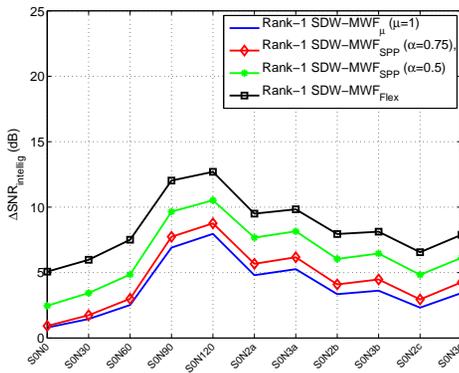
(b) Input SNR=0dB



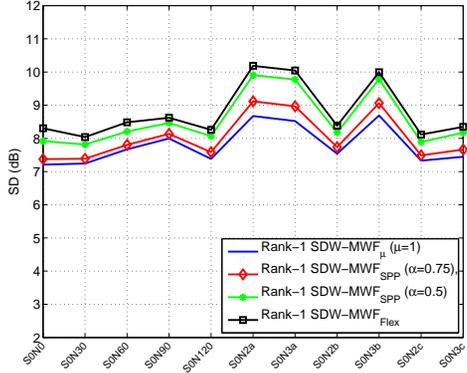
(c) Input SNR=-5dB



(d) Input SNR=-5dB



(e) Input SNR=-7.5dB



(f) Input SNR=-7.5dB

Figure 3.7: Comparison of rank-1 SDW-MWF_μ to rank-1 SDW-MWF_{SPP} and rank-1 SDW-MWF_{flex} for a high reverberation scenario using objective measures.

Chapter 4

SDW-MWF _{μ} based on robust estimation of the correlation matrices

This chapter addresses the issue of using correlation matrices that are kept fixed during speech-plus-noise periods and are updated during noise-only periods or vice versa. As mentioned this can limit the tracking both spectrally and spatially. The robustness of the correlation matrices can also be compromised especially when estimating the clean speech correlation matrix which requires an accurate estimation of the noise-only correlation matrix.

For this reason we once again turn our attention to single-channel NR where several attempts have been made to continuously track and update the noise power during periods of speech-plus-noise [133]. One of the interesting approaches is referred to as the improved minima controlled recursive averaging (IMCRA) noise estimation approach where the conditional SPP is used as a time-varying smoothing factor [35][36]. Inspired by this approach we propose a robust method to estimate and update the correlation matrices that exploits prior knowledge of the correlation matrices combined with a continuous update approach based on the conditional SPP [33][198].

Section 4.1 presents the proposed method to estimate and update the correlation matrices in a robust way. The approach is based on using the conditional SPP combined with prior knowledge of the correlation matrices.

Section 4.2 analyzes the performance of the proposed method to estimate and update the correlation matrices compared to the traditional method using a perfect

VAD.

Section 4.3 presents the experimental results using the SDW-MWF based NR introduced in Chapter 2 and 3 to verify the properties of the proposed correlation matrices.

4.1 Robust estimation of the correlation matrices

4.1.1 Uncertainty of the correlation matrices

The estimation of the correlation matrices is bound to introduce some errors due to, e.g., VAD errors, non-stationary noise, low input SNR, reverberation, averaging time etc. which can have a great influence on the MWF [38]. An error in the estimated correlation matrices can be written as

$$\hat{\mathbf{R}}_n(k, l) = \mathbf{R}_n(k, l) + \Delta_{\text{error}}^n \quad (4.1)$$

and

$$\hat{\mathbf{R}}_x(k, l) = \mathbf{R}_x(k, l) + \Delta_{\text{error}}^x \quad (4.2)$$

where $\mathbf{R}_n(k, l)$ and $\mathbf{R}_x(k, l)$ are the true (unknown) noise-only and speech-plus-noise correlation matrices, respectively, and Δ_{error}^n and Δ_{error}^x represent the error (also unknown) in the estimated correlation matrices. Obviously, the important factor is to compensate for the error introduced in (4.1) and (4.2). For this purpose, we propose that the correlation matrices are updated continuously, i.e., the $\hat{\mathbf{R}}_n(k, l)$ and $\hat{\mathbf{R}}_x(k, l)$ have to be estimated in both the $H_0(k, l)$ and $H_1(k, l)$ state. To achieve this we will introduce the use of prior knowledge of the correlation matrices. This is motivated by the fact that the correlation matrices used in (2.8) and (2.9) are not guaranteed to be valid from one state to another, and in that sense we can simply replace those correlation matrices with any prior knowledge that we have regarding the correlation matrices. Then (4.1) and (4.2) can be rewritten as

$$\mathbf{R}_n(k, l) = \bar{\mathbf{R}}_n(k, l) + \Delta_{\text{correction}}^n \quad (4.3)$$

and

$$\mathbf{R}_x(k, l) = \bar{\mathbf{R}}_x(k, l) + \Delta_{\text{correction}}^x \quad (4.4)$$

where $\bar{\mathbf{R}}_n(k, l)$ and $\bar{\mathbf{R}}_x(k, l)$ are the prior knowledge of the noise-only and speech-plus-noise correlation matrices, respectively. The term $\Delta_{\text{correction}}^n$ and $\Delta_{\text{correction}}^x$ can be considered as the parameters that compensate for the mismatch between the true and the prior correlation matrices. It is clear that the challenge is to select the prior correlation matrices and estimating $\Delta_{\text{correction}}^n$ and $\Delta_{\text{correction}}^x$ which will be explained in the following.

4.1.2 Continuous updating of the correlation matrices

To achieve a robust and accurate estimation of the correlation matrices we propose to combine the prior knowledge of the correlation matrices with a continuous updating approach exploiting the conditional SPP. The proposed noise correlation matrix estimation can then be written as

$$\begin{aligned}\hat{\mathbf{R}}_n(k, l + 1) &= \underbrace{p(k, l)\bar{\mathbf{R}}_n(k, l)}_{\text{prior}} + \underbrace{(1 - p(k, l))\left[\alpha_n\hat{\mathbf{R}}_n(k, l) + (1 - \alpha_n)\mathbf{X}(k, l)\mathbf{X}^H(k, l)\right]}_{\text{update}} \\ &= p(k, l)\bar{\mathbf{R}}_n(k, l) + (1 - p(k, l))\Delta_n(k, l).\end{aligned}\quad (4.5)$$

If the conditional SPP indicates a high probability of speech presence then greater weights are given to the prior knowledge $\bar{\mathbf{R}}_n(k, l)$. If the conditional SPP indicates a high probability of speech absence then greater weights are given to the instantaneous update. In a similar manner the speech-plus-noise correlation matrix can be estimated as

$$\begin{aligned}\hat{\mathbf{R}}_x(k, l + 1) &= \underbrace{(1 - p(k, l))\bar{\mathbf{R}}_x(k, l)}_{\text{prior}} + \underbrace{p(k, l)\left[\alpha_x\hat{\mathbf{R}}_x(k, l) + (1 - \alpha_x)\mathbf{X}(k, l)\mathbf{X}^H(k, l)\right]}_{\text{update}} \\ &= (1 - p(k, l))\bar{\mathbf{R}}_x(k, l) + p(k, l)\Delta_x(k, l).\end{aligned}\quad (4.6)$$

In [35][36][199] the $\bar{\mathbf{R}}_n(k, l)$ and the $\bar{\mathbf{R}}_x(k, l)$ are replaced by the previous estimate of the correlation matrices. Estimation of the correlation matrices with prior knowledge ensures that the estimated correlation matrices always have a certain structure which also makes the corresponding filter valid. The advantage of the continuous update is that during noise-only periods the $\hat{\mathbf{R}}_n(k, l)$ is updated with greater weight but more importantly the noise-level in $\hat{\mathbf{R}}_x(k, l)$ is also updated of course with less weight. Then during speech-plus-noise periods greater weight is on $\hat{\mathbf{R}}_x(k, l)$ which means that the noise-level in $\hat{\mathbf{R}}_x(k, l)$ and $\hat{\mathbf{R}}_n(k, l)$ should be better matched. It should also be emphasized that the conditional SPP is estimated for each frequency and for each frame such that certain frequencies can be updated more frequently than others. This can be a benefit if the speech and the noise are present at distinct frequencies. This problem is related to e.g. spectral subtraction based NR where an inaccurate estimation of the noise can lead to severe distortion [15][11][187].

Selecting the prior knowledge of the correlation matrices and estimating the conditional SPP is always associated with some errors and uncertainty. Therefore, it is desirable to always have the influence of the prior and the update during the estimation which can be ensured by constraining the conditional SPP to be $0 < p(k, l) < 1$. Excluding the scenarios where the speech is present ($p(k, l)=1$) or absent ($p(k, l)=0$) makes sense since in these cases the influence of $p(k, l)$ is

removed and correlation matrices are not continuously updated anymore. Actually when $(p(k, l)=0)$ the $\bar{\mathbf{R}}_x(k, l+1)$ in (2.8) is replaced with $\bar{\mathbf{R}}_x(k, l)$ and when $(p(k, l)=1)$ the $\bar{\mathbf{R}}_n(k, l+1)$ in (2.9) is replaced with $\bar{\mathbf{R}}_n(k, l)$.

4.1.3 Selection of prior correlation matrices

We propose to estimate the prior correlation matrices $\bar{\mathbf{R}}_x(k, l)$ and $\bar{\mathbf{R}}_n(k, l)$ using a batch procedure, i.e., the correlation matrices are estimated off-line and kept fixed during the NR process. The prior correlation matrices can be estimated as

$$H_0(k, l) : \bar{\mathbf{R}}_n(k, l+1) = \alpha_n \bar{\mathbf{R}}_n(k, l) + (1 - \alpha_n) \bar{\mathbf{X}}(k, l) \bar{\mathbf{X}}^H(k, l) \quad (4.7)$$

and

$$H_1(k, l) : \bar{\mathbf{R}}_x(k, l+1) = \alpha_x \bar{\mathbf{R}}_x(k, l) + (1 - \alpha_x) \bar{\mathbf{X}}(k, l) \bar{\mathbf{X}}^H(k, l) \quad (4.8)$$

where $\bar{\mathbf{X}}(k, l)$ is the signal used to estimate the prior correlation matrices. Since the correlation matrices contain both spectral and spatial signal characteristics the choice of $\bar{\mathbf{X}}(k, l)$ can result in both spectral and spatial mismatch. This is to be expected since the input SNRs and the spatial scenarios are typically unknown. Therefore, a spectral mismatch refers to a SNR mismatch between $\bar{\mathbf{X}}(k, l)$ and the actual signal $\mathbf{X}(k, l)$. For a typical hearing aid scenario we can assume that the desired speaker is located in the front of the hearing aid user [45][109], i.e., at an angle corresponding to 0° whereas the noise can be located at any spatial angle and can change over time. Therefore a spatial difference in the noise angles, between $\bar{\mathbf{X}}(k, l)$ and $\mathbf{X}(k, l)$ is referred to as a spatial mismatch.

The spectral mismatch of the prior is assumed to be more crucial than the spatial mismatch. This observation is related to the way that the speech correlation matrix is estimated. As mentioned in Chapter 2 the subtraction in (2.27) can lead to a poor estimate of the speech correlation matrix. To avoid this it is important that the SNR of $\bar{\mathbf{X}}(k, l)$ is not chosen too low, but if the SNR is chosen too high the NR may be compromised. However, this is where the expression in (4.5) and (4.6) is supposed to compensate for any mismatch between the prior and the true (unknown) correlation matrices defined in (4.3) and (4.4). On the other hand, if this compensation is not sufficient the single-channel postfilter defined in (3.22) can also compensate for any lack of spectral filtering by using the conditional SPP and choosing a proper value for α . This suggests, that the SNR of $\bar{\mathbf{X}}(k, l)$ can be selected higher than the actual input SNR in order to avoid the subtraction in (2.27) leading to a poor estimate.

4.2 Analysis of estimation errors

In this section we will compare the proposed estimation and update of the correlation matrices using prior knowledge and the conditional SPP with the traditional methods using a perfect VAD. As mentioned previously the main problem is the estimation of the clean speech correlation matrix, i.e.,

$$\hat{\mathbf{R}}_s(k, l) = \hat{\mathbf{R}}_x(k, l) - \hat{\mathbf{R}}_n(k, l). \quad (4.9)$$

In this analysis we are particularly interested in the spectral content of the estimated correlation matrices in (4.9). For this purpose we define the power of each correlation matrix as

$$\hat{P}_s(k, l) = \text{Tr}\{\hat{\mathbf{R}}_s(k, l)\} \quad (4.10)$$

$$\hat{P}_x(k, l) = \text{Tr}\{\hat{\mathbf{R}}_x(k, l)\} \quad (4.11)$$

$$\hat{P}_n(k, l) = \text{Tr}\{\hat{\mathbf{R}}_n(k, l)\} \quad (4.12)$$

In the following we will illustrate this problem by a series of examples. The estimation of $\hat{\mathbf{R}}_x(k, l)$ and $\hat{\mathbf{R}}_n(k, l)$, or more specifically $\hat{P}_x(k, l)$ and $\hat{P}_n(k, l)$, using a perfect VAD, based on a 0dB input SNR signal, is shown in Figure 4.1(a)-(d). These plots show the power of the correlation matrices as a function of the speech frames for selected frequencies. The first observation is that the noise-only correlation matrix is kept fixed during speech-plus-noise frames but more importantly most of the noise power $\hat{P}_n(k, l)$ is actually higher than the speech-plus-noise power $\hat{P}_x(k, l)$. The consequence of this is shown in Figure 4.1(e)-(f) where $\hat{P}_s(k, l)$ is estimated which shows that due to estimation errors the power of the estimated clean speech results in negative values. This will have a negative impact if these correlation matrices are used to form the MWF.

The same experiments are conducted for the proposed estimation of the correlation matrices (4.5)-(4.6) which is shown in Figure 4.2(a)-(f). It should also be mentioned that the SNR of $\bar{\mathbf{X}}(k, l)$ in (4.7)-(4.8) is set 5dB higher than the actual input SNR which is 0dB. A first observation is that the speech-plus-noise correlation matrices have a distinct shape, which is probably caused by the conditional SPP and the fact that a continuous update approach is used. This can be compared to the scenario with the perfect VAD which overall has a more monotonous shape for different frequencies across different speech frames. Another clear advantage is shown with the estimated noise-only correlation matrix which is now continuously updated and it is clear that the proposed estimation technique is able to track non-stationarity of the noise. The most important part here is the estimation of the clean speech correlation matrices and here Figure 4.2(e)-(f) clearly shows that the estimated $\hat{P}_s(k, l)$ does not result in negative values which is highly desirable.

The final design of the proposed SDW-MWF based NR is shown in Figure 4.3. Compared to the block diagrams shown in Figure 2.3 the binary VAD, that is typically used in a MWF based NR [49][207], is now removed since the correlation matrices are now jointly estimated. The estimation of the conditional SPP remains the same as in Figure 3.1 but the difference is that the conditional SPP is now also used in the estimation and in the update of the correlation matrices. As a new ingredient to the SDW-MWF based NR the use of the prior knowledge of the correlation matrices is introduced which allows for the continuous updating approach.

4.3 Experimental results

In this section, experimental results for the rank-1 SDW-MWF using the proposed robust estimation of the correlation matrices are presented and compared to a rank-1 SDW-MWF using the traditional method to estimate the correlation matrices based on a perfect VAD. As mentioned the rank-1 SDW-MWF has been selected since this formulation has shown to be less sensitive to estimation errors [38]. Therefore the proposed correlation matrices are also verified for the traditional formulation of the SDW-MWF [49][207][161].

4.3.1 Experimental set-up

The same simulation set-up is used as in Chapter 2 but only experiments with high reverberation and low input SNR are presented here. For the conditional SPP the same parameters are used as in Chapter 3. This means that only the selection of the prior correlation matrices has to be defined. As defined in (4.7)-(4.8) the prior correlation matrices are estimated based on $\bar{\mathbf{X}}(k, l)$. Since the SNR and the spatial scenarios are unknown the input SNR and spatial angles are selected differently compared to the actual test-setup, which as previously defined referred to as spectral and spatial mismatch. It should also be mentioned that the prior correlation matrices are estimated using different signals than the one used in the actual experiments. The spatial mismatch scenarios are shown in Table 4.1.

4.3.2 Results

In this experiment, the SNR of $\bar{\mathbf{X}}(k, l)$ is varied from -7.5dB to 7.5dB using the spatial scenarios defined in Table 2.1, i.e., no spatial mismatch is introduced between $\bar{\mathbf{X}}(k, l)$ and $\mathbf{X}(k, l)$. The effect of the spectral mismatch is shown in Figure 4.4(a)-(b). As a first observation it is clear that the rank-1 SDW-MWF_μ with the perfect VAD performs well for certain spatial scenarios such as S0N90

Case	Spatial mismatch 1	Spatial mismatch 2
S0N0	noise source(s) at 30°	noise source(s) at 60°
S0N30	noise source(s) at 90°	noise source(s) at 90°, 180°, 270°
S0N60	noise source(s) at 30°	noise source(s) at 90°, 180°
S0N90	noise source(s) at 30°, 60°	noise source(s) at 45°, 90°, 180°
S0N120	noise source(s) at 30°, 60°	noise source(s) at 45°, 90°, 180°
S0N2a	noise source(s) at 30°, 60°	noise source(s) at 120°
S0N3a	noise source(s) at 30°, 60°	noise source(s) at 45°, 90°
S0N2b	noise source(s) at 30°, 60°	noise source(s) at 120°
S0N3b	noise source(s) at 30°	noise source(s) at 90°
S0N2c	noise source(s) at 0°	noise source(s) at 120°
S0N3c	noise source(s) at 90°, 180°	noise source(s) at 120°

Table 4.1: Spatial mismatch compared to the actual spatial scenarios

and S0N120. The same trend is observed for rank-1 SDW-MWF $_{\mu}$ for the cases where the SNR of $\bar{\mathbf{X}}(k, l)$ is set to 7.5dB and 5dB but in these cases the SNR improvement is greater and the signal distortion is lower compared to the case with the perfect VAD. On the other hand, when the SNR of $\bar{\mathbf{X}}(k, l)$ to -7.5dB and -5dB the SNR improvement is poor even at S0N90 and S0N120 whereas spatial scenarios such as S0N0, S0N30, and S0N60 show an SNR improvement but this comes at the cost of a higher signal distortion. This suggests, that the noise has been over-estimated when the SNR of $\bar{\mathbf{X}}(k, l)$ is close to the input SNR and this results in a poor estimate of the speech correlation matrix. Overall, the prior correlation matrices with 7.5dB show the best performance.

In the next experiment, the SNR of $\bar{\mathbf{X}}(k, l)$ is fixed at 7.5dB since this shows the best SNR improvement with the lowest signal distortion. In order to show the influence of the single-channel postfilter α in the rank-1 SDW-MWF $_{\text{combined}}$ is varied from 1 to 0.5, 0.3, 0, and compared to rank-1 SDW-MWF $_{\mu}$ using a perfect VAD which is shown in Figure 4.4(c)-(d). A remarkable SNR improvement is observed with $\alpha=0$ for S0N90, S0N120, S0N2a, and S0N3a and impressively the signal distortion is lower compared to the performance using a perfect VAD. For other spatial scenarios the SNR improvement results in greater signal distortion but this can be avoided if α is set to 0.1. This suggests, that since the SNR of $\bar{\mathbf{X}}(k, l)$ is much higher than the actual input SNR the noise correlation matrix may not indicate the true noise-level and in such case the single-channel postfilter is able to compensate resulting in an overall SNR improvement. This clearly shows, that if the SNR of $\bar{\mathbf{X}}(k, l)$ is close to the actual input SNR the spatial filter can be negatively affected and in this case the single-channel postfilter will only make the performance worse.

In this experiment, the SNR of $\bar{\mathbf{X}}(k, l)$ is fixed at 7.5dB which is based on the

observation made above. The effect of the spatial mismatch for the rank-1 SDW-MWF _{μ} is shown in Fig. 4.5(a)-(b). It is clear that the spatial mismatch has an effect on the SNR improvement especially for spatial mismatch case 1, where the assumed location of the noise sources are chosen closer to the speech location. For spatial mismatch case 2 the SNR improvement is still better with a lower signal distortion compared to the performance with a perfect VAD. The effect of the spatial mismatch is also evaluated for the rank-1 SDW-MWF_{combined} when α is set to 0 and 0.1 which is shown Fig. 4.5(c)-(d). If the desire is to keep the signal distortion lower than in the case with a perfect VAD using the rank-1 SDW-MWF _{μ} , then α should be set to 0.1. Again it is observed that the SNR improvement for certain spatial scenarios results in an increased signal distortion especially when $\alpha=0$. However, it is clear that the proposed estimation of the correlation matrices with the rank-1 SDW-MWF_{combined} is able to outperform the rank-1 SDW-MWF _{μ} using a perfect VAD.

Simulation results for the traditional SDW-MWF _{μ} using the robust estimation of the correlation matrices is also compared to the rank-1 SDW-MWF _{μ} using a perfect VAD to estimate and update the correlation matrices, which is shown in Figure 4.6. Since the traditional SDW-MWF _{μ} jointly applies the spatial filter and the spectral filter it is clear that when $\mu = 1$ the SNR improvement is very low. However with $\mu = 3$ and 5 the SDW-MWF _{μ} outperforms the rank-1 SDW-MWF _{μ} both in terms of SNR improvement and signal distortion.

Simulation results for the SDW-MWF _{μ} with the robust estimation of the correlation matrices are also evaluated for different SNR of $\bar{\mathbf{X}}(k, l)$ and with spatial mismatch, which is shown in Figure 4.7. Again it is clear that a high SNR of $\bar{\mathbf{X}}(k, l)$ performs better and with $\mu = 5$ the SDW-MWF _{μ} still outperforms the rank-1 SDW-MWF _{μ} . The same trend is observed for the spatial mismatch cases.

4.4 Conclusion

In this chapter we have introduced an SDW-MWF-based NR that incorporates a robust method to estimate and update the correlation matrices. The robust estimation of the correlation matrices is based on introducing prior knowledge of the correlation matrices together with a continuous updating approach based on the conditional SPP. Combining this method to estimate and update the correlation matrices with a weighting factor to trade-off between NR and speech distortion that also varies for each frequency and for each frame, resulted in a novel SDW-MWF based NR that improves the robustness and the tracking capabilities.

Experimental results show that the proposed algorithm improves the SNR and the signal distortion compared to the traditional method with a perfect VAD used to estimate and update the correlation matrices for both the SDW-MWF

and the rank-1 SDW-MWF approaches. Analysis has shown that the estimated correlation matrices using a perfect VAD results in negative power in the estimated speech correlation matrix which in practice should not happen since the speech correlation matrix is estimated by subtracting the noise-only correlation matrix from the speech-plus-noise correlation matrix.

The SDW-MWF based NR proposed here has solved the problems of estimating and updating the correlation matrices in a robust way such that the speech correlation matrix can be reliably estimated. This is achieved by continuously estimating the noise-level in the speech-plus-noise and the noise-only correlation matrix during both speech-plus-noise and noise-only periods. It has also been shown how the conditional SPP can be used to further improve the single-channel postfilter by exploiting the proposed correlation matrices. Furthermore, the proposed correlation matrices also alleviate the sensitivity of the traditional SDW-MWF which was the reason to use the rank-1 SDW-MWF in the first place.

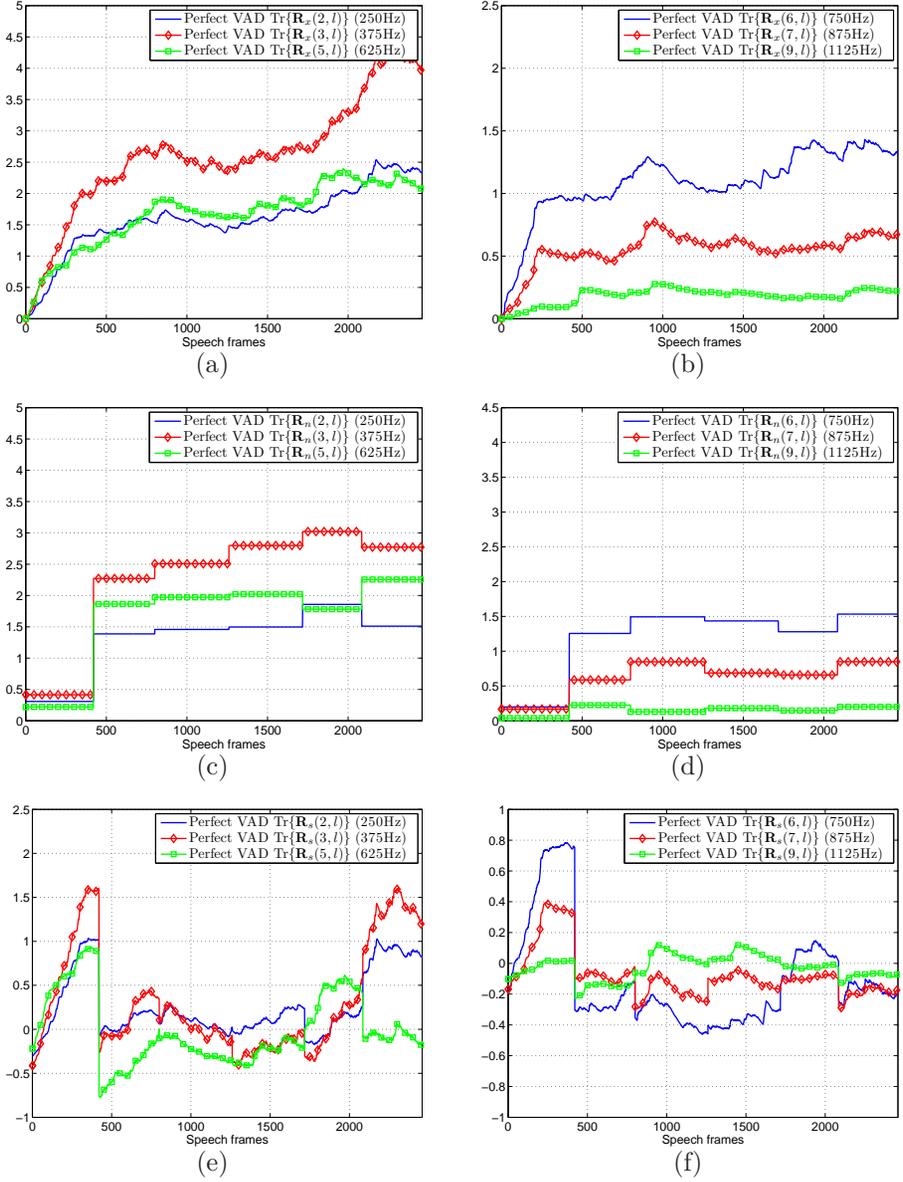
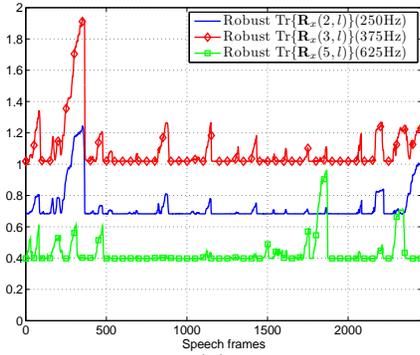
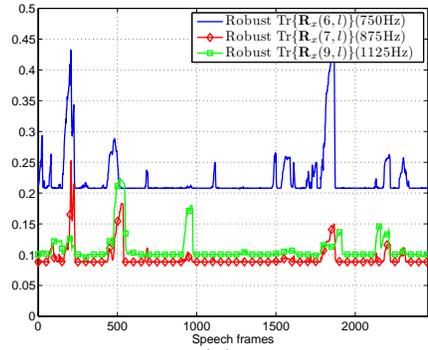


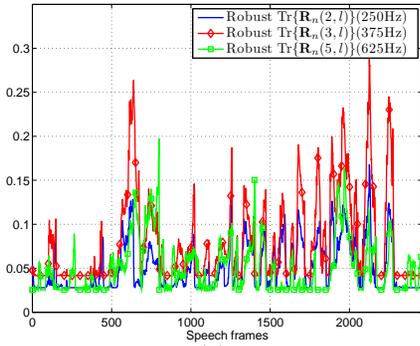
Figure 4.1: Illustration of the estimation errors in the correlation matrices using a traditional perfect VAD approach.



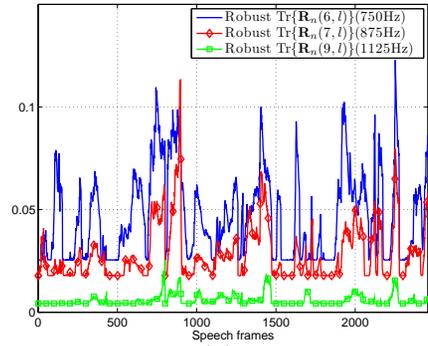
(a)



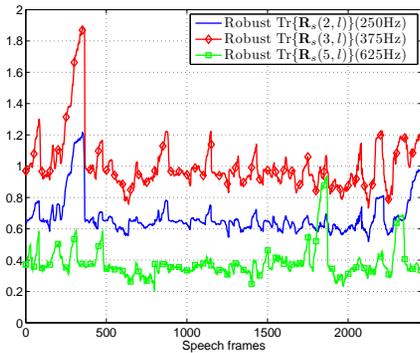
(b)



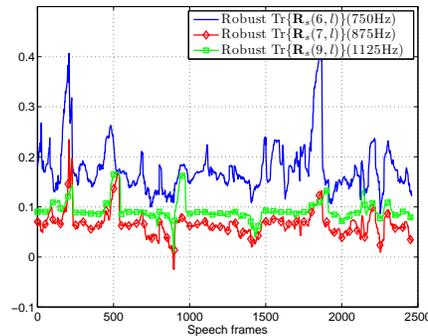
(c)



(d)



(e)



(f)

Figure 4.2: Illustration of the estimation errors in the correlation matrices using the proposed method.

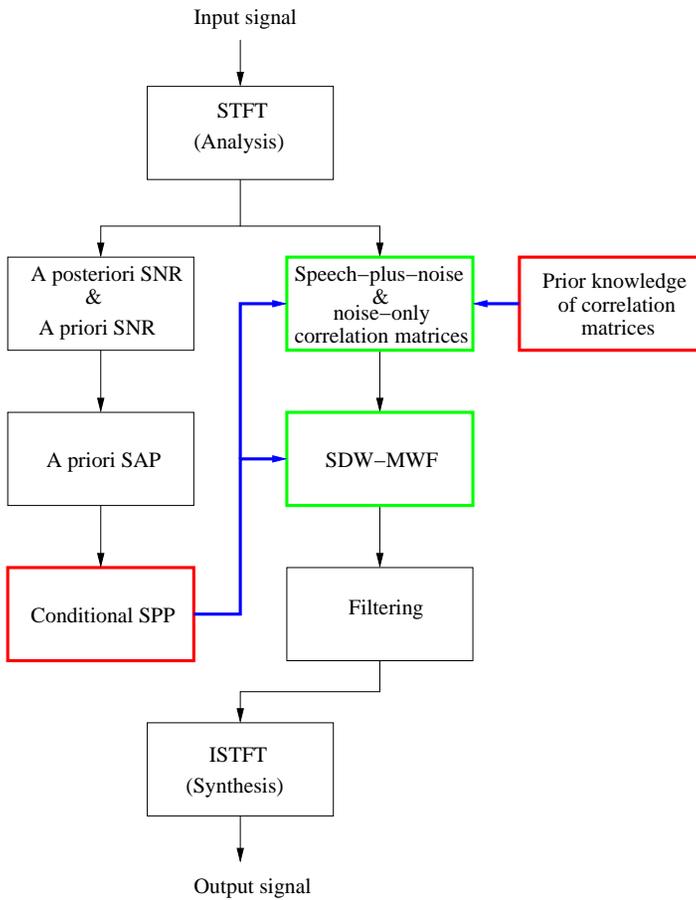


Figure 4.3: Block diagram of the SDW-MWF incorporating the proposed estimation of the correlation matrices.

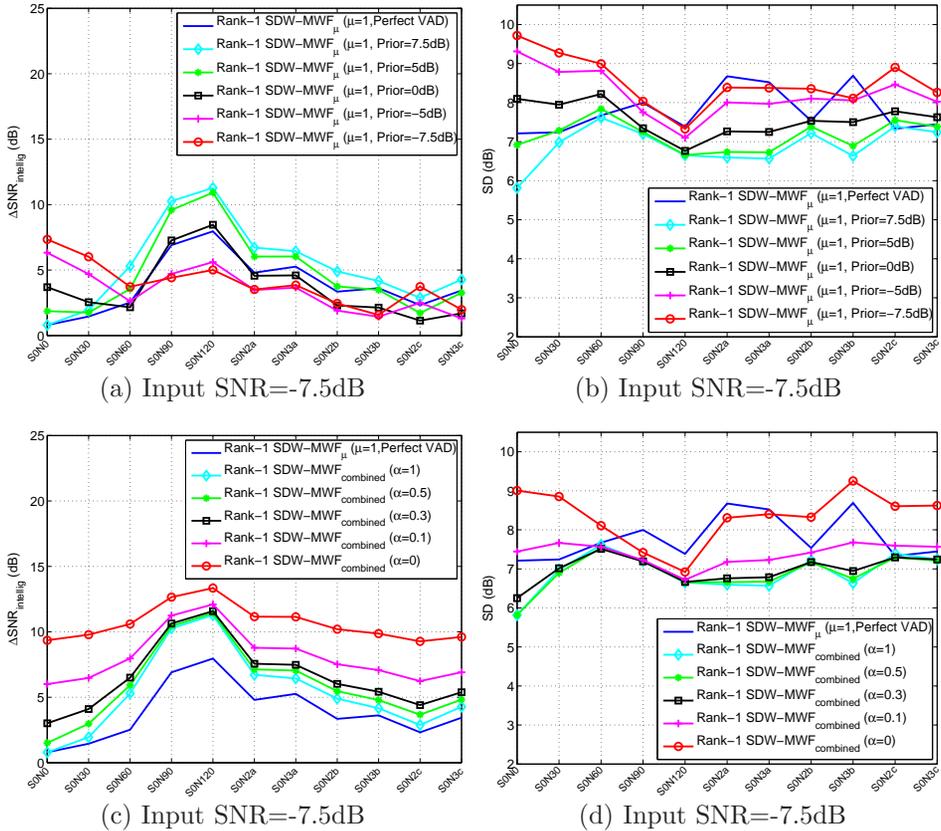


Figure 4.4: (a)-(b) SNR improvement and SD using a rank-1 SDW-MWF $_{\mu}$ for scenarios where the SNR of the priors are varied. (c)-(d) SNR improvement and SD using rank-1 SDW-MWF $_{\text{combined}}$ for scenarios where α is varied and the prior is fixed at SNR=7.5dB.

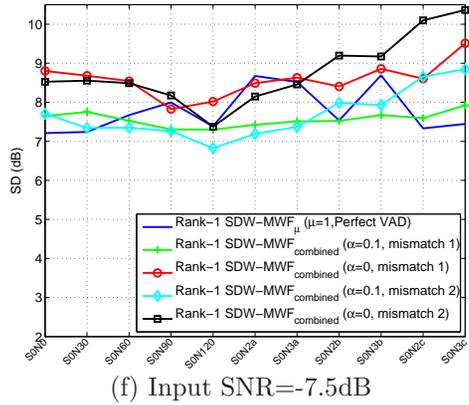
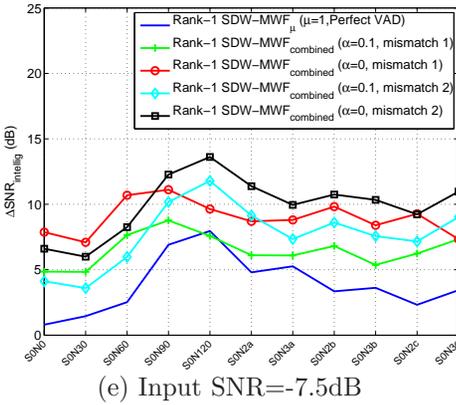
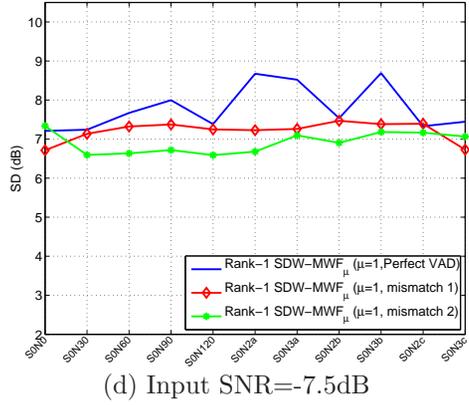
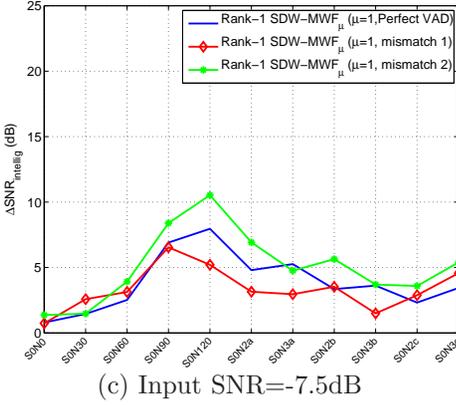


Figure 4.5: (a)-(b) SNR improvement and SD using a rank-1 SDW-MWF_μ for scenarios where spatial mismatch is introduced and the prior is fixed at SNR=7.5dB. (c)-(d) SNR improvement and SD using a rank-1 SDW-MWF_{combined} for scenarios where spatial mismatch is introduced and the α is set to 0.1 and 0

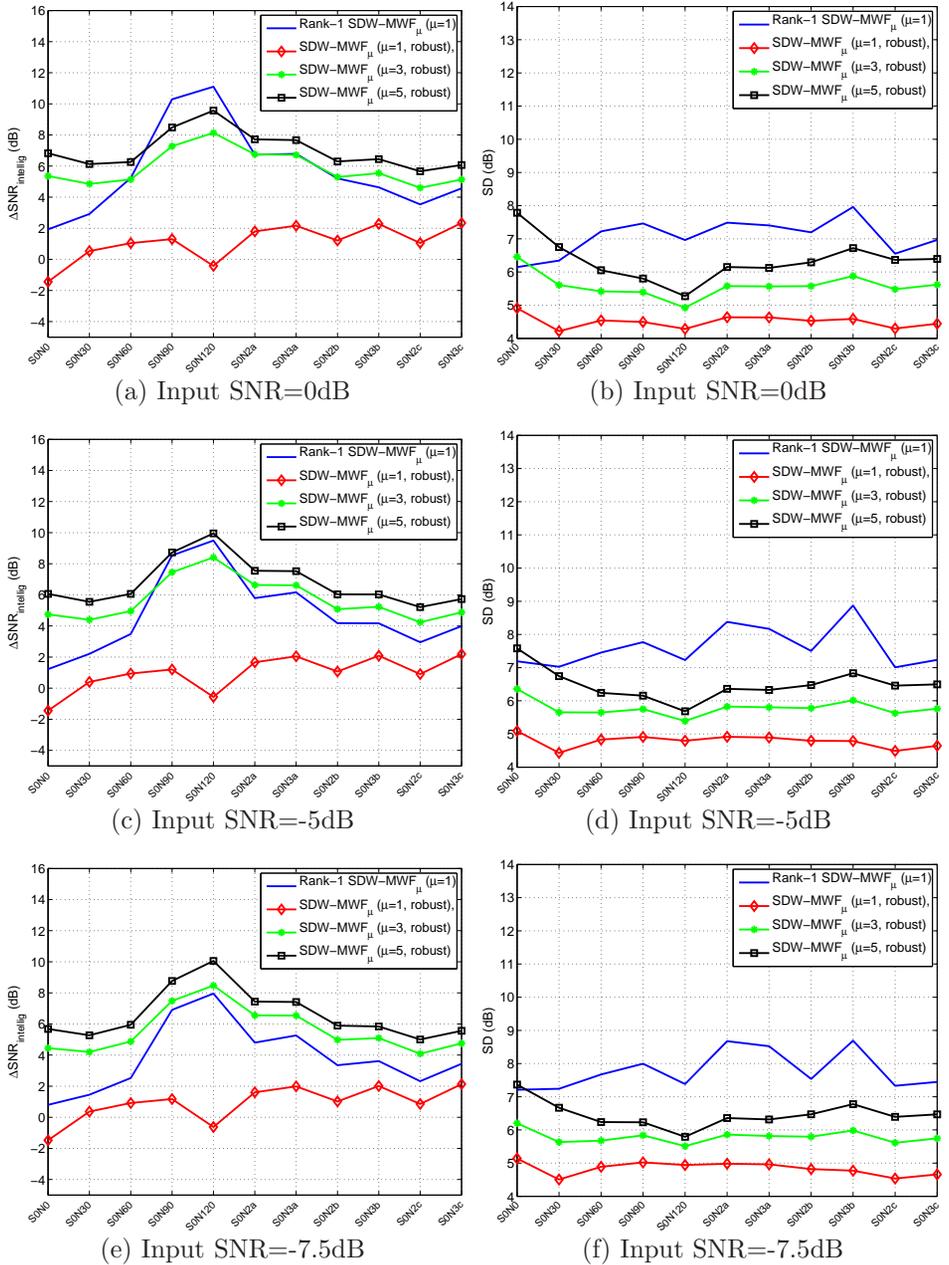


Figure 4.6: SNR improvement and SD for the comparison between the rank-1 SDW-MWF_μ, using a perfect VAD, to the SDW-MWF_μ using the robust correlation matrices.

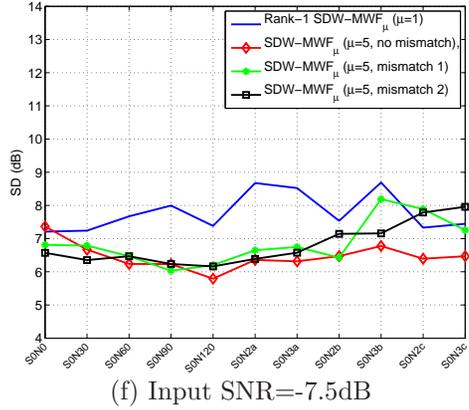
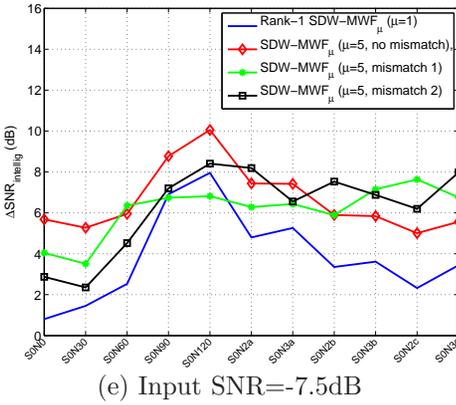
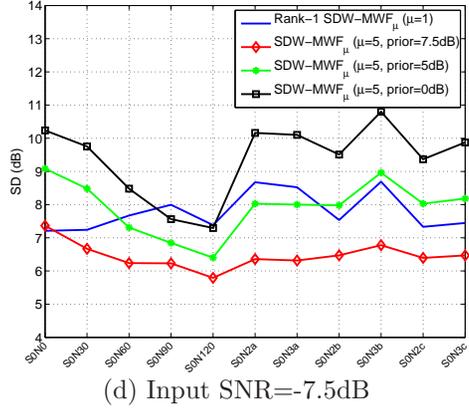
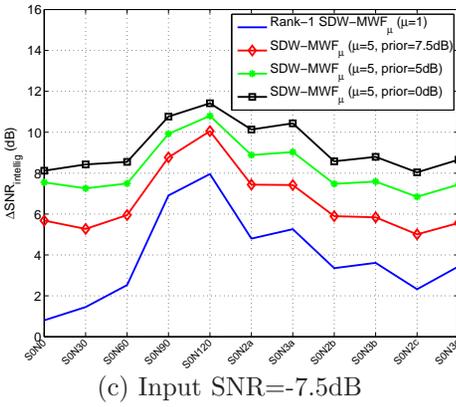


Figure 4.7: SNR improvement and SD for the comparison between the rank-1 SDW-MWF $_{\mu}$, using a perfect VAD, to the SDW-MWF $_{\mu}$ using the robust correlation matrices with spatial mismatch.

Chapter 5

Robust Capon beamforming for small arrays

This chapter presents a different multi-channel NR algorithm based on a standard Capon beamformer (SCB) also referred to as an MVDR beamformer. The main difference between the SCB and the MWF is that the SCB relies on a correct estimation of the steering vector of the target speech signal whereas the MWF is uniquely based on the estimated correlation matrices. This means that the estimated correlation matrices are not mainly responsible for the SCB performance but the target now is to find a robust method to estimate the steering vector. Therefore a robust Capon beamformer (RCB) is presented where the target is to adaptively estimate the steering vector in the presence of reverberation and noise. The proposed RCB is based on using prior knowledge of the steering vector combined with a steering vector uncertainty principle.

Section 5.1 gives a short introduction to the SCB problem and the motivation for the proposed RCB.

Section 5.2 introduces the concept behind the SCB together with the problem of having a mismatch between the presumed and the actual steering vector.

Section 5.3 reviews some previous popular methods for the RCB, where inspiration for the proposed RCB is taken from an approach related to the uncertainty based beamformers.

Section 5.4 presents the proposed RCB that has an adaptive estimation of the steering vectors which is based on using prior knowledge to constrain the steering vectors based on an uncertainty principle. The focus of the proposed RCB is on small arrays and low complexity and therefore the computational complexity is

also compared between the SCB and the RCB.

Section 5.5 presents the experimental results to confirm the robustness of the estimated steering vectors.

5.1 Introduction

The SCB [22] suffers from a substantial performance degradation when there is a mismatch between the presumed and the actual steering vector of the target signal. Therefore many approaches have been proposed to improve the robustness of the SCB. A variation of the SCB is known as the linearly constrained minimum variance (LCMV) beamformer [93] where a set of linear constraints is added. These constraints broaden the main beam by imposing a set of unity-gain constraints for steering vectors close to the presumed steering vector of the target signal such that robustness against a steering vector mismatch is achieved. A drawback with the LCMV is that each constraint removes one degree of freedom for interference suppression. Other robust extensions of the SCB have been based on diagonal loading of the sample correlation matrix [28][244]. The main problem with these approaches is to find the optimal value of the diagonal loading factor and that it reduces performance and the beam sharpness. Recent approaches estimate the diagonal loading factor based on the uncertainty region of the presumed steering vector of the target signal. These methods are robust against target signal suppression when the actual steering vector is within the predefined uncertainty region. Spherical [126][233], ellipsoid [7][122][125][134][214] and polyhedron [241] uncertainty regions have all been studied.

In [129][130] it is shown that a frequency-domain SCB outperforms a time-domain Frost beamformer and a generalized sidelobe canceler for a scenario with two or more nonstationary interfering speech sources and an array with two microphones. A frequency-domain SCB exploits the time-frequency sparseness of the sources better than a time-domain implementation.

In [130][129] a fixed steering vector is used for the target signal; the goal in this chapter is to extend the frequency-domain SCB to an adaptive frequency-domain RCB. The RCB proposed here is based on a gradient approach where the steering vector is adaptively estimated based on a predefined level of uncertainty in the steering vector. The proposed RCB offers a low complexity, simple implementation and suffers no loss of degrees of freedom for interference suppression.

5.2 Standard Capon Beamforming (SCB)

5.2.1 Optimization criterion for SCB

The goal of the SCB is to minimize the total beamformer output variance while constraining the target speech signal response to be unity to prevent speech signal suppression. The SCB design can be formulated as

$$\min_{\mathbf{W}(k,l)} \mathbf{W}^H(k,l) \mathbf{R}_x(k,l) \mathbf{W}(k,l), \quad \text{s. t.} \quad \bar{\mathbf{e}}^H(k,l) \mathbf{W}(k,l) = 1 \quad (5.1)$$

where $\bar{\mathbf{e}}(k,l)$ is the presumed steering vector of the target speech signal. The closed-form solution to (5.1) is given by Capon [22] as

$$\mathbf{W}_{\text{SCB}}(k,l) = \frac{\mathbf{R}_x^{-1}(k,l) \bar{\mathbf{e}}(k,l)}{\bar{\mathbf{e}}^H(k,l) \mathbf{R}_x^{-1}(k,l) \bar{\mathbf{e}}(k,l)}, \quad (5.2)$$

where the output power σ^2 is given by

$$\sigma^2 = \frac{1}{\bar{\mathbf{e}}^H(k,l) \mathbf{R}_x^{-1}(k,l) \bar{\mathbf{e}}(k,l)}. \quad (5.3)$$

5.2.2 Mismatch between presumed and actual steering vector

As mentioned, the SCB does not provide robustness against the case where there is a mismatch between the presumed and the actual steering vectors $\bar{\mathbf{e}}(k,l)$ and $\mathbf{e}_x(k,l)$, respectively. The mismatch between the presumed and the actual steering vector can be described as

$$\mathbf{e}_x(k,l) = \bar{\mathbf{e}}(k,l) + \Delta_{\bar{\mathbf{e}}} \quad (5.4)$$

where $\Delta_{\bar{\mathbf{e}}}$ is an unknown complex vector. Under such mismatch the constraint $\bar{\mathbf{e}}^H(k,l) \mathbf{W}_{\text{SCB}}(k,l) = 1$ leads to part of the speech signal being suppressed and hence a degradation in SNR and signal distortion. The goal is then to design a RCB that can compensate for the mismatch $\Delta_{\bar{\mathbf{e}}}$ by estimating the actual steering vector $\mathbf{e}_x(k,l)$.

5.3 Previous work on robust Capon beamformers

In the past, many RCB approaches have been proposed. In this section, we briefly review some popular methods such as the linearly constrained minimum variance (LCMV) beamformer, the diagonal-loading-based beamformer, and the uncertainty-based beamformer.

5.3.1 Linearly constrained minimum variance

In [93][222] the LCMV beamformer is proposed where the linear constraint in (5.1) is generalized such that the output power is minimized while restricting the filter weights to satisfy one or more linear equality constraints. The LCMV design can be formulated as

$$\min_{\mathbf{W}(k,l)} \mathbf{W}^H(k,l) \mathbf{R}_x(k,l) \mathbf{W}(k,l), \quad \text{s. t.} \quad \mathbf{C}^H \mathbf{W}(k,l) = \mathbf{f} \quad (5.5)$$

where \mathbf{C} is the constraint matrix and the vector \mathbf{f} specifies the corresponding constraint value for each vector. These additional linear constraints can be either directional constraints [220] or derivative constraints [3][18], where the core idea is to broaden the main beam of the beampattern so that it is more robust against steering vector mismatch. It should however be mentioned that for every additional linear constraint imposed, the beamformer loses one degree of freedom for interference suppression, which is undesirable for small arrays.

5.3.2 Diagonal-loading-based beamformer

In [23][125] the diagonal-loading-based RCB is proposed where a regularization of the correlation matrix is included. The diagonal-loading-based RCB design can be formulated as

$$\min_{\mathbf{W}(k,l)} \mathbf{W}^H(k,l) (\mathbf{R}_x(k,l) + \gamma \mathbf{I}) \mathbf{W}(k,l), \quad \text{s. t.} \quad \bar{\mathbf{e}}^H(k,l) \mathbf{W}(k,l) = 1 \quad (5.6)$$

where γ is the regularization factor and \mathbf{I} is the identity matrix. The problem here is the difficulty of specifying the optimal value of γ . If γ is set too large the beamformer loses interference suppression performance, and if γ is set too small the robustness is sacrificed.

5.3.3 Uncertainty-based beamformer

A recent robust beamforming approach suggests that the steering vector mismatch can be accounted for based on a predefined uncertainty region around the presumed steering vector. The uncertainty-based RCB design can be formulated as [134][122]

$$\min_{\mathbf{W}(k,l)} \mathbf{W}^H(k,l) \mathbf{R}_x(k,l) \mathbf{W}(k,l), \quad \text{s. t.} \quad |\mathbf{W}^H(k,l) (\bar{\mathbf{e}}(k,l) + \Delta)| \geq 1, \\ \text{for all } \|\Delta\| \leq \varepsilon \quad (5.7)$$

where Δ is the steering vector mismatch as in (5.4) and ε denotes the uncertainty region. This constraint forces the magnitude responses for an uncertainty set of

steering vectors to exceed unity. The drawback of the uncertainty-based RCB formulation in (5.7) is the infinite number of constraints. The performance is said to be optimal if ε is large enough to cover the mismatch between the presumed and the actual steering vector of the target signal. In [233] the problem is reformulated as a second-order cone programming problem, based on a spherical uncertainty set, which can be solved (although at very high computational complexity) using the well-established interior-point method. In order to use standard optimization techniques, the uncertainty region has often been generalized to an ellipsoid [134][125][7] or a polyhedron [241].

5.3.4 Max-min optimization

An alternative consists in estimating the steering vector \mathbf{e} which results in the maximal output power [244][174], i.e.,

$$\boxed{\max_{\mathbf{e}(k,l)} \min_{\mathbf{W}(k,l)} \mathbf{W}^H(k,l) \mathbf{R}_x(k,l) \mathbf{W}(k,l), \text{ s. t. } \mathbf{e}^H(k,l) \mathbf{W}(k,l) = 1, \|\mathbf{e}(k,l)\|^2 = 1} \quad (5.8)$$

where the unit norm is included to avoid a scaling ambiguity in the output power. With (5.3) the expression in (5.8) can be simplified to

$$\max_{\mathbf{e}(k,l)} \frac{1}{\mathbf{e}^H(k,l) \mathbf{R}_x^{-1}(k,l) \mathbf{e}(k,l)}, \text{ s. t. } \|\mathbf{e}(k,l)\|^2 = 1 \quad (5.9)$$

which is equivalent to

$$\min_{\mathbf{e}(k,l)} \mathbf{e}^H(k,l) \mathbf{R}_x^{-1}(k,l) \mathbf{e}(k,l), \text{ s. t. } \|\mathbf{e}(k,l)\|^2 = 1. \quad (5.10)$$

This is a principal eigenvector problem where the eigenvalue corresponding to the largest eigenvector of $\mathbf{R}_x(k,l)$ provides the estimate of the steering vector $\mathbf{e}(k,l)$. If the speech signal is not the dominant signal the solution in (5.9) leads to a wrong solution. In this case a subset of eigenvalues (speech-plus-noise) can be chosen at the cost of a reduced resolution [174]. The problem in (5.10) has been reformulated in [125] as

$$\boxed{\min_{\mathbf{e}(k,l)} \mathbf{e}^H(k,l) \mathbf{R}_x^{-1}(k,l) \mathbf{e}(k,l) \quad \text{s. t.} \quad \|\mathbf{e}(k,l) - \bar{\mathbf{e}}(k,l)\|^2 \leq \varepsilon.} \quad (5.11)$$

The idea behind (5.11) is illustrated in Figure 5.1. The radius ε of the sphere defines the uncertainty region. The solution to (5.11) is typically solved by observing that the solution to (5.11) will be on the boundary of the constraint, and the changing the inequality constraint to an equality constraint [125] i.e.,

$$\min_{\mathbf{e}(k,l)} \mathbf{e}^H(k,l) \mathbf{R}_x^{-1}(k,l) \mathbf{e}(k,l) \quad \text{s. t.} \quad \|\mathbf{e}(k,l) - \bar{\mathbf{e}}(k,l)\|^2 = \varepsilon. \quad (5.12)$$

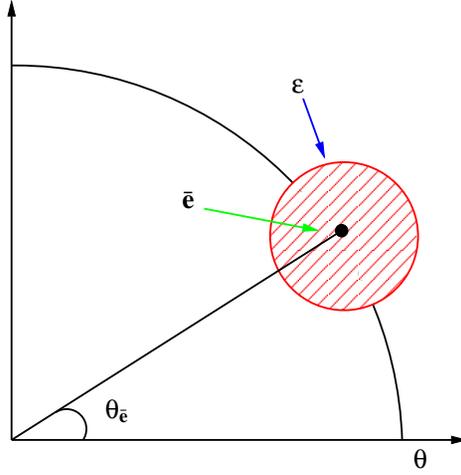


Figure 5.1: Hyperplane for the RCB defined in (5.11).

This problem can be solved by using the Lagrange multiplier methodology which is based on the function

$$f = \mathbf{e}^H(k, l) \mathbf{R}_x^{-1}(k, l) \mathbf{e}(k, l) + \lambda (\|\mathbf{e}(k, l) - \bar{\mathbf{e}}(k, l)\|^2 - \varepsilon) \quad (5.13)$$

where $\lambda \geq 0$ is the Lagrange multiplier. Differentiation of (5.13) with respect to $\mathbf{e}(k, l)$ gives the optimal steering vector

$$\hat{\mathbf{e}}(k, l) = \left(\frac{\mathbf{R}_x^{-1}(k, l)}{\lambda} + \mathbf{I} \right)^{-1} \bar{\mathbf{e}}(k, l) \quad (5.14)$$

$$= \bar{\mathbf{e}}(k, l) - (\mathbf{I} + \lambda \mathbf{R}_x(k, l))^{-1} \bar{\mathbf{e}}(k, l) \quad (5.15)$$

where the Lagrange multiplier $\lambda \geq 0$ can be estimated by solving the constraint equation

$$g(\lambda) \triangleq \|(\mathbf{I} + \lambda \mathbf{R}_x(k, l))^{-1} \bar{\mathbf{e}}(k, l)\|^2 = \varepsilon. \quad (5.16)$$

which is shown to have a unique solution. Then by replacing the expression in (5.14) with $\bar{\mathbf{e}}(k, l)$ in (5.2) gives

$$\mathbf{W}_{\text{DL}}(k, l) = \frac{(\mathbf{R}_x(k, l) + \frac{1}{\lambda} \mathbf{I})^{-1} \bar{\mathbf{e}}(k, l)}{\bar{\mathbf{e}}^H(k, l) (\mathbf{R}_x(k, l) + \frac{1}{\lambda} \mathbf{I})^{-1} \mathbf{R}_x(k, l) (\mathbf{R}_x(k, l) + \frac{1}{\lambda} \mathbf{I})^{-1} \bar{\mathbf{e}}(k, l)}. \quad (5.17)$$

This kind of RCB belongs to the class of diagonal loading based beamformers, which we will refer to as RCB-DL, where the diagonal loading factor is estimated

based on a given uncertainty set. The challenge is then again to find an optimal value for the radius ε . If ε is set too large the beamformer loses degrees of freedom for interference suppression, if ε is set too small the uncertainty region may not be sufficient to compensate for the steering vector mismatch.

5.4 Robust Capon beamforming (RCB)

Previous practical RCB methods mostly belong to the class of diagonal-loading-based beamformers [233][134][125][7], where the amount of diagonal loading may be calculated based on the uncertainty region around the presumed steering vector. This form of regularization reduces the beamformer's interference suppression capabilities, particularly for small arrays. In this section, we present an RCB that exploits the same uncertainty region but does not depend on a diagonal loading and where the steering vector is estimated based on a gradient algorithm. The new method can obtain both robustness and much greater noise suppression.

5.4.1 Proposed RCB formulation

The problem is that we obviously do not know the exact steering vector, only a region in which it should lie, i.e., if the selected vector from this region differs from the actual steering vector, the beamformer will attempt to minimize the output power by suppressing the target speech signal. If the selected vector is equal to the actual steering vector, the beamformer cannot cancel the target speech signal which leads to a larger output power. The proposed RCB formulation, when the uncertainty region is defined as a sphere, can be written as

$$\boxed{\begin{aligned} \max_{\mathbf{e}(k,l)} \min_{\mathbf{W}(k,l)} \mathbf{W}^H(k,l) \mathbf{R}_x(k,l) \mathbf{W}(k,l), \quad \text{s. t.} \quad \mathbf{e}^H(k,l) \mathbf{W}(k,l) = 1, \\ \|\mathbf{e}(k,l) - \bar{\mathbf{e}}(k,l)\|^2 \leq \varepsilon, \|\mathbf{e}(k,l)\|^2 = \|\bar{\mathbf{e}}(k,l)\|^2 \end{aligned}} \quad (5.18)$$

The proposed RCB differs from (5.7) in that instead of insisting to have a distortionless response for the entire uncertainty region an optimal steering vector is now estimated within the uncertainty region such that $\mathbf{e}^H(k,l) \mathbf{W}(k,l) = 1$ for that particular steering vector and hence the beamformer can do whatever is best elsewhere in the uncertainty region, thus allowing better performance in terms of suppressing noise. So, instead of having multiple constraints we now have a single adapted constraint. The proposed RCB also differs from (5.11) in that an additional constraint is included to avoid a scaling ambiguity (as in (5.8)), since the uncertainty region is defined as a sphere and certain steering vectors can therefore increase the power. The constraint $\|\mathbf{e}\|^2 = \|\bar{\mathbf{e}}\|^2$ therefore ensures that the estimated steering vector has the same power as the presumed steering vector.

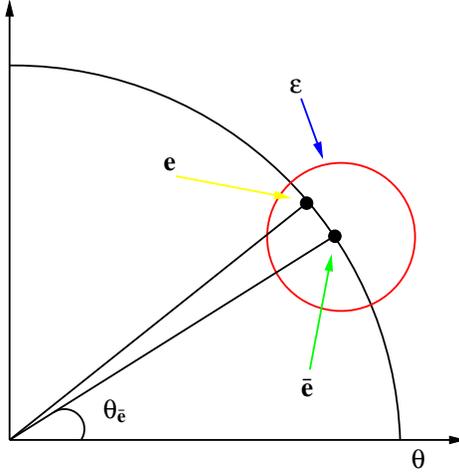


Figure 5.2: Hyperplane for the proposed RCB defined in (5.19).

The proposed RCB is shown in Figure 5.2 where the estimated steering vector lies within the uncertainty region, i.e., the beamformer is allowed to steer in certain directions as long as it remains within the predefined uncertainty region. Increasing or decreasing ε in our case only impacts the steering capabilities. The proposed RCB therefore suffers no loss of degrees of freedom for interference suppression, since the width of the beam is not increased. The proposed RCB also has the advantage of a low complexity and simple implementation, since the solution does not include an infinite number of constraints and hence a closed-form update expression can be derived.

Using the output power defined in (5.3), the optimization problem can be written as

$$\boxed{\begin{aligned} \min_{\mathbf{e}(k,l)} \mathbf{e}^H(k,l) \mathbf{R}_x^{-1}(k,l) \mathbf{e}(k,l), \quad \text{s. t.} \quad & \|\mathbf{e}(k,l) - \bar{\mathbf{e}}(k,l)\|^2 \leq \varepsilon, \\ & \|\mathbf{e}(k,l)\|^2 = \|\bar{\mathbf{e}}(k,l)\|^2. \end{aligned}} \quad (5.19)$$

As with all robust beamformers based on an uncertainty region, success depends on selecting a proper value of ε , since a large ε value may cause the steering vector to steer towards the interferences and a small ε value may not be sufficient to accommodate for the steering vector mismatch.

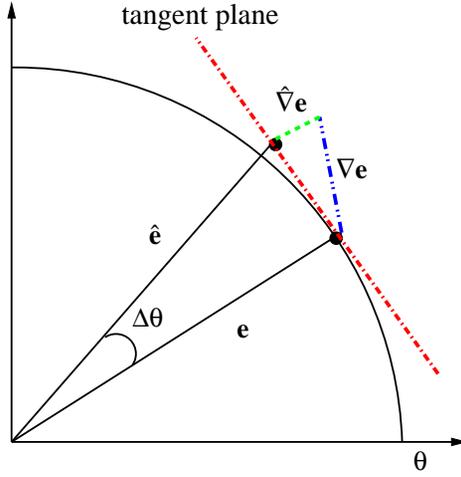


Figure 5.3: Gradient projected onto the tangent plane.

5.4.2 Gradient update of the steering vector

In this section, an algorithm is presented for adaptively estimating the steering vectors, based on (5.18) and a gradient approach. The gradient of the output power with respect to the estimated steering vector \mathbf{e} is:

$$\nabla \mathbf{e} = \frac{d}{d\mathbf{e}} (\mathbf{e}^H(k, l) \mathbf{R}_x^{-1}(k, l) \mathbf{e}(k, l)) = -2\mathbf{R}_x^{-1}(k, l) \mathbf{e}(k, l). \quad (5.20)$$

Additionally, we propose to project the gradient onto the tangent plane:

$$\hat{\nabla} \mathbf{e} = \nabla \mathbf{e} - \mathbf{e}(k, l) \frac{\mathbf{e}^H(k, l) \nabla \mathbf{e}}{\|\mathbf{e}(k, l)\|^2}. \quad (5.21)$$

The projection of the gradient onto the tangent plane is shown in Figure 5.3. Assuming that the gradient change is small, the benefit of the projection is that it remains close to the constrained steering vector norm, i.e., $\|\mathbf{e}(k, l)\|^2 = \|\hat{\mathbf{e}}(k, l)\|^2$. Furthermore, projecting the gradient onto the tangent plane before estimating the stepsize guarantees that any step remains within the tangent plane, which may not be the case if the projection is performed after the estimation of the stepsize. The stepsize in the gradient direction (tangent plane) can be calculated as follows:

$$\min_{\mu} (\mathbf{e}(k, l) + \mu \hat{\nabla} \mathbf{e})^H \mathbf{R}_x^{-1}(k, l) (\mathbf{e}(k, l) + \mu \hat{\nabla} \mathbf{e}), \quad \text{s. t.} \quad \|\mu \hat{\nabla} \mathbf{e}\|^2 \leq \|\alpha \mathbf{e}(k, l)\|^2 \quad (5.22)$$

where α controls the maximum stepsize change for each update. The constraint limits the change in the steering vector so as to avoid potential adaptation noise or

artifacts. The maximum stepsize yields a maximum reduction in the error criterion at each update. The stepsize μ can be found by differentiating (5.22) with respect to μ :

$$\frac{d}{d\mu}(\mathbf{e}(k, l) + \mu \hat{\nabla} \mathbf{e})^H \mathbf{R}_x^{-1}(k, l)(\mathbf{e}(k, l) + \mu \hat{\nabla} \mathbf{e}) \quad (5.23)$$

and setting the derivative to zero

$$\hat{\nabla} \mathbf{e}^H \mathbf{R}_x^{-1}(k, l) \mathbf{e}(k, l) + \mathbf{e}^H(k, l) \mathbf{R}_x^{-1}(k, l) \hat{\nabla} \mathbf{e} + 2\mu \hat{\nabla} \mathbf{e}^H \mathbf{R}_x^{-1}(k, l) \hat{\nabla} \mathbf{e} = 0 \quad (5.24)$$

gives

$$\mu = -\frac{\hat{\nabla} \mathbf{e}^H \mathbf{R}_x^{-1}(k, l) \mathbf{e}(k, l) + \mathbf{e}^H(k, l) \mathbf{R}_x^{-1}(k, l) \hat{\nabla} \mathbf{e}}{2\hat{\nabla} \mathbf{e}^H \mathbf{R}_x^{-1}(k, l) \hat{\nabla} \mathbf{e}}. \quad (5.25)$$

The estimated stepsize μ is only used if $\|\mu \hat{\nabla} \mathbf{e}\|^2 \leq \|\alpha \mathbf{e}(k, l)\|^2$ otherwise the stepsize is normalized as

$$\mu = \alpha \frac{\|\mathbf{e}(k, l)\|}{\|\hat{\nabla} \mathbf{e}\|}. \quad (5.26)$$

The steering vector is then updated as follows:

$$\hat{\mathbf{e}}(k, l) = \mathbf{e}(k, l) + \mu \hat{\nabla} \mathbf{e}, \quad (5.27)$$

and is selected if $\|\hat{\mathbf{e}}(k, l) - \bar{\mathbf{e}}(k, l)\|^2 \leq \varepsilon$ otherwise the previous estimated steering vector is selected. To satisfy the norm constraint in (5.19) the normalized updated steering vector can be written as

$$\hat{\mathbf{e}}(k, l) = \hat{\mathbf{e}}(k, l) \frac{\|\bar{\mathbf{e}}(k, l)\|}{\|\hat{\mathbf{e}}(k, l)\|}. \quad (5.28)$$

The proposed RCB can then be written as

$$\mathbf{W}_{\text{RCB}}(k, l) = \frac{\mathbf{R}_x^{-1}(k, l) \hat{\mathbf{e}}(k, l)}{\hat{\mathbf{e}}^H(k, l) \mathbf{R}_x^{-1}(k, l) \hat{\mathbf{e}}(k, l)}. \quad (5.29)$$

It is also clear that, when the sphere reduces to a single point, the proposed RCB is equivalent to the SCB.

5.4.3 Computational complexity

The computational complexity of the SCB is analyzed in [130] and here we will analyze the complexity of the proposed RCB. The comparison is made in terms of

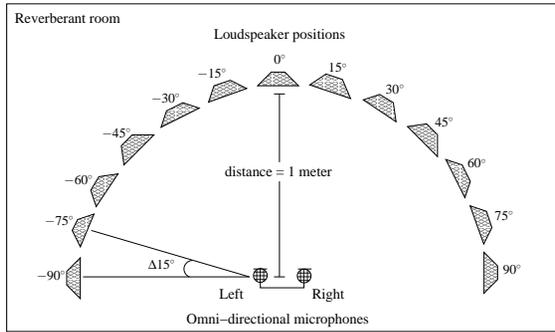


Figure 5.4: Simulation setup with loudspeaker positions.

real additions, real multiplications, and real divisions, by assuming that two real additions are required for each complex addition, four real multiplications and two real additions are required for each complex multiplication, and two complex multiplications and two real divisions for each complex division are required. Table 5.1 summarizes the computational complexity, for an example with $M=2$ (number of microphones) and $K=1024$ (FFT length). This shows that the increase in computational complexity for the RCB is not that large compared to the SCB. Furthermore the proposed RCB design gives a simple implementation and no loss of degrees of freedom for interference suppression. The performance of the proposed RCB is shown in the following section.

5.5 Experimental results

5.5.1 Experimental set-up

Simulations have been performed in a reverberant rectangular conference room with dimensions 9m x 4m x 2.75m (length x width x height) and a reverberation time, $T_{60} \approx 0.37s$, with two omnidirectional microphones in free-field separated by 15cm. Further details can be found in [130]. The loudspeakers are positioned at 1m from the microphones in the frontal horizontal plane at angles ranging from -90° to $+90^\circ$ relative to the microphones, with a spacing of 15° , which is illustrated in Figure 5.4. The speech signals consist of male and female sentences each approximately 2.5s. in duration. The signals are sampled at 22.05 kHz. The spatial scenarios used in the simulations are listed in Table 5.2. A total of three simulations are performed, i.e., the case with two, three and four interferences. For each simulation the first 2.5s sentence uses the configuration "a" and the next 2.5s sentence uses the configuration "b". The beamformer is evaluated for a different

Part of RCB algorithm	Complex additions	Complex multiplies	Real additions	Real multiplies	Real divisions
Gradient estimation	0	0	0	0	0
Projection	$M + (M - 1)\left(\frac{K}{2} + 1\right)$	$2M\left(\frac{K}{2} + 1\right)$	0	0	2
Stepsize	$M(M - 1)\left(\frac{K}{2} + 1\right)$	$M^2\left(\frac{K}{2} + 1\right)$	$2M\left(\frac{K}{2} + 1\right)$	$4M\left(\frac{K}{2} + 1\right)$	$\left(\frac{K}{2} + 1\right)$
Gradient update	$M\left(\frac{K}{2} + 1\right)$	0	0	$2M\left(\frac{K}{2} + 1\right)$	0
Stepsize constraint	0	0	$3M - 1\left(\frac{K}{2} + 1\right)$	$4M + 2\left(\frac{K}{2} + 1\right)$	$\left(\frac{K}{2} + 1\right)$
Steering vector constraint	$2M\left(\frac{K}{2} + 1\right)$	0	$2M - 1\left(\frac{K}{2} + 1\right)$	$2M + 1\left(\frac{K}{2} + 1\right)$	0
SCB [130]	-	-	55355	59455	1026
Proposed RCB	-	-	74861	83578	1030

Table 5.1: Total computational complexity for the proposed RCB compared to SCB.

Notation	Spatial angle of source(s)
1a	Speech at 0° , noise at -30° and 30°
1b	Speech at 0° , noise at -45° and 45°
2a	Speech at 0° , noise at -90° , -30° and 45°
2b	Speech at 0° , noise at -90° , 45° and 75°
3a	Speech at 0° , noise at -90° , -45° , 45° and 90°
3b	Speech at 0° , noise at -90° , -30° , 30° and 90°

Table 5.2: Spatial scenarios for the experimental evaluation

range of input SNR's and for different values of α_x defined in (2.9). An FFT length of 1024 with 50% overlap was used in all simulations. The radius ε of the uncertainty region is defined as $\varepsilon = \|\beta \bar{\mathbf{e}}(k, l)\|^2$, where different values of β are considered and since the target speech signal is coming from 0° angle then $\bar{\mathbf{e}} = \mathbf{e}_1$.

5.5.2 Results

The SNR and SD for the three cases are shown in Figure 5.5, 5.6 and 5.7. The performance benefits in terms of output SNR of the proposed RCB is clear, especially when $\alpha_x = 0.75$ and $\alpha_x = 0.85$ with $\beta = 0.35$, and furthermore the distortion is also much lower. When $\alpha_x = 0.95$ the performance of the proposed RCB is closer to the SCB (especially at low input SNR) but the benefit still remains at higher input SNR. It is worth noting that in the case with two interferers the SCB outperforms the RCB with $\beta = 0.35$ at low input SNR, but this can be avoided when $\beta = 0.15$. This does not seem to happen for the case with three and four interferers, which suggests that for the case with two interferers, $\alpha_x = 0.95$ was sufficient to track the non-stationarity of the sources. It is clear that the performance of the SCB highly depends on the accuracy of the estimated correlation matrices, whereas the proposed RCB is more robust. The reduced performance when $\alpha_x = 0.95$ for the proposed RCB can be caused by the limited tracking performance.

Overall it seems that $\beta = 0.35$ is a reasonable value and therefore it would be interesting to evaluate the RCB performance for larger values of β . This would then correspond to an estimation of the steering vector that is not constraint. For this purpose the experiment with four interferers is repeated with different values of β which is shown in Figure 5.8. From this it is clear that $\beta = 0.35$ gives the best performance in terms of SNR and SD. It is especially worth noting the increase in distortion when $\beta = 0.95$. This shows the importance of estimating robust steering vectors.

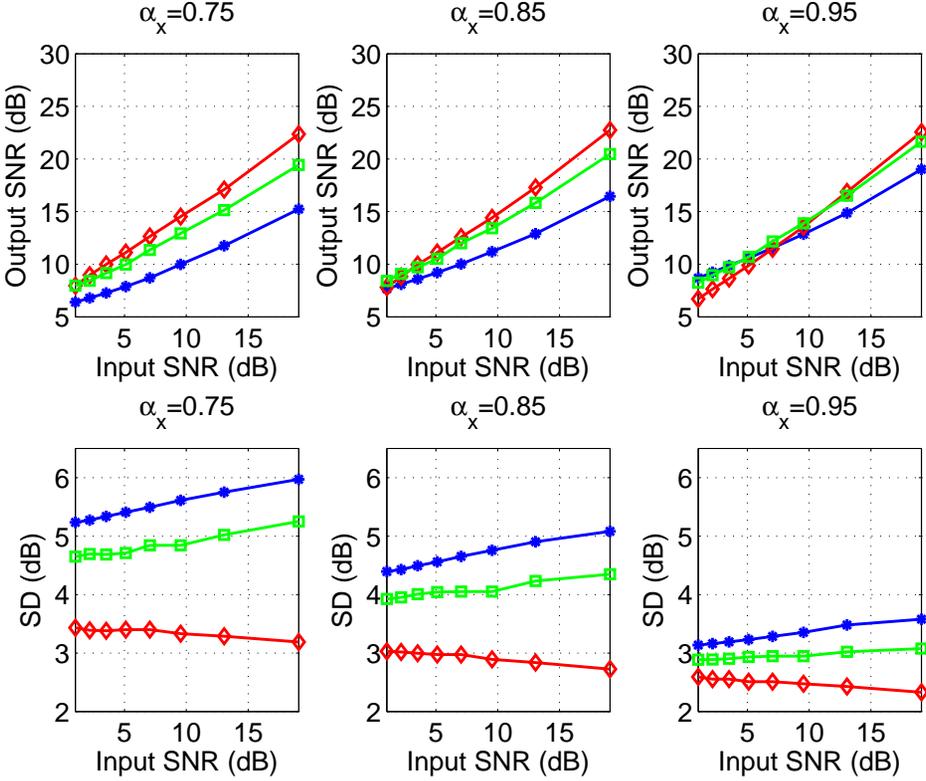


Figure 5.5: SNR and SD for two interferers for different values of α_x and. (*) SCB, (\square) RCB ($\beta = 0.15$), (\diamond) RCB ($\beta = 0.35$).

The proposed RCB is also compared to a diagonal loading based beamformer referred to as RCB-DL. The diagonally loaded correlation matrix can then be written as

$$\hat{\mathbf{R}}_{DL} = \hat{\mathbf{R}}_x(k, l) + \delta \mathbf{I} \quad (5.30)$$

where δ is the value of diagonal loading factor and \mathbf{I} is the identity matrix. In this case we define δ to be a fraction of the largest eigenvalue. The comparison between the proposed RCB and the RCB-DL is shown in Figure 5.9. The results show a small but consistent SNR benefit for the proposed RCB compared to the RCB-DL. At the same time the proposed RCB maintains a low distortion compared to the RCB-DL.

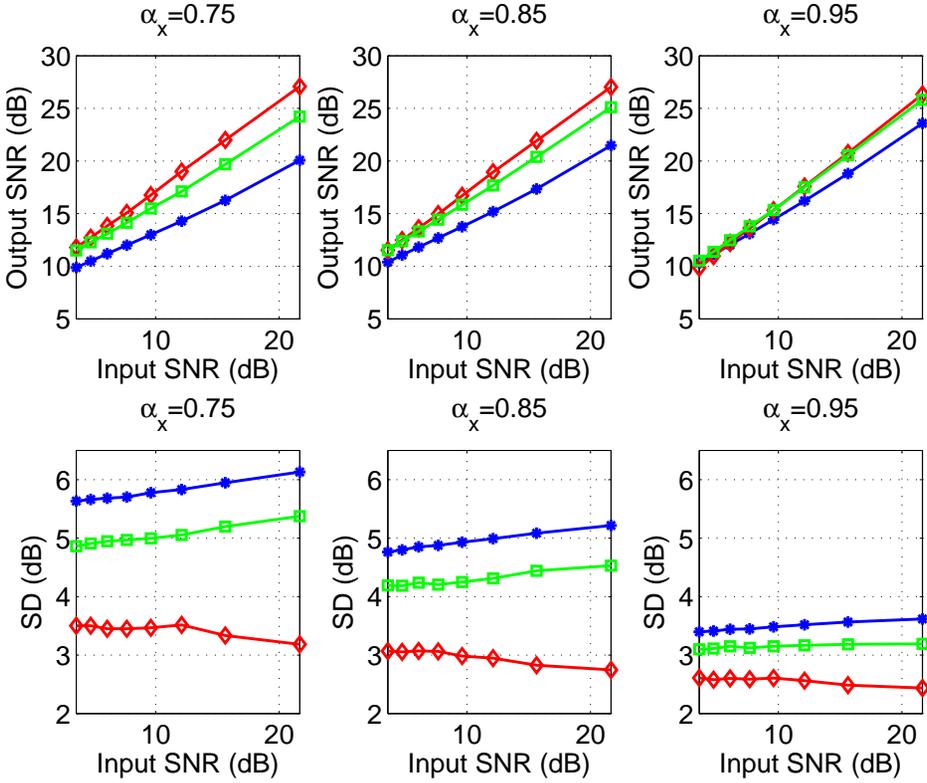


Figure 5.6: SNR and SD for three interferers for different values of α_x . (*) SCB, (\square) RCB ($\beta = 0.15$), (\diamond) RCB ($\beta = 0.35$).

5.6 Conclusion

In this chapter, a frequency-domain RCB is presented which is based on a gradient approach where the steering vector is estimated adaptively based on a predefined uncertainty region of the steering vector mismatch. Experimental results for a scenario with a microphone array containing two omnidirectional microphones and with competing speakers in a reverberant room demonstrate that the proposed RCB outperforms the SCB in terms of output SNR and signal distortion. The proposed RCB showed robustness especially in the case where a small forgetting factor was used to estimate the correlation matrices, whereas the SCB showed a significant performance degradation. The proposed RCB also showed a small SNR benefit compared to a diagonal loading based beamformer. However the SNR benefit was achieved at a lower distortion. Furthermore, the proposed RCB offers

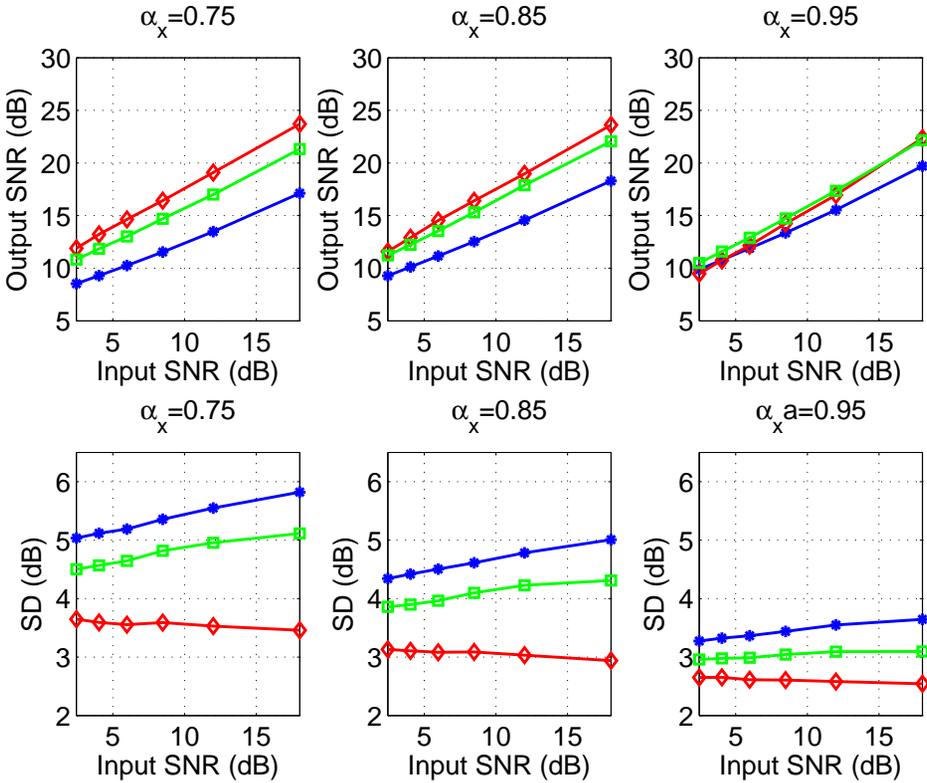


Figure 5.7: SNR and SD for four interferers for different values of α_x . (*) SCB, (□) RCB ($\beta = 0.15$), (◇) RCB ($\beta = 0.35$).

a low complexity, simple implementation and no loss of degrees of freedom for interference suppression.

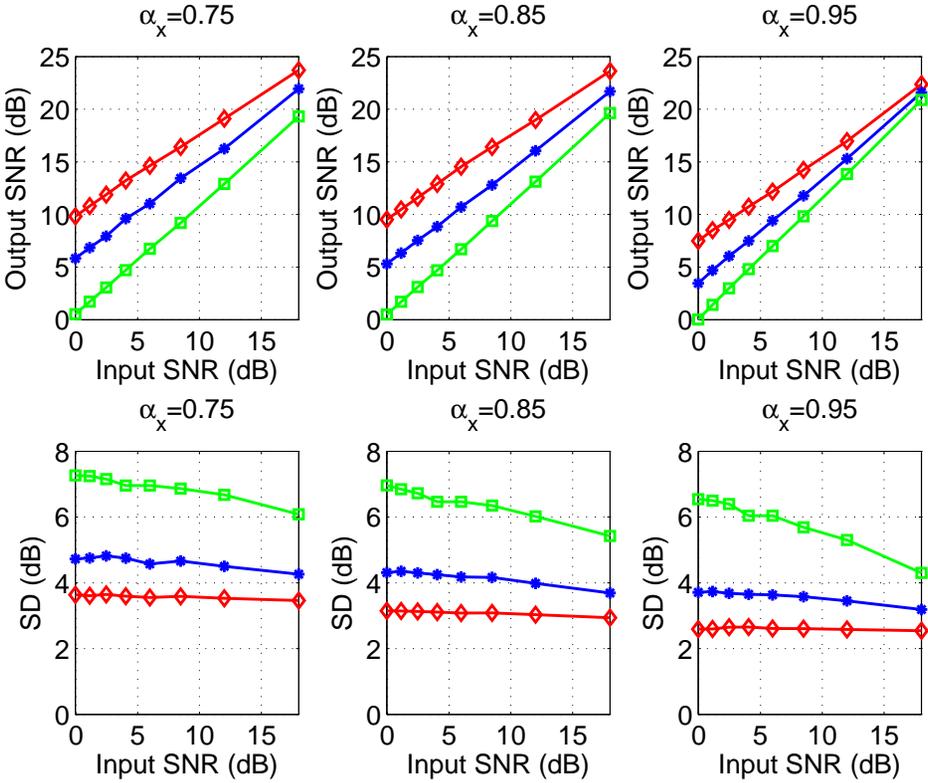


Figure 5.8: SNR and SD for four interferers for different values of α_x . (*) RCB ($\beta = 0.75$), RCB ($\beta = 0.95$), (\diamond) RCB ($\beta = 0.35$).

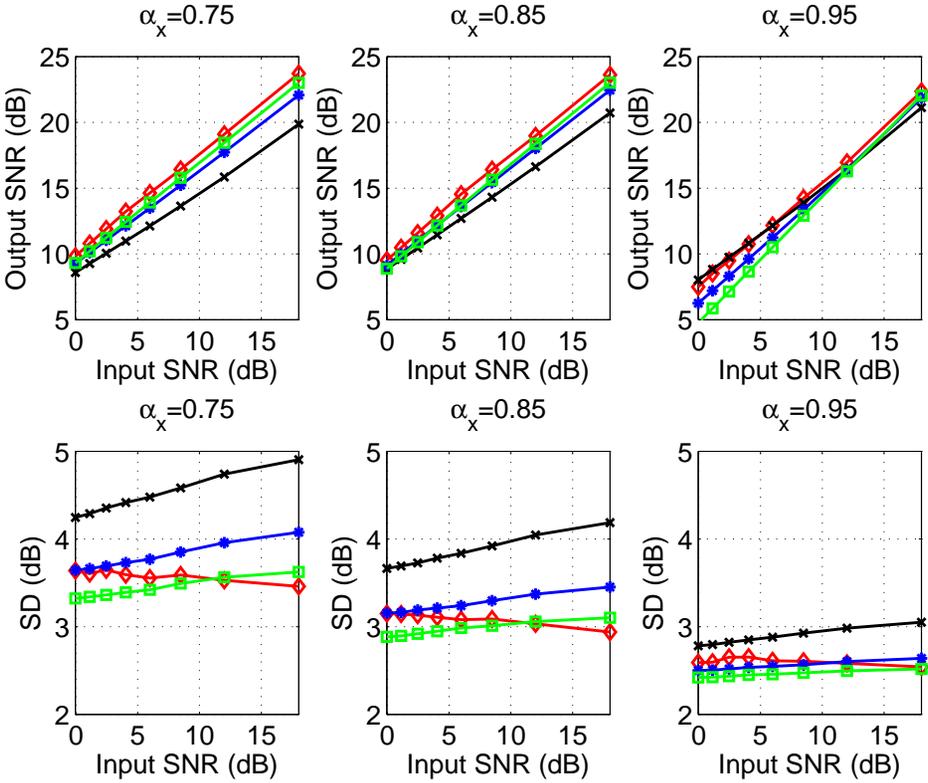


Figure 5.9: SNR and SD for four interferers for different values of α_x . (x) RCB-DL ($\delta = 0.05$), (*) RCB-DL ($\delta = 0.10$), (\square) RCB-DL ($\delta = 0.15$), (\diamond) RCB ($\beta = 0.35$).

Chapter 6

Dynamic range compression (DRC)

This chapter introduces the DRC algorithm used in this dissertation. DRC is a basic component in digital hearing aids and the use of DRC in hearing aids has increased over the years [150][201]. The role of the DRC is to estimate a desirable gain to map the wide dynamic range of an input audio (e.g. speech) signal into the reduced dynamic range of a hearing impaired listener. DRC is a signal processing strategy that makes speech audible over a wide range of sound levels and reduces the dynamic range of speech signals. Basically, a DRC is an automatic gain control, where the gain is automatically adjusted based on the intensity level of the input signal. Typically the design of DRC is based on clean speech scenarios without considering the presence of background noise. Therefore the work here is focussed on the design of DRC operating in the presence of background noise, i.e., to analyze how the DRC reacts to the background noise compared to clean speech scenarios.

Section 6.1 presents the design of the DRC algorithm used in this work together with the typical parameters that are involved in a DRC algorithm.

Section 6.2 analyzes the effect that the background noise has on the DRC. First it is shown why a typical DRC algorithm is bound to degrade the SNR especially when noise dominant segments receives more amplification compared to speech dominant segments.

Section 6.3 presents the experimental results to confirm the observation made in Section 6.2 and to properly state the problem and the negative effect that the background noise has on the DRC.

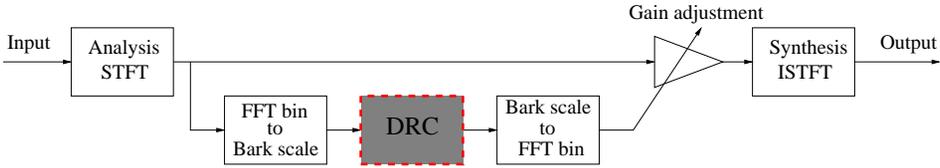


Figure 6.1: Block diagram of the multi-band DRC.

6.1 Design of DRC algorithms

Reduced audibility and reduced dynamic range between the hearing threshold and the uncomfortable level are some of the problems that people with a sensorineural hearing loss are dealing with [45][100]. The role of dynamic range compression (DRC) algorithms in hearing aids is to map the wide dynamic range of speech signals into the reduced dynamic range of hearing impaired listeners. Hearing aids or more specifically DRC should enhance the speech signal such that all of the important features of the speech signal are above the hearing threshold but at the same time below the discomfort level [149]. This is achieved by allowing more gain at low input levels and less at higher input levels which means that the DRC provides comfort for loud sounds and audibility for soft sounds.

Even though DRC is a main component in hearing aids there is still a disagreement about the best way to incorporate DRC in hearing aids [45][100]. In the past extensive work has analyzed the challenges and difficulties in the design of DRC algorithms [21][53][120][149][154][215][175][231]. The general design of different DRC algorithms can be found in [13][87][110][127][193]. The aim here is to show the effect that background noise has on DRC, and to discuss the problems and challenges when designing DRC algorithms in the presence of background noise.

6.1.1 Multi-band compression

The developments on DRC have mainly been focussed on multi-band DRC since the hearing loss and the dynamic range of speech varies markedly with frequency [148][193][26]. This can be achieved by using filter banks [110][109] or an FFT approach. In this work the critical bands [247] are realized using an FFT approach such that the FFT bins are combined to produce a critical band spectrum, i.e., by using individual FFT bins at low frequencies and by combining FFT bins at higher frequencies [110][61][96]. A block diagram of the multi-band DRC is shown in Figure 6.1. First the input signal is divided into frames using either an overlap-add or an overlap-save procedure with a window function. Then an FFT is performed on each frame and as input to the DRC the FFT bins are combined to produce a critical band spectrum. The DRC block then estimates

the required DRC gain based on the input level from each critical band. The estimated DRC gain is then converted back to the linear frequency and applied to the input spectrum. The final step is the reconstruction and the ISTFT operation. In the next section we will describe the operation inside the DRC block.

6.1.2 DRC parameters

The DRC is typically defined by the following parameters:

- Compression threshold (CT).
- Compression ratio (CR).
- Attack (at) and release time (rt).
- DRC gain G_{dB}^s .

The CT is defined in dB and corresponds to the point where the DRC becomes active, i.e., where the gain is reduced. The CR determines the degree of compression. A CR of 2 (i.e. 2:1) means that for every 2dB SPL increase in the input signal, the output signal increases by 1dB SPL. The attack and release time are defined in milliseconds and specify how fast the gain is changed according to changes in the input signal. The attack time is defined as the time taken for the compressor to react to an increase in input signal level. The release time is the time taken for the compressor to react to a decrease in input SPL and G_{dB}^s is defined as the speech DRC gain. For the DRC the input level for each critical band in dB SPL is defined as

$$P_{\text{DRC,dB}}^{\text{in,s}}(k', l) = 20 \log_{10} \left(\frac{|P_{\text{DRC}}^{\text{in}}(k', l)|}{P_{\text{ref}}} \right) \quad (6.1)$$

where k' is used to indicate that the linear frequency is now mapped to the Bark scale and P_{ref} is the reference sound pressure (20 micro Pascal). The DRC curve is defined based on a linear curve and a compression curve defined in (6.2) and (6.3), respectively:

$$P_{\text{lin,dB}}(k', l) = P_{\text{DRC,dB}}^{\text{in,s}}(k', l) + G_{\text{dB}}^s \quad (6.2)$$

$$P_{\text{cp,dB}}(k', l) = \text{CT} + \frac{1}{\text{CR}} \cdot (P_{\text{DRC,dB}}^{\text{in,s}}(k', l) - \text{CT}) + G_{\text{dB}}^s \quad (6.3)$$

The output level in dB SPL is then given by

$$P_{\text{DRC,dB}}^{\text{out,s}}(k', l) = \begin{cases} P_{\text{lin,dB}}(k', l) & \text{if } P_{\text{DRC,dB}}^{\text{in,s}}(k', l) < \text{CT} \\ P_{\text{cp,dB}}(k', l) & \text{if } P_{\text{DRC,dB}}^{\text{in,s}}(k', l) \geq \text{CT} \end{cases} \quad (6.4)$$

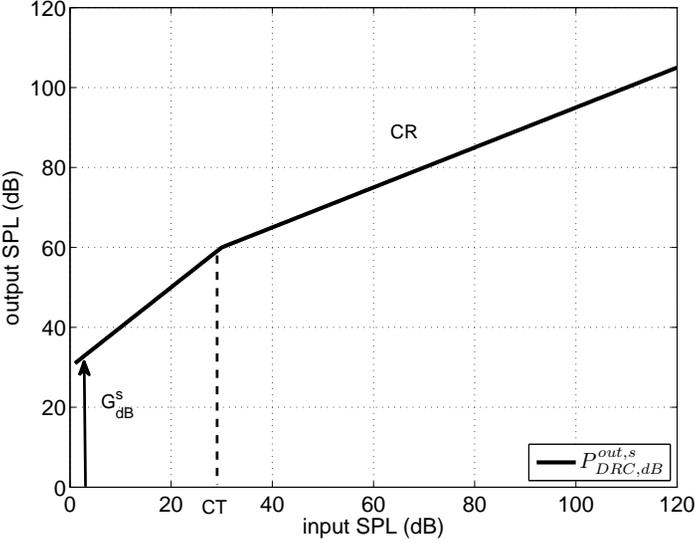


Figure 6.2: DRC curve (CR defines how the slope is changed and CT is the point at which the slope changes).

A DRC curve that shows the output SPL as a function of the input SPL with $CR=2$, $CT=30\text{dB}$ and $G_{dB}^s=30\text{dB}$ is shown in Figure 6.2. Finally the DRC gain in dB is calculated as the output level minus the input level, i.e.,

$$G_{DRC,dB}(k', l) = P_{DRC,dB}^{out,s}(k', l) - P_{DRC,dB}^{in,s}(k', l). \quad (6.5)$$

The attack and release time are then applied to the DRC gain $G_{DRC,dB}(k', l)$ typically using a first-order recursive averaging filter which can be written as

$$\hat{G}_{DRC,dB}(k', l) = \begin{cases} G_{diff}(k', l) \cdot \lambda_{rt} + G_{DRC,dB}(k', l), & \text{if } G_{diff}(k', l) \geq 0 \\ G_{diff}(k', l) \cdot \lambda_{at} + G_{DRC,dB}(k', l), & \text{else} \end{cases} \quad (6.6)$$

where λ_{rt} and λ_{at} represents the release (rt) and the attack (at) time, respectively, and

$$G_{diff}(k', l) = \hat{G}_{DRC,dB}(k' - 1, l) - G_{DRC,dB}(k', l). \quad (6.7)$$

represent the difference between the DRC gain estimated in (6.5) and (6.6). A more detailed illustration of the DRC block in Figure 6.1 is shown in Figure 6.3. This shows that the DRC operates on each critical band independently and the DRC characteristic (input-output mapping) is equal for each critical band. Therefore the DRC only depends on the intensity level (input SPL), i.e., the DRC curve shown in Figure 6.2 does not change its characteristic depending on speech dominant segments or noise dominant segments.

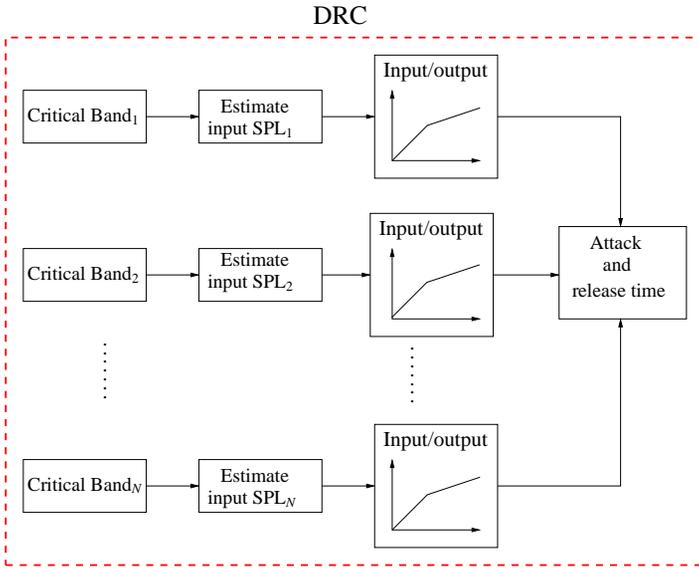


Figure 6.3: Illustration of the overall steps in the multi-band DRC

6.2 The effect of background noise on DRC

In the past, DRC algorithms have mainly been designed under the assumption that the input signal does not contain background noise which in many cases is not a valid assumption. In this section, the effect of background noise on DRC is illustrated through a number of examples where the idea is to show how the DRC reacts to different input SNRs both for speech dominant segments and for noise dominant segments. The evaluation is based on objective measures such as intelligibility-weighted SNR and frequency-weighted log-spectral signal distortion measure.

6.2.1 Undesired amplification over frequencies

Figure 6.4 shows the estimated input SPL and the corresponding DRC gain for each critical band. The DRC gain in this case is estimated with the compression characteristic shown in Figure 6.2. The first observation made is that for a 0dB input SNR the noisy speech input SPL is higher compared to the clean speech input SPL. This leads to a lower DRC gain for the noisy speech, e.g., at critical band 5, 11 and 15 the DRC gain is 11.5dB, 13.15dB and 6.20dB lower compared to the clean speech DRC gain, respectively. This means that for low input SNR the estimated gain may not be sufficient to compensate for the hearing loss. This

could also explain that in the presence of background noise a linear amplification is typically preferred [92][154]. When the input SNR is increased the noisy speech DRC gain approaches the clean speech DRC gain. This shows that for speech dominant segments it is desirable to have the noisy speech as close as possible to the clean speech, which is a typical objective for noise reduction algorithms.

A second observation can be made by using critical band 10 and 14 for the 0dB scenario which suggests that these two critical bands do not contain noise. The problem is that for critical band 10 and 14 the DRC gain is much higher compared to the neighboring critical bands. This can be a major problem if we consider the NR problem and the case where the noise has been reduced for certain frequencies. The DRC will then consider the noise as a low input signal and apply higher amplification compared to the speech which is considered a high input signal. This could lead to certain frequencies containing speech being masked by the noise.

A third observation that can be made is that the short-time spectrum of the speech is bound to be more flat after the DRC has been applied. In [236] it was concluded that multi-band DRC unnecessarily attenuates important information regarding the shape of the short-time speech spectrum, i.e., the height of spectral peaks is reduced and the floor of the spectral valleys is increased which flattens the short-time speech spectrum, resulting in poor speech perception. However this problem is not directly related to the presence of the noise but rather the design of the DRC.

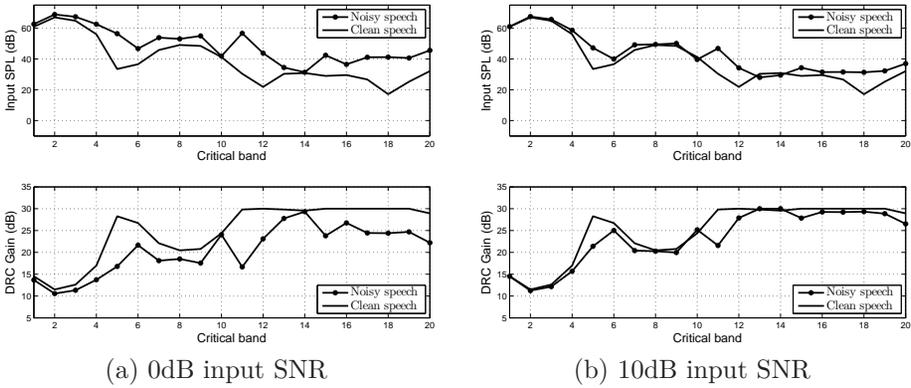


Figure 6.4: Illustration of the estimated input SPL and the corresponding DRC gain for a speech dominant frame.

For a noise dominant segment the situation will be much different as shown in figure 6.5. First of all it is clear that the clean speech input SPL is very low and

therefore from the DRC point of view maximum amplification is applied. When the input SNR increases, e.g., when the input SNR is 10dB, the DRC gain approaches maximum amplification which is highly undesired when the noise is so dominant which also suggests that for higher input SNR the DRC may amplify the noise more compared to low input SNR scenarios. This can potentially lead to a greater SNR degradation. In speech dominant segments the lower critical bands, e.g., 1 to 5 contain a high input SPL and therefore a lower DRC gain is applied. This is the opposite for the noise dominant segments where a much higher DRC gain is applied for the lower critical bands, see Figure 6.4(a) and Figure 6.5(a). This is highly undesired if the noise dominant frame occurs right before a speech dominant frame which again could mask some features in the speech dominant frames.

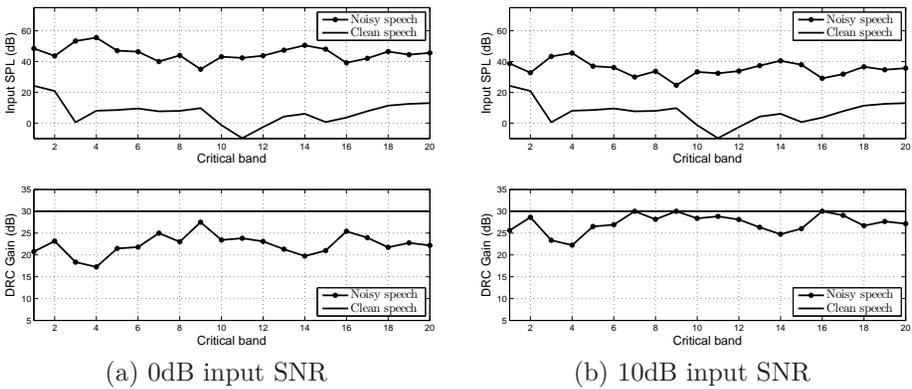


Figure 6.5: Illustration of the estimated input SPL and the corresponding DRC gain for a noise dominant frame.

6.2.2 Undesired amplification over time

In this section we will analyze how the DRC reacts to the background noise by analyzing different critical bands for various frames. The aim is to show how the DRC can negatively affect different frequencies over time, e.g., if certain frequencies are amplified more compared to neighboring frequencies or frames.

The input SPL and the estimated DRC gain for the clean speech, noisy speech, and the noise-only signals are shown in Figure 6.6 and 6.7. An interesting observation can be made using critical band 7 shown in Figure 6.6(e)-(f) which shows that for frames up to 30 the noise is very dominant (noise dominant segments) and for frames between 30-40 the clean speech is mixed with the noise (speech-plus-

noise segments) whereas for frames between 40-50 the speech is dominant (speech dominant segments). Looking at the corresponding DRC gain the problem is that the gain applied to the noise dominant segments is more or less equal to the speech dominant segments. In this case it would be better to actually apply the DRC gain estimated for the noise-only signal. The same observation can be made with critical band 12. Looking at the DRC gain for the clean speech it is also clear that the DRC kind of flattens the level between the low input SPL and high input SPL which is the purpose of the DRC, i.e., making soft sounds audible while avoiding loud sounds becoming too loud. This can have an effect on the speech if the signal information over frequency and time is smoothed. It is also worth noting that at critical band 1 and 4 the speech and the noise are better mixed, since speech contains more energy at lower frequencies, resulting in less difference between the DRC gain.

Intuitively these experiments show that it would be better to apply less gain to the noise dominant segments compared to the speech dominant segments in order not to degrade the SNR. At the same time a smoothing between the noise dominant segments and speech dominant segments could also be avoided if knowledge is available regarding which critical bands contain speech or noise.

6.2.3 Compensation of speech and noise dominant segments

In this section we will analyze how the DRC gain can be modified if the SNR should be preserved. This of course can only be possible if knowledge regarding the speech and the noise-only contributions is available. For the analysis we will exploit the knowledge of the clean speech and the noise-only input SPL as shown in Figure 6.6 and 6.7 to mark out the speech dominant, noise dominant, and speech-plus-noise segments. Once this is done the DRC can be modified in the following way:

- For speech dominant segments set $G_{\text{dB}}^s=30\text{dB}$
- For speech-plus-noise segments set $G_{\text{dB}}^s=25\text{dB}$
- For noise dominant segments set $G_{\text{dB}}^s=20\text{dB}$

This of course would correspond to the hearing aid user manually adjusting the volume based on the acoustic environment but for this analysis the aim is to show how the overall DRC gain over the various frames can be modified to avoid a SNR degradation. The estimated DRC gain for the noisy speech with and without the modified DRC gain is shown in Figure 6.8 for 0dB input SNR together with the distinction of speech dominant segments=1, noise dominant segments=0, and speech-plus-noise segments=0.5.

As an example we can again use critical bands 7 and 12 which show that for frames up to 30 the noise is very dominant and then the following frames indicate that the speech is dominant. Basically in this simple experiment the goal is to show that the gain applied to noise dominant segments should never be higher than the gain applied to the speech dominant segments. The effect of being able to make a distinction between speech and noise and to be able to reduce the gain properly can be illustrated by looking at the average DRC gain applied for the speech dominant segments and for the noise dominant segments. The average DRC gain across all frames for different critical bands is shown in Table 6.1. This shows that with the modified DRC gain the noise receives much less amplification but an important observation here is that with the noisy speech with the standard DRC the gain is almost equal independently of whether speech or noise is present. This problem is further emphasized with the DRC estimated on the clean speech which shows that the noise dominant segments at all time receive higher amplification compared to the speech dominant segments.

In practice the clean speech and the noise-only signals are of course not available and furthermore it is not practical to manually reduce the gain dependent on whether speech or noise is present. Therefore in Chapter 7 we will present a modified DRC algorithm that automatically can reduce the gain based on exploiting the conditional SPP which has already proven to be valuable in the SDW-MWF based NR algorithms, see Chapter 3 and Chapter 4.

6.3 Experimental results

This section presents the experimental results for the analysis of the background noise on the DRC and to confirm the observations made in Section 6.2.

6.3.1 Experimental set-up

Simulations have been performed with speech signals from the HINT-database [167] and the noise signals consisting of a multi-talker babble from Auditec [5]. The signals are sampled at 16kHz. The DRC is implemented based on 20 critical bands [247] which is realized by using individual FFT bins at low frequencies and by combining FFT bins at higher frequencies [110]. The following parameters are used during the simulations:

- Input level is set to 65dB SPL at the microphone.
- Attack and release time are set to $a_t=10\text{ms}$ and $r_t=20\text{ms}$.
- Compression ratio is varied from $CR=1.5, 2, 2.5$ and 3 .

Modified DRC gain (Noisy speech)			
Critical band	Speech	Speech-plus-noise	Noise
1 (avg. gain)	16.0dB	10.9dB	4.0dB
4 (avg. gain)	15.6dB	14.2dB	8.8dB
7 (avg. gain)	17.9dB	16.6dB	7.7dB
9 (avg. gain)	21.4dB	18.2dB	10.5dB
12 (avg. gain)	15.1dB	17.9dB	13.3dB
15 (avg. gain)	22.5dB	17.5dB	12.4dB
DRC gain (Noisy speech)			
Critical band	Speech	Speech-plus-noise	Noise
1 (avg. gain)	16.0dB	15.9dB	14.0dB
4 (avg. gain)	15.6dB	19.2dB	18.8dB
7 (avg. gain)	17.9dB	21.6dB	17.7dB
9 (avg. gain)	21.4dB	23.2dB	20.5dB
12 (avg. gain)	15.1dB	22.9dB	23.3dB
15 (avg. gain)	22.5dB	22.5dB	22.4dB
DRC gain (Clean speech)			
Critical band	Speech	Speech-plus-noise	Noise
1 (avg. gain)	15.5dB	16.8dB	17.5dB
4 (avg. gain)	15.7dB	19.8dB	23.7dB
7 (avg. gain)	18.2dB	23.1dB	27.4dB
9 (avg. gain)	21.1dB	25.8dB	29.1dB
12 (avg. gain)	15.2dB	26.6dB	29.9dB
15 (avg. gain)	24.9dB	28.4dB	29.8dB

Table 6.1: Average DRC gain for speech dominant, noise dominant, and speech-plus-noise segments.

- Compression threshold is set to $CT=30$ dB.

6.3.2 Analysis procedure

The effect of the background noise on the DRC is analyzed using the procedure shown in Figure 6.9 in which three different experiments will be conducted.

In the first experiment, the DRC gain is based on the noisy speech signal, i.e.,

$$P_{\text{DRC,dB}}^{\text{in}}(k', l) = 20 \log_{10} \left(\frac{|X(k', l)|}{P_{\text{ref}}} \right) \quad (6.8)$$

which is illustrated with the shaded DRC block which is the true signal processing path when background noise is present in the input signal. The estimated DRC

gain based on the noisy speech signal is then applied to the clean speech and the noise-only signal. The SNR improvement is then estimated based on $\hat{\mathbf{X}}^s$ and $\hat{\mathbf{X}}^n$. The signal distortion is estimated between \hat{X}^s and \hat{X}^n , which will be defined later.

In the second experiment, the DRC gain is based on the clean speech, i.e.,

$$\bar{P}_{\text{DRC,dB}}^{\text{in}}(k', l) = 20 \log_{10} \left(\frac{|X^s(k', l)|}{P_{\text{ref}}} \right) \quad (6.9)$$

which is then applied to both $X^s(k', l)$ and $X^n(k', l)$. The purpose here is to show that even when access to the clean speech is available this does not result in the optimal DRC gain.

In the third experiment, a separate DRC is applied to the clean speech as in (6.9) and for the noise-only signal

$$\tilde{P}_{\text{DRC,dB}}^{\text{in}}(k', l) = 20 \log_{10} \left(\frac{|X^n(k', l)|}{P_{\text{ref}}} \right). \quad (6.10)$$

The purpose here is that, since we have access to the clean speech and the noise-only signal we can reduce the hearing aid gain for the noise-only signal such that $G_{\text{dB}}^n < G_{\text{dB}}^s$ where G_{dB}^n is the gain applied to the noise-only signal. The goal is then to reduce G_{dB}^n until the SNR is preserved. The SNR improvement is then estimated based on \bar{X}^s and \bar{X}^n .

6.3.3 Results

The results from the first experiment are shown in Figure 6.10. Overall, the SNR degradation is less significant at low input SNR which is supported by Figure 6.5(a). At high input SNR the noise dominant segments receive maximum amplification which leads to the large SNR degradation. The distortion decreases at high input SNR which also correlates well with Figure 6.4(c) where the DRC gain is approaching the clean speech DRC gain at high input SNR. It is also clear the SNR degradation is worse at higher CR.

The results for the second experiment are shown in Table 6.2. As expected the SNR degradation is even worse when the clean speech DRC gain is applied to the noisy input signal, which again is explained by the excessive amplification of noise dominant segments. Since the DRC is based on the clean speech the input SNR does not play a role here. Basically the results in Table 6.2 correspond to an input SNR of 30dB in Figure 6.10. This shows that even if the NR algorithm is able to significantly improve the SNR this may just be compromised by the DRC.

The results for the third experiment are shown in Figure 6.11 which shows the SNR degradation when $G_{\text{dB}}^n=30\text{dB}$ compared to the case where G_{dB}^n is reduced.

CR	1.5	2	2.5	3
$\Delta\text{SNR}_{\text{intellig}}$	-4.7dB	-6.7dB	-7.9dB	-8.4dB

Table 6.2: The SNR improvement when the clean speech DRC gain is applied to the noisy input signal.

The results show that in order to compensate for the SNR degradation the gain G_{dB}^n for the noise DRC needs to be reduced further at high input SNR. This of course only holds if the input SNR is fixed otherwise the gain G_{dB}^n needs to be changed accordingly.

6.4 Conclusion

In this chapter the DRC is analyzed when operating in the presence of background noise. Typically DRC algorithms are designed under the assumption of clean speech. It is shown that when the DRC gain for each critical band is based solely on the input level a significant SNR degradation is observed which is highly undesired. The SNR degradation is more severe at higher input SNR since low level noise signal in this case the noise are now amplified more compared to the high level speech signals.

On the other hand, when the DRC is based on a low input SNR signal the DRC gain is much lower compared to the DRC gain estimated on the clean speech signal and therefore the estimated DRC gain may be too low to compensate for the hearing loss. Through a number of examples it has also been shown that a multi-band DRC can reduce the shape of the short-time speech spectrum by making the spectrum more flat. Through experiments it has been shown that in order to preserve the SNR, knowledge of the speech and the noise preferably for different frequencies and for different frames is required such that the DRC gain for each critical band can vary depending on the speech and the noise and not only on the input level.

The experimental results show that when background noise is present an SNR degradation of 2-3dB is observed whereas at high input SNR an SNR degradation of 4-8dB is observed depending on the CR. The signal distortion is also higher for low input SNR which shows that the noise indeed has an influence on the DRC since the signal distortion is very low at high input SNR.

To summarize, it has been shown that the background noise indeed has a negative effect on the DRC. First of all a significant SNR degradation is observed and secondly the background noise also has an influence on the level of the estimated DRC gain. This problem can be solved by reducing the noise, i.e., increasing the

SNR. However this is where the interaction between the DRC and the background noise is most severe. The challenge is therefore to estimate the speech dominant segments and the noise dominant segments and to automatically adjust the DRC gain based on this information.

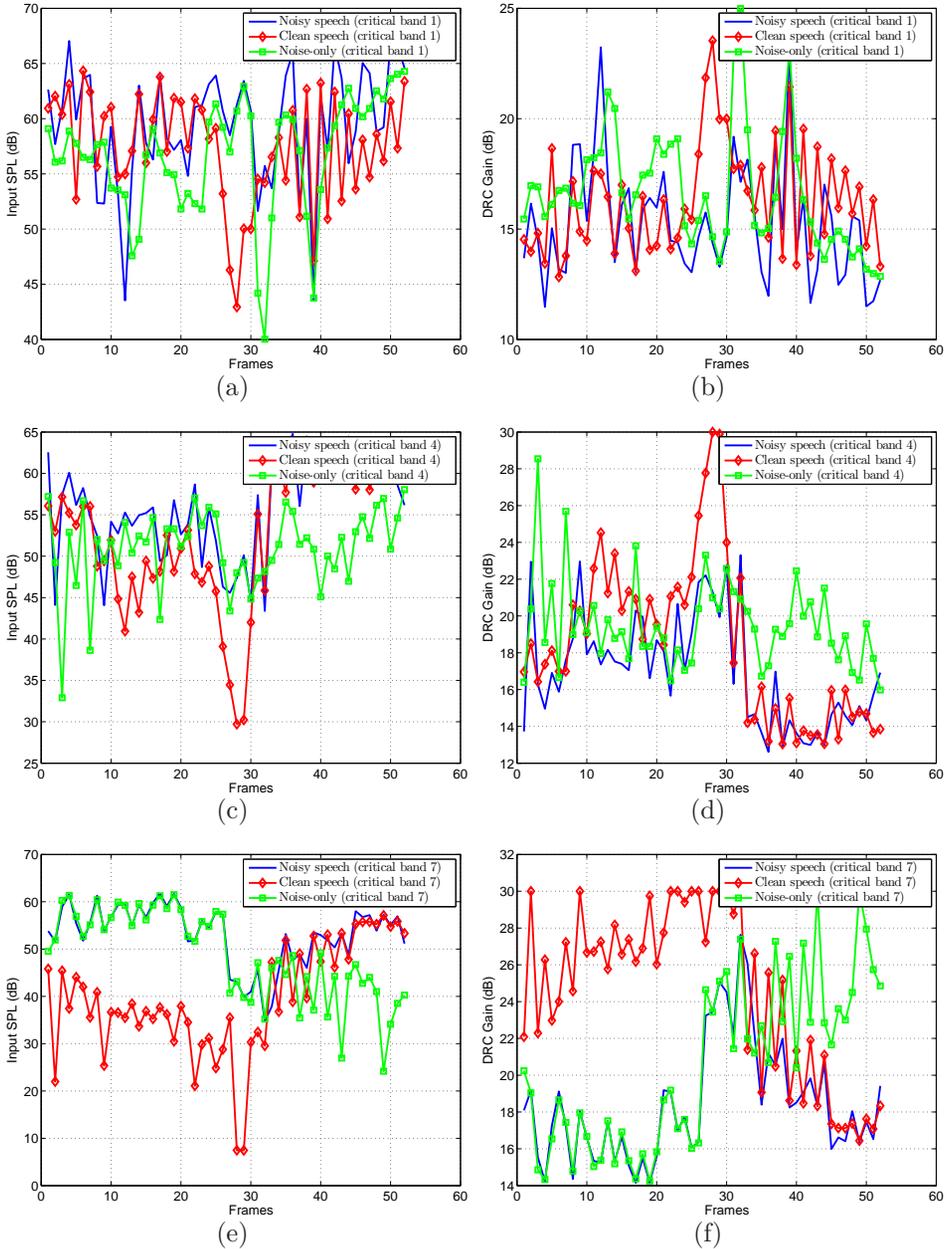


Figure 6.6: Estimated input SPL and DRC gain for the noisy speech, the clean speech, and the noise-only signals for critical band 1, 4, and 7.

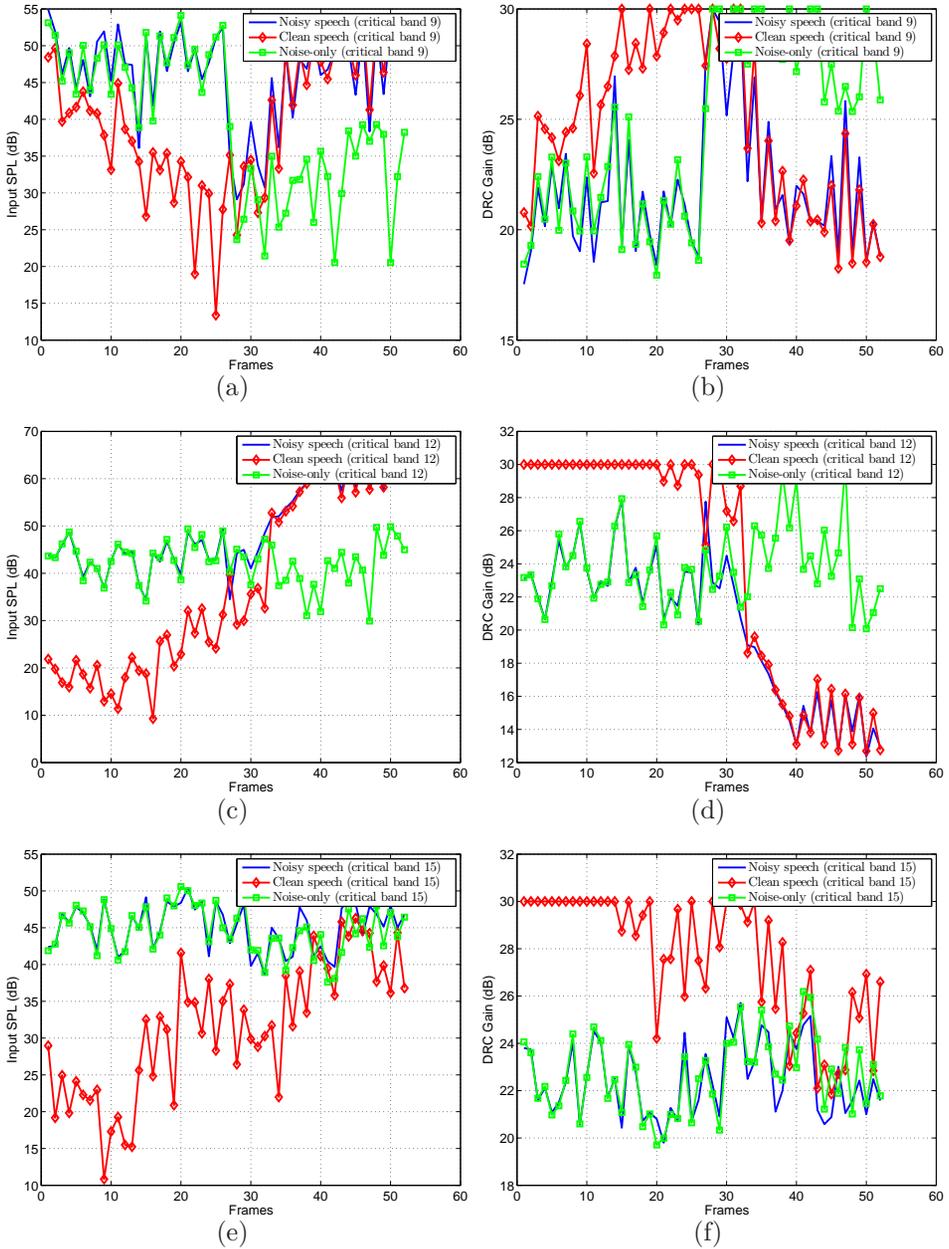


Figure 6.7: Estimated input SPL and DRC gain for the noisy speech, the clean speech, and the noise-only signals for critical band 9, 12, and 15.

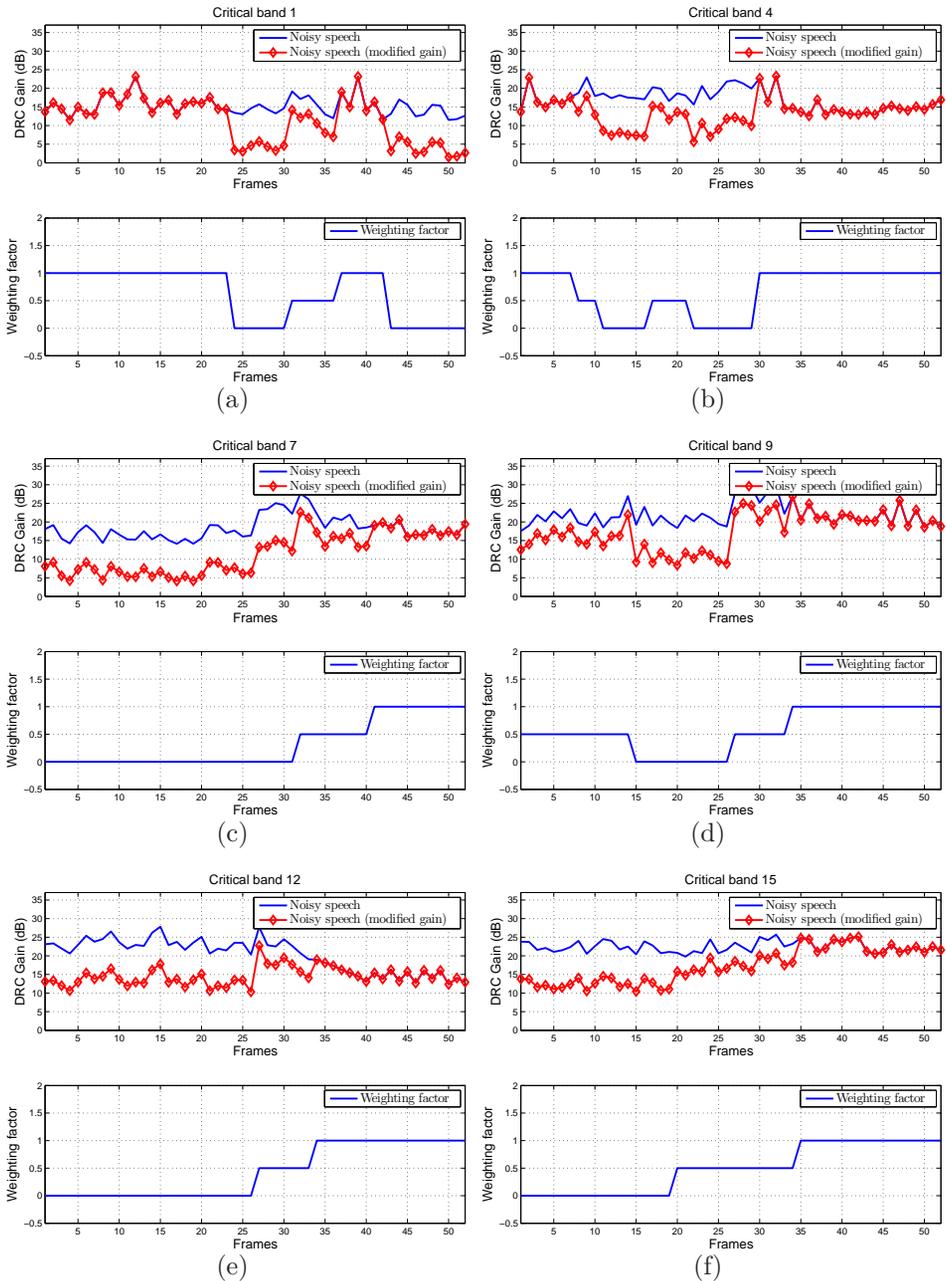


Figure 6.8: Example of compensating for undesired amplification of noise by defining speech dominant segments and noise dominant segments.

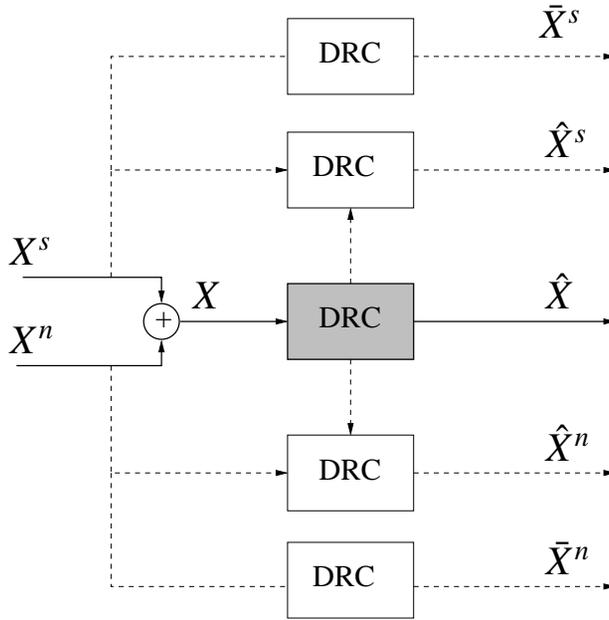


Figure 6.9: Evaluation set-up for analyzing the effect of background noise on DRC.

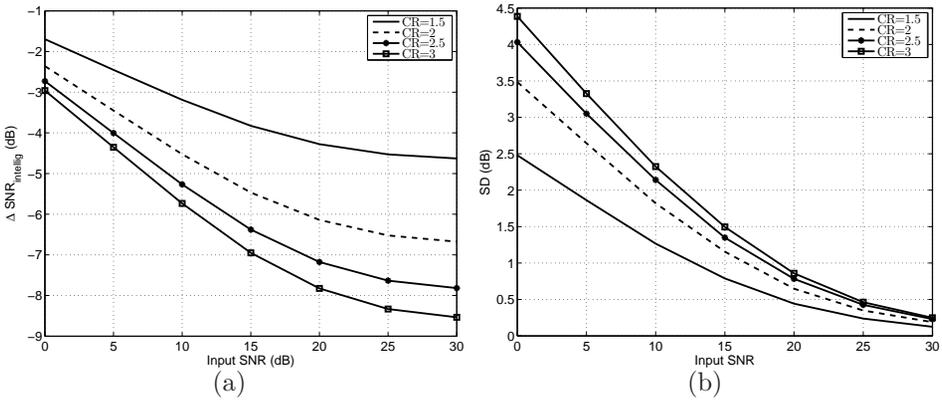


Figure 6.10: SNR improvement and signal distortion for different CR at various input SNR.

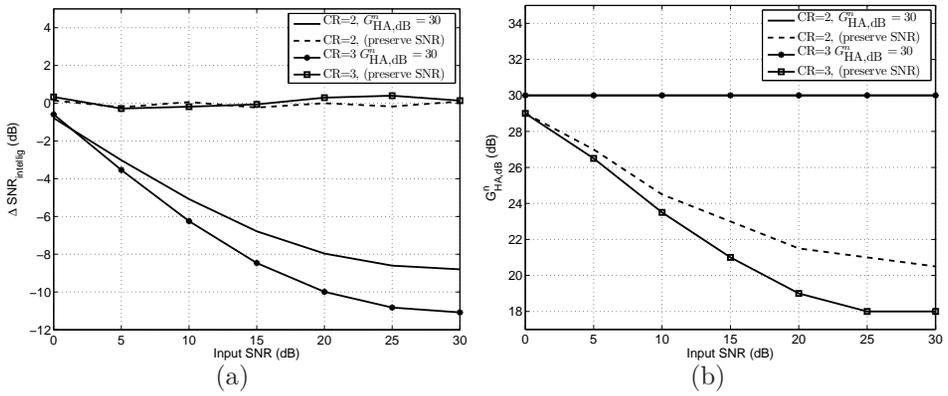


Figure 6.11: SNR preservation when knowledge of clean speech and noise-only contributions are available together with the noise DRC gain required to preserve the SNR.

Chapter 7

SDW-MWF based noise reduction and dynamic range compression

This chapter addresses the issue of having NR and DRC combined. NR and DRC are basic components in hearing aids, but generally these components are developed and evaluated independently of each other. Hearing aids typically use a serial concatenation of NR and DRC. However, the DRC in such a concatenation negatively affects the performance of the NR stage: the residual noise after NR receives more amplification compared to the speech, resulting in an SNR degradation. The integration of NR and DRC has not received a lot of attention so far.

In this work, an MWF based approach is presented for speech and noise scenarios, where an SDW-MWF based NR algorithm is combined with DRC. The proposed solution is based on modifying the SDW-MWF and the DRC to incorporate the conditional SPP in order to avoid residual noise amplification. The approach is analyzed to verify if there are any undesired interaction effects between the NR and the DRC. The work is evaluated by means of objective measures.

Section 7.1 explains the problem statement and motivation for having NR and DRC combined. The evaluation of the combined scheme is also introduced which is based on introducing the concept of using a speech DRC and a noise DRC.

Section 7.2 presents a serial concatenation of a SDW-MWF _{μ} based NR and DRC which will serve as a baseline for the proposed combined approaches.

Section 7.3 presents a combined SDW-MWF_{SPP} based NR and dual-DRC. Here

it is shown how the conditional SPP can be reused from the SDW-MWF_{SPP} such that the standard DRC can be extended to a dual-DRC approach.

Section 7.4 presents a combined SDW-MWF_{Flex} based NR and dual-DRC. Here the concept of a flexible weighting factor from the SDW-MWF_{Flex} is reused to avoid the noise DRC in the dual-DRC compromising the speech DRC.

Section 7.5 presents the experimental results to confirm that the serial concatenation of NR and DRC leads to a performance degradation and to verify that the combined approach is able to compensate for this undesired degradation.

7.1 Problem statement and motivation

The design and benefits of single-channel and multi-channel NR algorithms have been widely studied [15][49][56][74][93][207]. The same goes for the design and evaluation of different DRC algorithms [13][87][110][127][193]. Although sophisticated algorithms for NR and DRC exist there is still a question as to how these algorithms should be combined, which unfortunately, has not received a lot of attention so far. Combining hearing aid algorithms in general is indeed a challenging task since each algorithm can counteract and limit the functionality of other algorithms.

In [2][30] experiments have been conducted to evaluate different combinations of NR and DRC. One of the main conclusions was that a serial concatenation of NR and DRC performs suboptimally due to the interaction effects between the NR and the DRC. In [29] it was shown that the NR algorithm does enhance the modulation depth of a noisy speech but when the DRC is activated the modulation depth of the speech envelope is greatly reduced. This indicates that the noise level is increased compared to the speech level, which is clearly undesirable.

An important issue is the evaluation of such combined and integrated schemes, where the lack of an overall design criterion indeed makes the evaluation more difficult. In the evaluation the crucial question will be as to which effects are most damaging to speech intelligibility, e.g., the amount of background noise, distortion or the audibility.

When combining NR and DRC a main problem is that each algorithm serves a different purpose. The objective of the NR algorithm in speech and noise scenarios is to maximally reduce the noise while minimizing speech distortion, e.g., based on temporal, spectral and spatial signal information. The DRC on the other hand is designed to amplify sounds based on their intensity level and a compression characteristic. Figure 7.1(a)-(b) shows the two ways to serially concatenate NR and DRC. The main issues can be stated as follows:

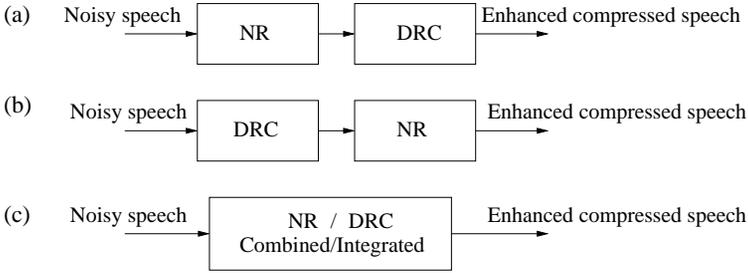


Figure 7.1: NR and DRC in a serial concatenation compared to an integrated scheme.

- When NR is performed before DRC, as in Figure 7.1(a), the residual noise receives more amplification compared to the speech, which consequently defeats the purpose of using NR. From a DRC point of view there is no distinction between speech dominant segments and noise dominant segments, so all low intensity segments are amplified equally. This means that the reduced noise level, from the DRC point of view, is considered a low level signal which is then amplified while the speech is considered a high level signal, and receives less amplification. This leads to the undesired SNR degradation.
- When DRC is performed before NR, as in Figure 7.1(b), the DRC can negatively affect the NR especially so in a multi-channel NR where the correlation between the microphone signals can be affected by the independent DRC on the microphone signals. Furthermore in this set-up the DRC is based on the speech-plus-noise level rather than the speech-plus-residual noise level and so the applied gain in this case may be too small to make the soft speech segments audible.

To avoid any undesired interaction effects it is desirable to combine NR and DRC in an integrated scheme, as in Figure 7.1(c), which is the goal of this work. In the sequel, the serial concatenation shown in Figure 7.1(a) will serve as a reference system and the proposed solution will be referred to as the combined approach.

A combined NR and DRC system that could be viewed as the ideal system is shown in Figure 7.2. The idea here is that if the clean speech and the noise-only contribution can be perfectly extracted then a speech DRC can be applied to the clean speech and a noise DRC to the noise-only contribution. The gain difference between the speech DRC and the noise DRC indicates a target noise suppression which means that the noise DRC gain can be set to zero, i.e., to suppress all noise, or it can be a scaled version of the speech DRC gain, i.e., $G_{dB}^n < G_{dB}^s$. The gain

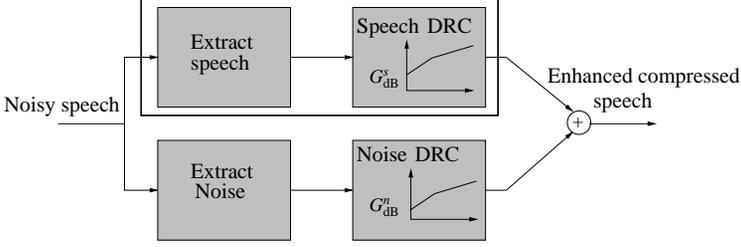


Figure 7.2: Ideal system where the speech and the noise-only contributions are extracted which then are compressed separately.

difference between the speech DRC curve and the noise DRC curve is defined as

$$\Delta G_{\text{dB}} = G_{\text{dB}}^s - G_{\text{dB}}^n. \quad (7.1)$$

Finally, the overall output signal is the sum of the two compressed components. Since the ideal case does not contain residual noise then the SNR will improve when the noise DRC gain G_{dB}^n decreases compared to the speech DRC gain G_{dB}^s . The goal is then to compare the performance of the combined approach against this ideal performance, and any deviation from this will be considered as an undesired effect of having a NR and DRC combined.

Next we presents three different approaches to combine a SDW-MWF based NR and DRC. A SDW-MWF $_{\mu}$ serially concatenated with a DRC is described first and is considered to be the baseline system. The SDW-MWF $_{\mu}$ is then replaced by the SDW-MWF $_{\text{SPP}}$ and SDW-MWF $_{\text{Flex}}$ together with a dual-DRC approach.

7.2 Combined SDW-MWF $_{\mu}$ based NR and DRC

First the perfect extraction of the clean speech and the noise-only contribution in Figure 7.2 is replaced with a SDW-MWF $_{\mu}$ based NR. The estimated speech component can then be written as

$$\begin{aligned} Z^s(k, l) &= \mathbf{W}^H(k, l)(\mathbf{X}^s(k, l) + \mathbf{X}^n(k, l)) \\ &= Z^{ss}(k, l) + Z^{sn}(k, l) \end{aligned} \quad (7.2)$$

where $Z^{ss}(k, l)$ is the speech component in $Z^s(k, l)$ and $Z^{sn}(k, l)$ is residual noise. This is where the usual problem with a cascade of NR and DRC appears since the estimated speech component $Z^s(k, l)$ is indeed bound to have residual noise, which then could be amplified by the DRC. i.e.,

$$\hat{Z}^s(k, l) = Z^s(k, l)G_{\text{DRC, dB}}^s(k, l). \quad (7.3)$$

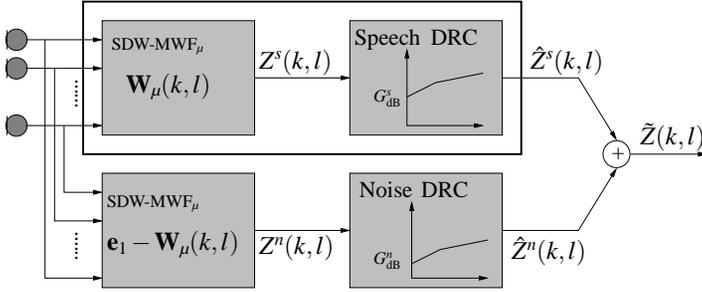


Figure 7.3: A serial concatenation of a SDW-MWF_μ based NR and DRC.

Any such residual noise, from the speech DRC point of view, is now considered a low level signal which is then amplified, while the actual speech component is considered a high level signal which is then compressed. This leads to the undesired SNR degradation.

In a similar manner the noise component $X_1^n(k, l)$ in the first microphone signal can be estimated with a SDW-MWF_μ given as

$$\begin{aligned} \mathbf{V}_\mu(k, l) &= (\mathbf{R}^s(k, l) + \mu \mathbf{R}^n(k, l))^{-1} \mu \mathbf{R}^n(k, l) \mathbf{e}_1 \\ &= \mathbf{e}_1 - \mathbf{W}_\mu(k, l) \end{aligned} \quad (7.4)$$

which leads to the estimated noise component

$$\begin{aligned} Z^n(k, l) &= \mathbf{V}_\mu^H(k, l) \mathbf{X}(k, l) \\ &= X_1(k, l) - Z^s(k, l). \end{aligned} \quad (7.5)$$

The combined SDW-MWF based NR and DRC is shown in Figure 7.3. At this point it is important to emphasize that the main challenge is the estimation of the speech component, which is shown with the solid box in Figure 7.3. On the other hand, the estimated noise component $Z^n(k, l)$ in (7.5) is better controlled since the noise DRC can be set to zero, i.e., to suppress all noise, or it can be a scaled version of the speech DRC as explained in section 7.1.

The speech and the noise problem in DRC is shown with an example, see Figure 7.4 where the speech and the noise input SPL are located at 50dB and 30dB, respectively. This shows that with the given DRC curve the output SPL between the speech and the noise is reduced by 10dB which is obviously undesired.

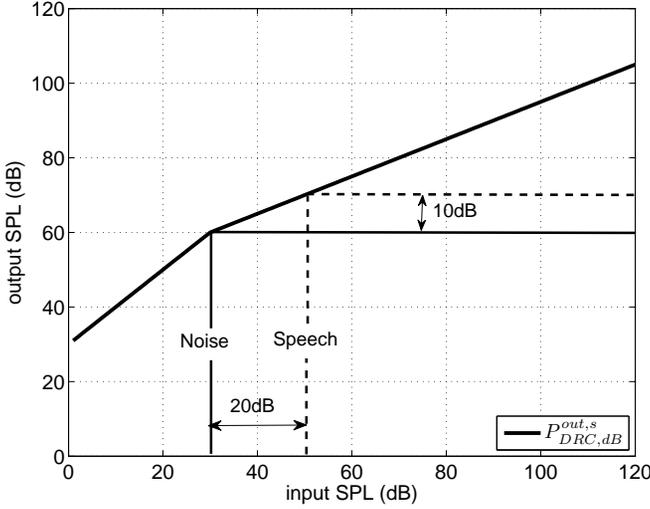


Figure 7.4: Illustration of the output SPL after the DRC with the noise located at 30dB input SPL and the speech at 50dB SPL.

7.3 Combined SDW-MWF_{spp} based NR and dual-DRC

The DRC described in Chapter 6 amplifies signals based on their intensity level and makes no distinction between speech dominant segments and noise dominant segments. The aim could then be to identify the speech dominant segments and the noise dominant segments such that the residual noise amplification can be avoided. By reusing the conditional SPP $p(k, l)$ estimated in the SDW-MWF_{spp} a dual-DRC approach is introduced such that a different DRC curve is applied to the speech dominant segments and to the noise dominant segments. The two DRC curves are defined similarly as in (6.2)-(6.4) and the overall DRC output power is then defined as

$$P_{\text{dual-DRC,dB}}^{\text{out}}(k, l) = p(k, l) \cdot P_{\text{DRC,dB}}^{\text{out,s}}(k, l) + (1 - p(k, l)) \cdot P_{\text{DRC,dB}}^{\text{out,n}}(k, l) \quad (7.6)$$

where $P_{\text{DRC,dB}}^{\text{out,s}}(k, l)$ and $P_{\text{DRC,dB}}^{\text{out,n}}(k, l)$ are defined by the speech DRC curve and the noise DRC curve, respectively. The dual-DRC gain is then defined as

$$G_{\text{dual-DRC,dB}}(k, l) = P_{\text{dual-DRC,dB}}^{\text{out}}(k, l) - P_{\text{DRC,dB}}^{\text{in,s}}(k, l). \quad (7.7)$$

The dual-DRC approach is illustrated in Figure 7.5 with an example where the input SPL is 60dB and the output SPL now depends on the conditional SPP $p(k, l)$. The procedure is as follows:

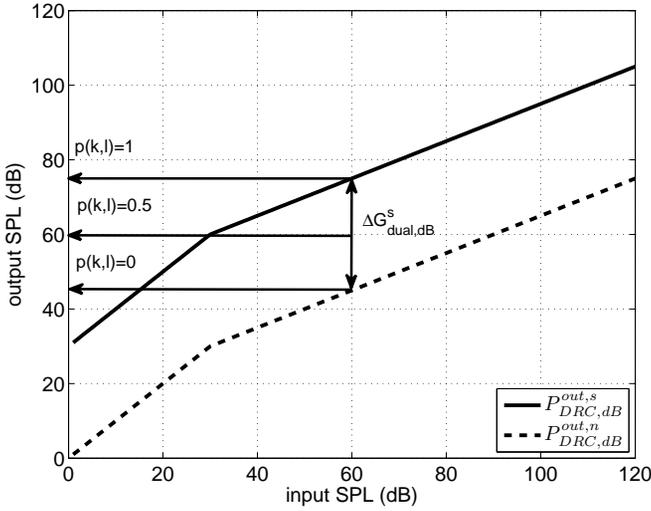


Figure 7.5: Dual-DRC with the conditional speech presence probability $p(k, l)$ to provide a weighting between the two DRC curves.

- If speech is present ($p(k, l)=1$) the speech DRC curve is applied.
- If speech is absent ($p(k, l)=0$) it is undesirable to amplify the residual noise compared to the speech and therefore a lower gain is applied, i.e., the noise DRC curve is applied.
- For the in-between cases a weighted sum of the two DRC curves is used.

The rationale behind the noise DRC curve is that it results in a lower gain compared to the speech DRC curve, as the goal indeed is to apply a lower gain to the noise dominant segments compared to the speech dominant segments.

The proposed MWF based NR and dual-DRC using SDW-MWF_{SPP} is shown in Figure 7.6. The main difference between this approach and the MWF based NR and DRC using SDW-MWF_μ is that the speech DRC in Figure 7.3 implicitly assumes that the estimated speech component does not contain residual noise. The gain difference between the noise DRC curve and the speech DRC curve in the dual-DRC is given by

$$\Delta G_{\text{dual,dB}} = G_{\text{dB}}^s - G_{H_1,\text{dB}}^n \tag{7.8}$$

where $G_{H_1,\text{dB}}^n$ is the noise DRC curve in the dual-DRC approach. Based on the example given in Figure 7.4 it is shown in Figure 7.7 that the noise DRC gain $G_{H_1,\text{dB}}^n$ has to be 10dB lower than G_{dB}^s to compensate for the 10dB reduction

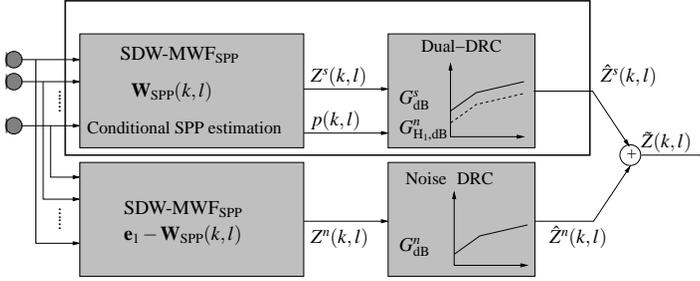


Figure 7.6: A combined approach of a SDW-MWF_{SPP} based NR and dual-DRC.

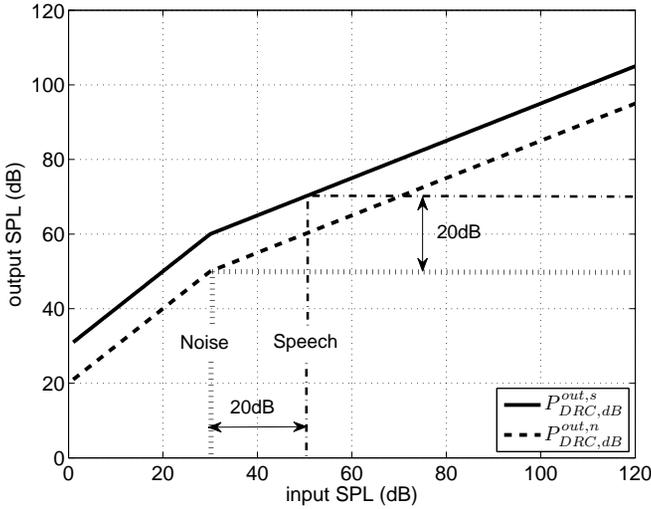


Figure 7.7: Illustration of the output SPL after the dual-DRC with the noise located at 30dB input SPL and speech at 50dB SPL.

between the speech and the noise output SPL. The properties of $G_{H_1,dB}^n$ can be summarized as follows:

- If $G_{H_1,dB}^n$ is set too low the desired hearing aid gain G_{dB}^s may be compromised.
- If $G_{H_1,dB}^n$ is set too high the impact of $p(k,l)$ may be too small to compensate for the residual noise amplification.

The goal of the dual-DRC is thus to find a proper trade-off between NR and DRC, i.e., SNR improvement and the desired DRC gain.

7.4 Combined SDW-MWF_{flex} based NR and flex dual-DRC

Following the above discussion it is desirable to minimize the term in (7.8) without sacrificing the SNR improvement. This can be achieved by not only using the conditional SPP $p(k, l)$ introduced in the SDW-MWF_{SPP} but also the H_0 and H_1 state detection $P(l)$ introduced in the SDW-MWF_{Flex}. A flexible dual-DRC can then be written as

$$\begin{aligned}
 P_{\text{flex-DRC,dB}}^{\text{out}}(k, l) &= P(l) \left[p(k, l) P_{\text{DRC,dB}}^{\text{out,s}}(k, l) + (1 - p(k, l)) P_{\text{DRC,dB}}^{\text{out,n}}(k, l) \right] \\
 &+ (1 - P(l)) P_{\text{DRC,dB}}^{\text{out,n}}(k, l) \\
 &= \begin{cases} H_1 : p(k, l) P_{\text{DRC,dB}}^{\text{out,s}}(k, l) + (1 - p(k, l)) P_{\text{DRC,dB}}^{\text{out,n}}(k, l) \\ H_0 : P_{\text{DRC,dB}}^{\text{out,n}}(k, l) \end{cases} \quad (7.9)
 \end{aligned}$$

where the noise DRC curve $P_{\text{DRC,dB}}^{\text{out,n}}(k, l)$ in the H_1 and H_0 states can be either similar or in the H_0 state the gain can be set lower. The flexible dual-DRC gain is given by

$$G_{\text{flex-DRC,dB}}(k, l) = P_{\text{flex-DRC,dB}}^{\text{out}}(k, l) - P_{\text{DRC,dB}}^{\text{in,s}}(k, l). \quad (7.10)$$

The rationale behind the flexible dual-DRC is:

- When a H_1 state is detected, i.e., $P(l)=1$, a dual-DRC is applied using G_{dB}^s and $G_{H_1,\text{dB}}^n$.
- When a H_0 state is detected, i.e., $P(l)=0$, a DRC is applied with $G_{H_0,\text{dB}}^n \leq G_{H_1,\text{dB}}^n$.

The DRC gain difference between the noise DRC curve and the speech DRC curve in the flexible dual-DRC is then given by

$$\begin{aligned}
 \Delta G_{\text{flex,dB}} &= P(l) \left[G_{\text{dB}}^s - G_{H_1,\text{dB}}^n \right] + (1 - P(l)) G_{H_0,\text{dB}}^n \\
 &= \begin{cases} H_1 : G_{\text{dB}}^s - G_{H_1,\text{dB}}^n \\ H_0 : G_{H_0,\text{dB}}^n \end{cases} \quad (7.11)
 \end{aligned}$$

The proposed MWF based NR and the flexible dual-DRC using SDW-MWF_{Flex} is shown in Figure 7.8.

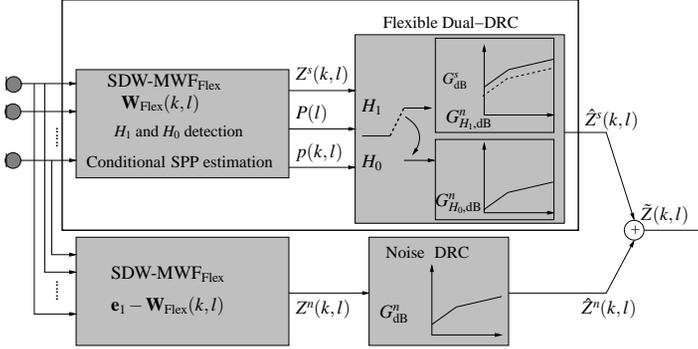


Figure 7.8: A combined approach of a SDW-MWF_{Flex} based NR and a flexible dual-DRC.

7.5 Experimental results

In this section, experimental results for the combined approaches are presented. The simulations aim at showing the undesired interaction effects when a MWF based NR and DRC are serially concatenated, and to compare this approach to the proposed combined approaches using the introduced dual-DRC.

7.5.1 Experimental set-up

Both the MWF based NR and the DRC are implemented using an FFT length of 128 with half overlapping frames. The DRC is implemented based on critical bands [247] which is realized by using individual FFT bins at low frequencies and by combining FFT bins at higher frequencies [110]. The following parameters are fixed during all simulations:

- Input level is set to 65dB SPL at the microphone.
- Attack and release time are set to $at=10\text{ms}$ and $rt=20\text{ms}$.
- Compression ratio $CR=2$.
- Compression threshold is set to $CT=30\text{dB}$.

In order to evaluate the effect from the DRC on the different SDW-MWF based NR algorithms and to make a fair comparison each SDW-MWF algorithm is adjusted such that the SNR improvement and SD are as similar as possible, see table 7.1.

Method	SDW-MWF _{μ}	SDW-MWF _{SPP}	SDW-MWF _{Flex}
Input SNR	0dB	0dB	0dB
Δ SNR	13.1dB	13.2dB	13.9dB
SD	4.2dB	4.3dB	4.2dB

Table 7.1: SNR improvement and SD of the different SDW-MWF based NR.

	G_{dB}^s (7.1)	G_{dB}^n (7.1)	$G_{H_1,\text{dB}}^n$ (7.8)	$G_{H_0,\text{dB}}^n$ (7.11)
SDW-MWF _{μ}	30dB	0	N/A	N/A
SDW-MWF _{SPP}	30dB	0	10dB-30dB	N/A
SDW-MWF _{Flex}	30dB	0	27.5dB	20dB-25dB

Table 7.2: Gain settings for first experiment.

Figure 7.9 shows how the signals used in the objective measures is estimated. In the first experiment the aim is to show the degradation of the NR performance which means that the signals \bar{X}^s and \bar{X}^n are compared to the input signals X^s and X^n . The signal distortion is estimated between \hat{X}^s and \bar{X}^s .

In the second experiment, the performance of the different schemes is compared to the ideal performance which is shown with the signals \check{X}^s and \check{X}^n . This is an ideal performance since the DRC gain applied to the X^s does not contain noise and the SNR would then improve when the DRC gain is reduced for X^n . This ideal performance is then compared to $\hat{Z}^s = \bar{X}^s + \bar{X}^n$ and $\hat{Z}^n = \check{X}^s + \check{X}^n$.

7.5.2 Results

The gain settings in the first experiment are shown in Table 7.2. Notice that G_{dB}^n is set to zero since the aim is to show the effect of the DRC on the SNR improvement for the NR performance shown in Table 7.1. The results for these experiments are shown in Figure 7.10 and 7.11. This shows that the DRC degrades the SNR improvement of the SDW-MWF _{μ} and the SDW-MWF_{SPP} by 6dB which is illustrated at $\Delta G_{\text{dual,dB}}=0\text{dB}$ compared to Table 7.1. The dotted line shows the SNR improvement for SDW-MWF_{SPP} and dual-DRC as a function of $\Delta G_{\text{dual,dB}}$. Better performance is achieved when $\Delta G_{\text{dual,dB}}$ increases as this increases the impact of the dual-DRC. The SDW-MWF_{Flex} based NR and the flexible dual-DRC is seen to achieve a larger SNR improvement at a small increase in SD as low as 1dB.

The gain settings in the second experiment are shown in Table 7.3. In this

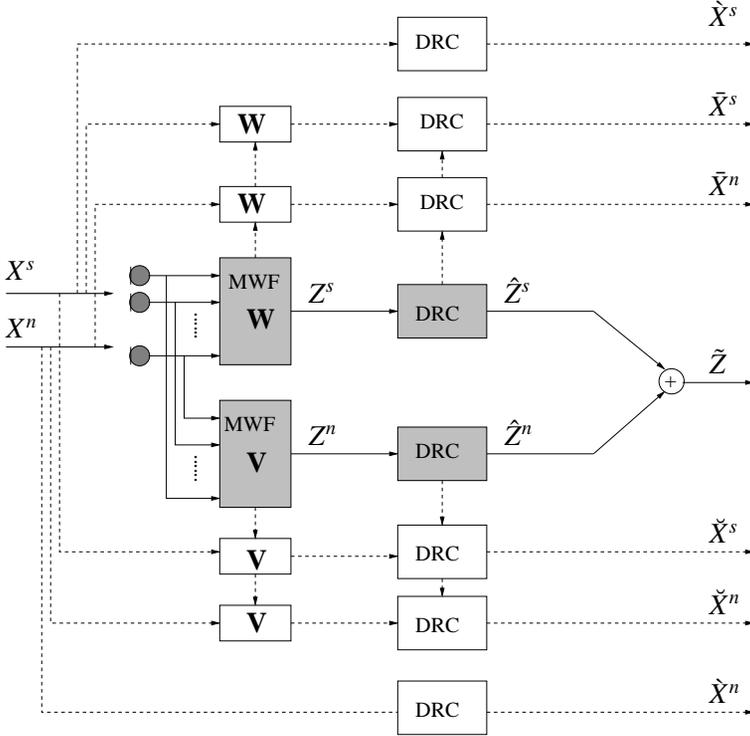


Figure 7.9: Illustration of the signals used in estimating the objective measures.

	G_{dB}^s (7.1)	G_{dB}^m (7.1)	$G_{H_1, \text{dB}}^m$ (7.8)	$G_{H_0, \text{dB}}^m$ (7.11)
SDW-MWF $_{\mu}$	30dB	0dB-30dB	N/A	N/A
SDW-MWF $_{\text{SPP}}$	30dB	0dB-30dB	20dB-27.5dB	N/A
SDW-MWF $_{\text{Flex}}$	30dB	0dB-30dB	27.5dB	20dB-25dB

Table 7.3: Gain settings for second experiment.

experiment, the performance of the different schemes is compared to the ideal performance, i.e., when the speech DRC is applied to the clean speech and the noise DRC is applied to the noise-only signal, see section 7.1. The results for these experiments are shown in Figure 7.12 and 7.13. The dashed line shows the ideal output SNR which as expected improves when ΔG_{dB} is increased. For the combined schemes the output SNR also improves but at the same time it is clear that the SNR improvement is smaller than in the ideal case. This happens because the signal filtered by $\mathbf{W}_{\mu}(k, l)$, $\mathbf{W}_{\text{SPP}}(k, l)$ and $\mathbf{W}_{\text{Flex}}(k, l)$ contains residual noise

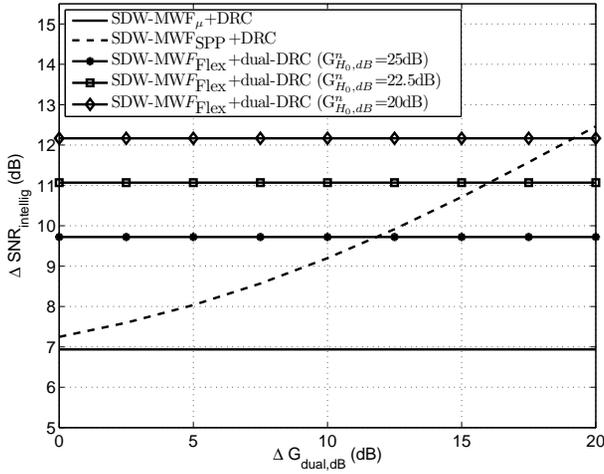


Figure 7.10: The SNR improvement for the different SDW-MWF based NR and DRC algorithms.

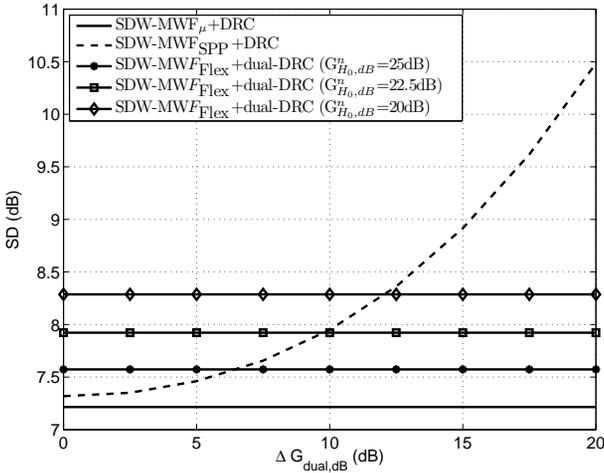


Figure 7.11: The distortion for the different SDW-MWF based NR and DRC algorithms.

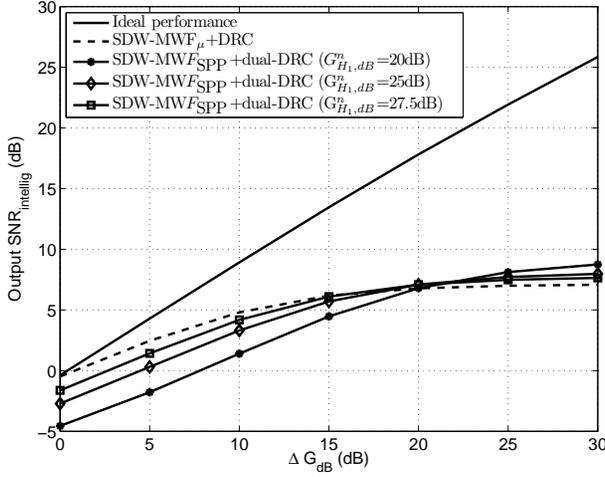


Figure 7.12: The effect of ΔG_{dB} on the output SNR for SDW-MWF $_{\mu}$ and SDW-MWF $_{SPP}$ based NR and DRC.

which subsequently receives more amplification compared to the speech. It is also worth noting that when $\Delta G_{dB} < 20dB$ the output SNR is higher for SDW-MWF $_{\mu}$ and DRC which is due to the fact that the overall gain with the DRC is higher than with the dual-DRC gain. Using the flexible dual-DRC improves the output SNR but it is still far from the ideal performance.

7.6 Conclusion

In this chapter, the undesired interaction effects in a serial concatenation of a MWF based NR and DRC are analyzed. First of all it is shown that having a traditional SDW-MWF $_{\mu}$ based NR and DRC leads to a SNR degradation. The reason for this is that a traditional DRC only uses the input level of a signal segment to estimate the gain independently of whether a speech dominant segment or a noise dominant segment is present. This is highly undesirable since this consequently defeats the purpose of having a NR algorithm, as the residual noise receives more amplification compared to the speech after the NR stage.

The combined solutions proposed here are based on two modifications both in the MWF based NR process and in the DRC. The first modification is to incorporate the conditional SPP in the NR process, which is referred to as SDW-MWF $_{SPP}$. Using the conditional SPP serves the purpose of identifying the speech dominant segments and the noise dominant segments. The second modification is based on

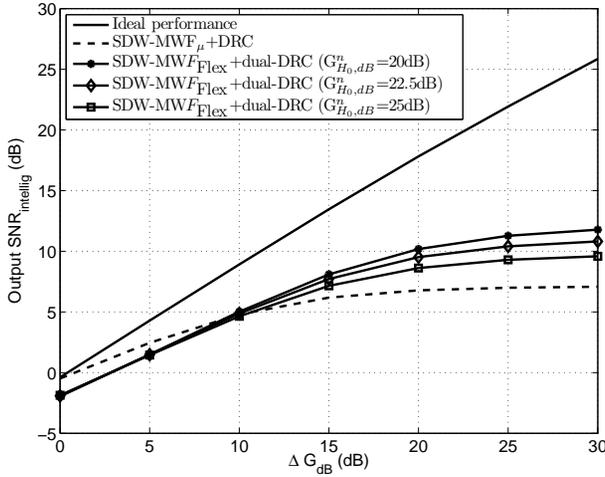


Figure 7.13: The effect of ΔG_{dB} on the output SNR for SDW-MWF $_{\mu}$ and SDW-MWF $_{Flex}$ based NR and DRC.

reusing the conditional SPP estimated in the SDW-MWF $_{SPP}$ to change the DRC into a dual-DRC that incorporates the conditional SPP. The dual-DRC uses two compression curves instead of one compression curve in a traditional DRC. The two compression curves allow for a switchable compression characteristic based on the conditional SPP, i.e., a smaller gain is applied to the noise dominant segments whereas in the speech dominant segments the aim is to apply a gain similar to a traditional DRC. Another solution has been proposed based on the SDW-MWF $_{Flex}$ where the conditional SPP is combined with a H_0 and H_1 state detection in order to incorporate a flexible weighting factor in the NR process. This principle has been reused in the dual-DRC referred to as the flexible dual-DRC. The idea is that in the H_0 a lower gain can be applied without compromising the amplification in the H_1 state.

Experimental results indeed confirm that a serial concatenation of NR and DRC degrades the SNR improvement provided by the NR, whereas the combined approach proposed here shows less degradation of the SNR improvement at a low increase in distortion compared to a serial concatenation. It has been shown that even with the SDW-MWF $_{SPP}$ and the dual-DRC the SNR is still far from the ideal performance. The approach with the SDW-MWF $_{Flex}$ and the flexible dual-DRC is able to get closer to the ideal performance. Therefore the integration of NR and DRC still remains an open problem and in this work we have shown how significant the NR performance can be degraded.

Chapter 8

Prediction error method-based adaptive feedback cancellation

This chapter introduces the prediction error method-based adaptive feedback cancellation (PEM-based AFC) together with the idea of using a near-end signal model. In [208] four commercial hearing aids were evaluated and compared to the PEM-based AFC [209][210]. It was shown that the PEM-based AFC offered a high added stable gain (ASG) compared to certain commercial hearing aids. However the PEM-based AFC was more sensitive towards tonal input signals. This is mainly due to the near-end signal model used which in this case was based on linear prediction (LP). For this reason a cascaded near-end signal model is introduced.

The notation related to PEM-based AFC will be given together with the evaluation using objective measures such as maximum stable gain (MSG) and filter misadjustment which differ from the objective measures used to evaluate the NR algorithms. Furthermore the advantage and disadvantage of the current PEM-based AFC will be discussed and the motivation for further research on the PEM-based AFC will be explained.

Section 8.1 introduces the idea behind AFC with emphasis on the PEM which we will refer to as PEM-based AFC. It is also explained why the PEM-based AFC leads to an unbiased estimate of acoustic feedback path.

Section 8.2 introduces the use of a near-end signal model based on LP. Furthermore it is discussed why an LP model in certain cases will fail to provide a proper decorrelation when the near-end signal is tonal such as speech and music.

Section 8.3 introduces a cascaded near-end signal model based on LP and a

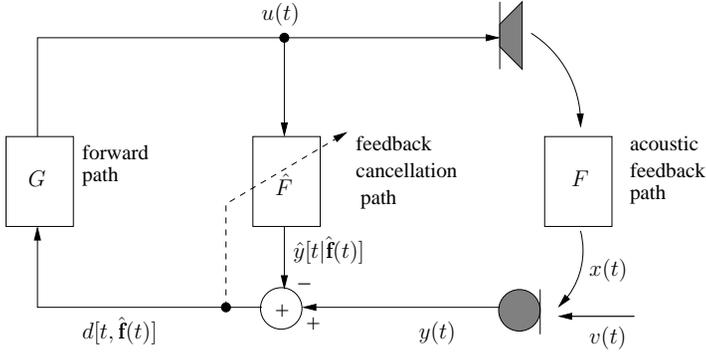


Figure 8.1: Concept of an adaptive feedback cancellation (AFC) algorithm.

pole zero LP (PZLP) model. The idea is that the LP can model the noise component whereas the PZLP can model the tonal components. In this way a better decorrelation effect is achieved with tonal input signals. Section 8.4 presents the experimental results for the PEM-based AFC using a single near-end signal model and a cascaded near-end signal model. These results serve as the baseline for further developments on the PEM-based AFC.

8.1 Adaptive feedback cancellation (AFC)

The adaptive feedback cancellation concept is set-up in Figure 8.1. The microphone signal is given by

$$y(t) = v(t) + x(t) = v(t) + F(q, t)u(t) \tag{8.1}$$

where q denotes the time shift operator and t is the discrete time variable. $F(q, t)$ is the feedback path between the loudspeaker and the microphone, $v(t)$ is the near-end signal, $x(t)$ is the feedback signal. The forward path $G(q, t)$ maps the microphone signal $y(t)$, possibly after AFC, to the loudspeaker signal $u(t)$. It typically consists of an amplifier with time-varying gain $K(t)$ cascaded with a linear equalization filter $J(q, t)$, such that

$$G(q, t) = K(t)J(q, t). \tag{8.2}$$

The aim of the AFC is to place an estimated finite impulse response (FIR) adaptive filter $\hat{F}(q, t)$ in parallel with the feedback path, having the loudspeaker signal as input and the microphone signal as the desired output. The feedback canceller $\hat{F}(q, t)$ produces an estimate of the feedback signal $x(t)$ which is then subtracted

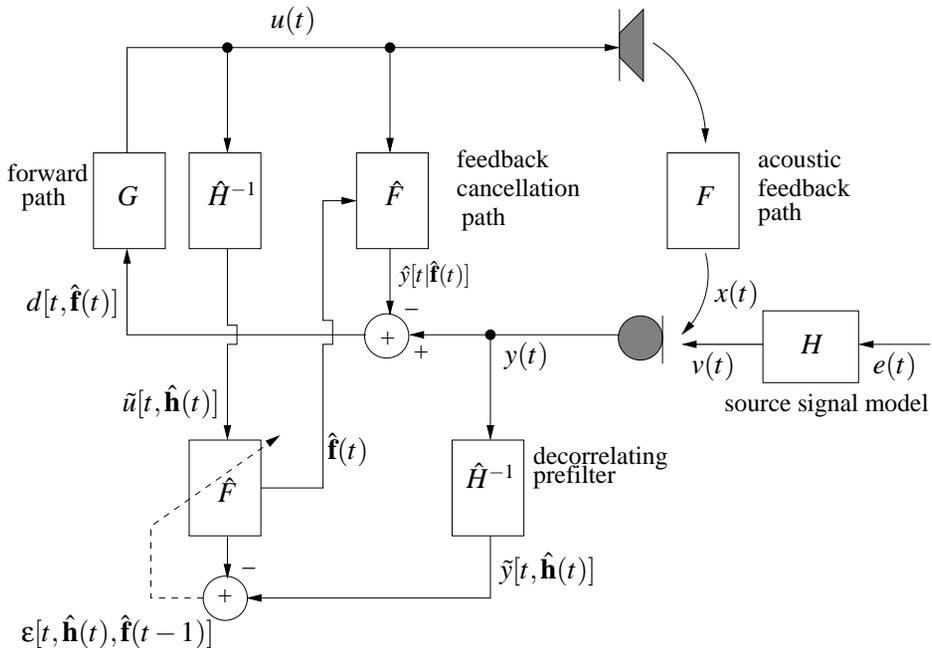


Figure 8.2: AFC with decorrelating prefilters in the adaptive filtering circuit.

from the microphone signal $y(t)$. The feedback-compensated signal is given by

$$d(t) = v(t) + [F(q, t) - \hat{F}(q, t)]u(t). \quad (8.3)$$

8.1.1 Prediction error method

The main problem in identifying the feedback path model is the correlation between the near-end signal and the loudspeaker signal, due to the forward path $G(q, t)$, which causes standard adaptive filtering algorithms to converge to a biased solution [229][85]. This means that the adaptive filter does not only predict and cancel the feedback component in the microphone signal, but also part of the near-end signal, which results in a distorted feedback-compensated signal $d(t)$. The PEM-based AFC is shown in Figure 8.2. An unbiased identification of the feedback path model can be achieved by applying decorrelation in the adaptive filtering circuit, i.e., by first prefiltering the loudspeaker and the microphone signals with the inverse near-end signal model $\hat{H}^{-1}(q, t)$ (see Figure 8.2) before feeding these signals to the adaptive filtering algorithm. The near-end signal model and the feedback path model can be jointly estimated using the PEM [229][209][228]. The PEM delivers an unbiased estimate of the feedback path coefficient vector

$\mathbf{f}(t) = [f_0(t) \ f_1(t) \ \dots \ f_{n_F}(t)]$, where n_F is the feedback path model order, by minimization of the prediction error criterion

$$\min_{\hat{\mathbf{f}}(t)} \sum_{k=1}^t \varepsilon^2(k) \quad (8.4)$$

if the prediction error is calculated as

$$\varepsilon(t) = H^{-1}(q, t) [y(t) - \hat{F}(q, t)u(t)] \quad (8.5)$$

where $H(q, t)$ is a linear model for the source signal $v(t)$. The model structure of $H(q, t)$ and the estimation of its parameters will be discussed in details in the remainder of this chapter.

8.2 Single near-end signal model

In PEM-based AFC the near-end signal is typically modelled with an LP model [209][210], i.e.,

$$\begin{aligned} v(t) &= H_{\text{LP}}(q, t)r(t) \\ &= \frac{1}{1 + \sum_{i=1}^{n_C} c_i(t)q^{-i}} r(t), \end{aligned} \quad (8.6)$$

where $r(t)$ is a white noise signal and n_C is the model order. The prediction error for the PEM-based AFC using an LP model is then

$$\boxed{\varepsilon(t) = H_{\text{LP}}^{-1}(q, t) [y(t) - F(q, t)u(t)]}. \quad (8.7)$$

The drawback with an LP model for the near-end signal is that the white noise assumption is not valid for periodic signals such as voiced speech where the excitation $r(t)$ is an impulse train [209].

8.3 Cascaded near-end signal model

The idea behind using a cascaded near-end signal model is that the tonal components can be represented by one model and the noise components by another model. In [228] it has been proposed that a PZLP model of order $2P$ [24] can be used to represent P tonal components. Still, by constraining the poles and the zeros to lie on a common radial line in the z -plane, the number of unknown parameters in the pole-zero model can be limited to P and the LP parameters

can be uniquely related to the unknown frequencies [227]. The constrained PZLP (CPZLP) model can be written as

$$v(t) = \left(\prod_{n=1}^P \frac{1 - 2\rho \cos \omega_n z^{-1} + \rho^2 z^{-2}}{1 - 2 \cos \omega_n z^{-1} + z^{-2}} \right) e(t) \quad (8.8)$$

where ω_n denotes the frequencies and ρ the pole radius. The CPZLP minimization criterion is given by

$$\min_{\omega} V(\omega) = \min_{\omega} \frac{1}{M} \sum_{t=1}^M e^2(t, \omega) \quad (8.9)$$

with the residual signal defined as the output from the prediction error filter (PEF)

$$e(t, \omega) = \left(\prod_{n=1}^P \frac{1 - 2 \cos \omega_n z^{-1} + z^{-2}}{1 - 2\rho \cos \omega_n z^{-1} + \rho^2 z^{-2}} \right) v(t) = H_{\text{CPZLP}}^{-1}(q, t) v(t) \quad (8.10)$$

and $\omega = [\omega_1 \dots \omega_P]^T$. The CPZLP minimization in (8.9)-(8.10) can be solved in a decoupled fashion, using an iterative line search optimization [227]. Using the CPZLP model for the tonal components and an LP model for the noise components the prediction error using a cascaded near-end signal model can then be written as

$$\boxed{\varepsilon(t) = H_{\text{LP}}^{-1}(q, t) H_{\text{CPZLP}}^{-1}(q, t) [y(t) - F(q, t)u(t)]} \quad (8.11)$$

with the PEF for the noise component $r(t)$ defined as $H_{\text{LP}}^{-1}(q, t) = C(q, t) = 1 + \sum_{i=1}^{n_C} c_i(t)q^{-i}$ which is straightforward using LP on the output signal of the first PEF $H_{\text{CPZLP}}^{-1}(q, t)$.

8.4 Experimental results

In this section, experimental results for the PEM-based AFC with the decorrelation prefilters $H_{\text{LP}}^{-1}(q, t)$ and $H_{\text{CPZLP}}^{-1}(q, t)$ are presented both when using a single-near end signal model, referred to as AFC-LP, and a cascaded near-end signal model, referred to as AFC-CPZLP.

8.4.1 Experimental set-up

The near-end noise model order is fixed to $n_C = 30$ and the near-end speech model is varied from $P=15, 10,$ and 5 . Both near-end signal models are estimated using 50% overlapping data windows of length $M = 320$ samples. The NLMS adaptive filter length is set equal to the acoustic feedback path length, i.e., n_F

= 200 (measured hearing aid feedback path). The near-end signal is a 30 second male speech signal sampled at $f_s = 16$ kHz. The speech signal is taken from an interview with two male Dutch-speaking subjects that was digitally broadcasted by the Flemish Radio and Television Network (VRT). The forward path gain $K(t)$ defined in (8.2) is set 3 dB below the maximum stable gain (MSG) without feedback cancellation.

8.4.2 Performance measures

To assess the performance of the AFC algorithm the following measures are used. The achievable amplification before instability occurs is measured by the MSG, which is defined as

$$\text{MSG}(t) = -20 \log_{10} \left[\max_{\omega \in \mathcal{P}} |J(\omega, t)[F(\omega, t) - \hat{F}(\omega, t)]| \right] \quad (8.12)$$

where $J(q, t) = \frac{G(q, t)}{K(t)}$ denotes the forward path transfer function without the amplification gain $K(t)$, and \mathcal{P} denotes the set of frequencies at which the feedback signal $x(t)$ is in phase with the near-end signal $v(t)$. The misadjustment between the estimated feedback path $\hat{\mathbf{f}}(t)$ and the true feedback path \mathbf{f} represents the accuracy of the feedback path estimation and is defined as

$$\text{MA}_F = 20 \log_{10} \frac{\|\hat{\mathbf{f}}(t) - \mathbf{f}\|_2}{\|\mathbf{f}\|_2}. \quad (8.13)$$

8.4.3 Results

The first simulation, is performed with a single near-end signal model using an LP model since this is the motivation for introducing the cascaded near-end signal model. The instantaneous value of the $\text{MSG}(t)$ and the corresponding misadjustment is shown in Figure 8.3. The $\text{MSG}(t)$ curves have been smoothed with a one-pole low-pass filter to improve the clarity of the figures. The instantaneous value of the forward path gain $20 \log_{10} K(t)$ and the MSG without acoustic feedback control ($\text{MSG} F(q)$) are also shown. It is clear that the LP model is initially able to provide sufficient decorrelation resulting in increased MSG and misadjustment. However over time it can be observed that the LP model fails which can be due to the tonal components in the speech. Once instability occurs it can be a problem to recover due to the closed loop.

The second simulation is performed with a cascaded near-end signal model using a CPZLP model combined with a LP model for different order P which can be seen in Figure 8.4. In general the MSG is higher for AFC-CPZLP compared to AFC-LP

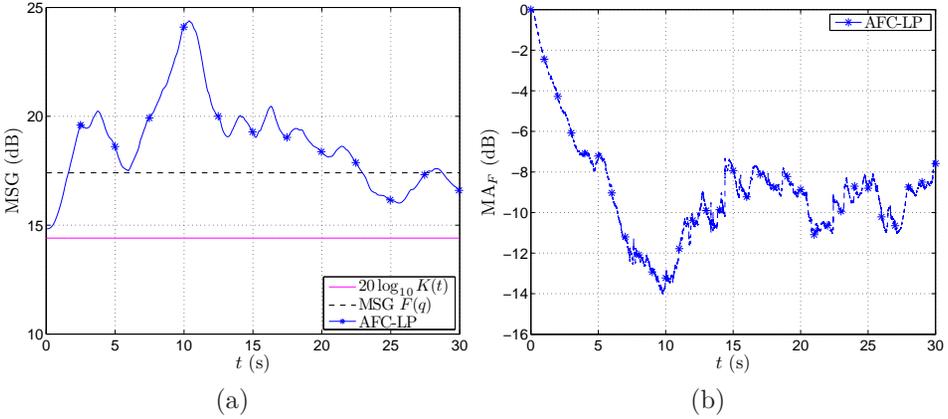


Figure 8.3: Instantaneous MSG vs. time for simulations with speech for PEM-based AFC in hearing aids and misadjustment between the estimated feedback path $\hat{\mathbf{f}}(t)$ and the true feedback path \mathbf{f} using AFC-LP.

and the corresponding misadjustment is also lower. An important observation here is that the AFC-CPZLP provides a strong decorrelation such that a stable MSG and misadjustment is achieved. A lower order also seems to perform better both in terms MSG and misadjustment. However when $P=5$ the AFC performance drops below the AFC performance of $P=10$. This of course leads to the question as to which order P is the optimal for speech in terms of the AFC performance.

8.5 Conclusion

In this chapter we have introduced the PEM-based AFC using a near-end signal model based on LP. Another solution is based on using a cascaded near-end signal model based on a LP and a PZLP model such that the tonal and the noise components are modelled separately.

Experimental results show that the PEM-based AFC with LP is not able to provide sufficient decorrelation and therefore a low MSG is achieved with a rather high filter misadjustment. The cascaded near-end signal model is able to outperform the single near-end signal model both in terms of MSG and filter misadjustment.

It has been shown that the PEM-based AFC performance highly depends on the near-end signal model. Especially the modelling of the tonal components seems to play a crucial role since the tonal components are more correlated compared to

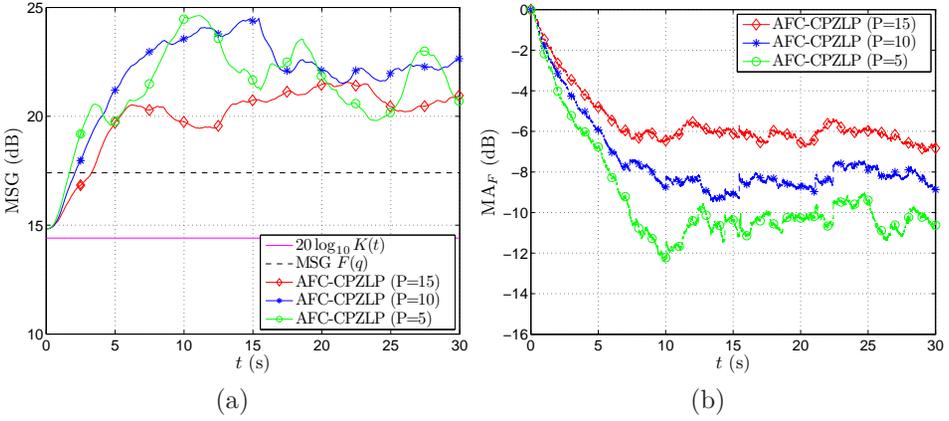


Figure 8.4: Instantaneous MSG vs. time for simulations with speech for PEM-based AFC in hearing aids and misadjustment between the estimated feedback path $\hat{\mathbf{f}}(t)$ and the true feedback path \mathbf{f} using AFC-CPZLP.

the noise components. In Chapter 9 we will therefore improve the modelling of the tonal components by exploiting typical information regarding the speech, e.g., pitch, amplitude, order, voiced and unvoiced segments. Furthermore the use of a sinusoidal near-end signal model is proposed.

Chapter 9

PEM-based AFC using a harmonic sinusoidal near-end signal model

This chapter addresses the issue of designing improved prediction error filters for PEM-based AFC. To this aim a harmonic sinusoidal near-end signal model is introduced together with various pitch estimation techniques. Basically the idea is to find an improved method to represent the near-end signal by using the knowledge that we have regarding speech signals. For this purpose we have turned our attention to the research area of speech and audio coding based on harmonic sinusoidal based pitch estimation techniques. The reproduction of speech signals highly depends on an accurate estimation of parameters such as pitch, amplitude, model order, and the selection of voiced-unvoiced frames. It is therefore interesting to analyze if a more accurate estimation of the near-end signal model would result in improved PEM-based AFC performance.

Section 9.1 introduces the harmonic sinusoidal near-end signal model together with three methods to estimate the pitch. Furthermore it is also shown the model order and the amplitude of the harmonics can be estimated.

Section 9.2 presents the PEF that incorporates the pitch, harmonics amplitudes, and the model order in the PZLP model.

Section 9.3 explains how voiced-unvoiced frames affect the AFC correlation problem which is measured using the spectral flatness measure (SFM). The idea here is to use a voiced-unvoiced detection in the PEF to switch between a single near-end signal model and a cascaded near-end signal model. This can be beneficial

for two reasons. First of all unvoiced frames do not produce a high signal correlation in the AFC compared to voiced frames. Furthermore the complexity can be reduced by exploiting the unvoiced frames and hence using a single near-end signal model for unvoiced frames and a cascaded near-end signal model for voiced frames.

Section 9.4 presents the experimental results for the PEM-based AFC using the improved near-end signal model based on harmonic sinusoidal modelling.

9.1 Harmonic sinusoidal near-end signal model

In a harmonic sinusoidal near-end signal model [163][164] the tonal components are represented as a sum of real harmonically related sinusoids. This means that the sinusoids are assumed to have frequencies that are integer multiples of a fundamental frequency ω_0 , i.e., $\omega_n = n\omega_0$. This follows naturally from voiced speech being quasi-periodic. The near-end signal $v(t)$ consisting of a sum of real harmonically related sinusoids and additive noise can be written as,

$$v(t) = \sum_{n=1}^P a_n \cos(n\omega_0 t + \phi_n) + r(t), \quad t = 1, \dots, M \quad (9.1)$$

where P is the model order, $\omega_0 \in [0, \pi]$ is the pitch frequency, a_n the amplitude, and $\phi_n \in [0, 2\pi)$ the phase of the n th sinusoid, and $r(t)$ the noise which is assumed to be autoregressive, i.e., $r(t) = \frac{1}{C(q,t)}e(t)$, with $C(q,t) = 1 + \sum_{i=1}^{n_C} c_i(t)q^{-i}$.

The pitch estimation technique used here is primarily based on optimal filtering (optfilt) [27] of the feedback-compensated signal $d(t)$, which ideally corresponds to the near-end signal $v(t)$. However we will also include some other pitch estimation methods in order to show how the different pitch estimation techniques affect the AFC performance. The different pitch estimation techniques together with the estimation of the harmonic amplitudes and the model order are explained in the following sections.

9.1.1 Optimal-filtering based pitch estimation

The idea behind pitch estimation based on filtering is to find a filter that passes power undistorted at the harmonic frequencies $n\omega_0$, while minimizing the power at all other frequencies. This filter design problem can be stated mathematically as [27]

$$\min_{\mathbf{h}} \mathbf{h}^H \mathbf{R} \mathbf{h} \quad \text{s.t.} \quad \mathbf{h}^H \mathbf{Z} = \mathbf{1}, \quad (9.2)$$

with $\mathbf{1} = [1 \ \dots \ 1]^T \in \mathbb{R}^{2P}$, $\mathbf{Z} = [\mathbf{z}(\omega_0) \ \mathbf{z}^*(\omega_0) \ \dots \ \mathbf{z}(\omega_0 P) \ \mathbf{z}^*(\omega_0 P)]$, \mathbf{h} is the length- N impulse response of the filter, and \mathbf{R} is the covariance matrix defined as

$$\mathbf{R} = \mathbb{E}\{\mathbf{d}(t)\mathbf{d}^H(t)\}, \quad (9.3)$$

where $(\cdot)^H$ denotes Hermitian transpose and $\mathbf{d}(t)$ is a vector containing M consecutive samples of the feedback-compensated signal $d(t)$ [27]. Using the Lagrange multiplier method, the optimal filter can be shown to be

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}\right)^{-1}\mathbf{1} \quad (9.4)$$

This filter is signal adaptive and depends on the unknown fundamental frequency. Intuitively, one can obtain a fundamental frequency estimate by filtering the signal using the optimal filters for various fundamental frequencies and then picking the one for which the output power is maximized, i.e.,

$$\begin{aligned} \hat{\omega}_0 &= \arg \max_{\omega_0} \mathbb{E}\{|\mathbf{h}^H\mathbf{d}(t)|^2\} \\ &= \arg \max_{\omega_0} \mathbf{1}^H \left(\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}\right)^{-1} \mathbf{1}. \end{aligned} \quad (9.5)$$

This method has demonstrated to have a number of desirable features, namely excellent statistical performance and robustness against periodic interference [27].

9.1.2 Subspace-orthogonality based pitch estimation

The idea behind subspace methods is to divide the full space into a so-called signal subspace containing the signal of interest and its orthogonal complement, the noise subspace. The subspace orthogonality method is based on the observation that the sinusoids in (9.1) are all orthogonal to the noise. The covariance matrix of the observed signal in (9.1) can be shown to be

$$\mathbf{R} = \mathbb{E}\{\tilde{\mathbf{d}}(t)\tilde{\mathbf{d}}^H(t)\} \quad (9.6)$$

$$= \mathbf{Z}\mathbf{P}\mathbf{Z}^H + \sigma^2\mathbf{I} \quad (9.7)$$

where $\tilde{\mathbf{d}}(t)$ is a vector containing M consecutive samples of the analytical counterpart of the feedback-compensated signal $d(t)$ [27] since the subspace methods assumes a complex signal model. Furthermore, \mathbf{Z} is a Vandermonde matrix containing the sinusoids of the model in (9.1), and \mathbf{P} is the covariance matrix of the amplitudes, which can be shown to be diagonal under certain conditions. Finally, σ^2 denotes the variance of the additive noise, and \mathbf{I} is the identity matrix. In the presence of colored noise, it is required that pre-whitening is applied, as the model in (9.7) would otherwise be invalid. Exploiting the fact

that the noise subspace eigenvectors \mathbf{G} are orthogonal to the columns of the matrix \mathbf{Z} , it follows that the fundamental frequency ω_0 can be estimated as

$$\hat{\omega}_0 = \arg \min_{\omega_0} \|\mathbf{Z}^H \mathbf{G}\|_F^2, \quad (9.8)$$

where \mathbf{Z} depends on ω_0 . More specifically, the matrix \mathbf{G} is constructed from the $M - 2P$ least significant eigenvectors of \mathbf{R} .

9.1.3 Subspace-shift-invariance based pitch estimation

The next method is based on a particular property of the signal subspace generated by signals as in (9.1), namely the shift-invariance property. The signal subspace is spanned by the columns of the matrix \mathbf{S} formed from the $2P$ most significant eigenvectors of \mathbf{R} . Two matrices $\underline{\mathbf{S}}$ and $\overline{\mathbf{S}}$ are constructed by removing the last and first row of the matrix \mathbf{S} which can be shown to be related by a linear transform as $\overline{\mathbf{S}} = \underline{\mathbf{S}}\mathbf{\Xi}$. The problem of finding the fundamental frequency can then be seen as a fitting problem, i.e.

$$\overline{\mathbf{S}} \approx \underline{\mathbf{S}}\mathbf{Q}\tilde{\mathbf{D}}\mathbf{Q}^{-1} \quad (9.9)$$

where $\tilde{\mathbf{D}} = \text{diag}([e^{j\omega_0} \ e^{-j\omega_0} \ \dots \ e^{j\omega_0 P} \ e^{-j\omega_0 P}])$ is a diagonal matrix containing the unknown fundamental frequency. The matrix \mathbf{Q} contains the eigenvectors of the matrix $\hat{\mathbf{\Xi}} = (\underline{\mathbf{S}}^H \underline{\mathbf{S}})^{-1} \underline{\mathbf{S}}^H \overline{\mathbf{S}}$. The fundamental frequency can then be estimated as

$$\hat{\omega}_0 = \arg \min_{\omega_0} \|\overline{\mathbf{S}} - \underline{\mathbf{S}}\mathbf{Q}\tilde{\mathbf{D}}\mathbf{Q}^{-1}\|_F^2, \quad (9.10)$$

which can be simplified significantly, as shown in [27].

9.1.4 Amplitude and models order estimation

Once ω_0 is known, the amplitude of the sinusoids can be estimated using a least squares approach:

$$\hat{\mathbf{a}} = \left(\mathbf{Z}^H \mathbf{Z}\right)^{-1} \mathbf{Z}^H \mathbf{d} \quad (9.11)$$

with $\hat{\mathbf{a}} = [\hat{a}_1 \ \dots \ \hat{a}_P]$. Finally, the number of harmonics P can be determined by using a maximum a posteriori (MAP) criterion [27][47],

$$\hat{P} = \arg \min_P M \log \hat{\sigma}_P^2 + P \log M + \frac{3}{2} \log M \quad (9.12)$$

where the first term is a log-likelihood term which comprises a noise variance estimate that depends on the candidate model order, the second term is the penalty associated with the amplitude and phase, while the third term is due to the fundamental frequency.

9.2 PZLP using pitch estimation based PEF

Inserting the estimated pitch of the model (9.1) in (8.10) the PEF $H_{\text{Pitch}}^{-1}(q, t)$ for the harmonic sinusoidal near-end signal can be written as a cascade of second-order sections:

$$H_{\text{Pitch}}^{-1}(q, t) = \prod_{n=1}^P \frac{1 - 2\nu_n \cos n\omega_0 z^{-1} + \nu_n^2 z^{-2}}{1 - 2\rho_n \cos n\omega_0 z^{-1} + \rho_n^2 z^{-2}} \quad (9.13)$$

where the poles and zeros are on the same radial lines, with the poles positioned between the zeros and the unit circle, i.e., $0 \ll \rho_n < \nu_n \leq 1$. The advantage of $H_{\text{Pitch}}^{-1}(q, t)$ is that the frequencies are assumed to be an integer multiple of a fundamental frequency which follows naturally from voiced speech being quasi-periodic. This assumption is not made in the general CPZLP model (9.1) where all the frequencies are estimated independently. Furthermore, the pitch estimation also offers a method to incorporate the amplitude a_n and the order P , see below.

In [163] the performance of the PEF in (9.13) was analyzed in terms of different pitch estimation techniques, incorporating optimal filtering and subspace methods. The use of the estimated pitch, i.e., $\omega_0 = \hat{\omega}_0$ in general resulted in better performance in the PEM-based AFC. The PEF in (9.13) is generally designed with the zero radii fixed to $\nu_n = 1$ and the pole radii fixed to $\rho_n = 0.95$, and with a fixed order P . This would result in a PEF that applies equal (infinity) suppression for all frequencies $\omega_n = n\omega_0$ by placing all the zeros on the unit circle. However, speech is a non-stationary signal with time-varying pitch, harmonic amplitudes, and number of harmonics. Here, the pitch PEF is further improved by including the amplitudes a_n and the estimated order $P = \hat{P}$ which again outperformed the CPZLP model.

9.2.1 Incorporating amplitude, order and pitch information

For an example speech frame, with a spectrum shown in Figure 9.1, the corresponding PEF response is shown in Figure 9.2 (when CPZLP (8.8) is used for frequency estimation) and Figure 9.3(top, when pitch estimation (9.13) is used). A first observation is that the PEF applies equal (infinity) suppression for all frequencies when all the zeros are placed on the unit circle. The PEF using pitch estimation in Figure 9.3(top) shows that the PEF has a more dense structure in the low frequency region when harmonicity is assumed.

Pitch and variable model order estimates are straightforward to include in the PEF, by setting $\omega_0 = \hat{\omega}_0$ and $P = \hat{P}$. From the design of the PEF it is clear that the zero radius determines the notch depth, which should correspond to the inverse of estimated amplitudes. Incorporating the amplitude in the PEF then

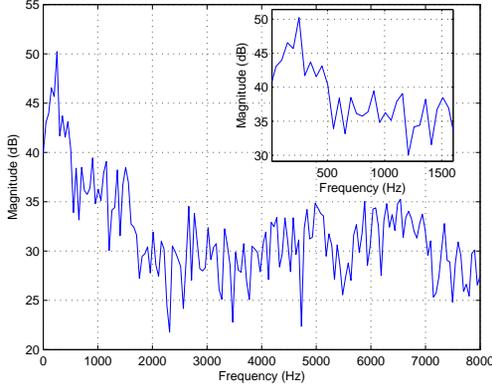


Figure 9.1: Speech spectrum used to estimate the PEF.

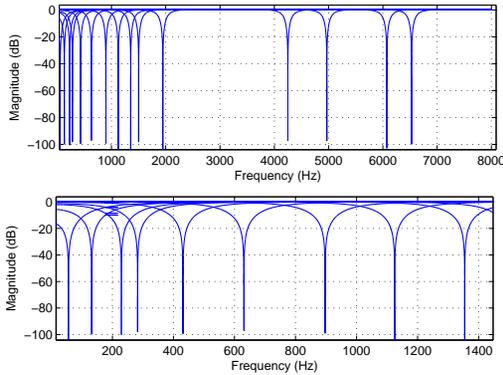


Figure 9.2: PEF using CPZLP for frequency estimation, (top) notch filters up to 8000Hz and (bottom) notch filters up to 1400Hz.

follows from the design rule in [226], i.e.,

$$\nu_n = \max \left(\rho_n, 1 - \frac{1 - \rho_n}{\hat{a}_n} \right). \tag{9.14}$$

For the harmonic amplitudes estimated in (9.11) and the pitch estimated in (9.5) the PEF is shown in Figure 9.3(bottom) which shows a more signal dependent behavior, when compared to the corresponding speech spectrum shown in Figure 9.1.

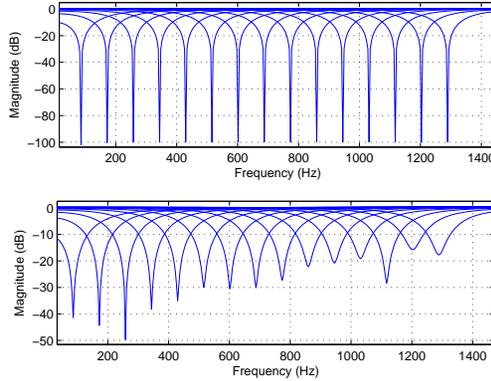


Figure 9.3: PEF using optimal filtering based pitch estimation (top) without incorporating amplitude information and (bottom) with incorporating amplitude information.

In previous work [163], besides assuming infinite notch depth, the model order is also assumed to be equal for every speech frame. This may not be the optimal solution since speech generally can be considered voiced or unvoiced, resulting in different harmonic amplitudes and number of harmonics. A histogram of the estimated number of harmonics using (9.12) for the speech signal used in the evaluation in Section 9.4 is shown for different frames in Figure 9.4. This indeed suggests, that the harmonic sinusoidal near-end signal model order varies across different frames and that the fixed model order of $P = 15$ used in [163] indeed is too high compared to the estimated model order \hat{P} .

The prediction error using a cascaded harmonic sinusoidal near-end signal model can then be written as

$$\varepsilon(t) = H_{\text{LP}}^{-1}(q, t) H_{\text{Pitch}}^{-1}(q, t) [y(t) - F(q, t)u(t)]. \quad (9.15)$$

9.3 Voiced-unvoiced detection

Previous design of the PEF has been focused on time-varying parameters such as frequency, amplitude and the number of harmonics of a typical speech signal. However, the motivation behind the cascaded near-end signal model is due to the fact that a single near-end signal model, e.g., an LP model can not model the tonal components such as voiced speech. As an additional ingredient we will analyze the performance of the PEF by introducing a voiced-unvoiced detector.

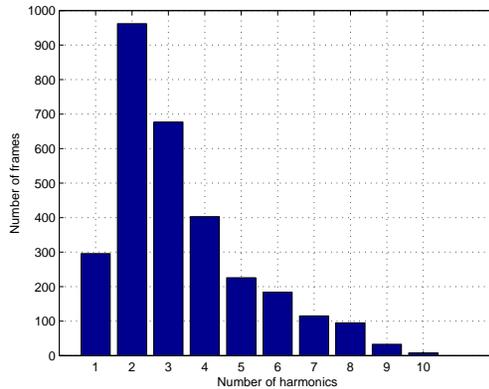


Figure 9.4: Number of harmonics \hat{P} for each frame.

9.3.1 ZCR and energy based voiced-unvoiced detection

It is well-known that a speech signal can be divided into voiced and unvoiced frames. The voiced-unvoiced parts of a speech signal can be extracted using information such as zero crossing rate (ZCR) and the energy in a given frame [6][4][178][192][133]. The ZCR indicates the number of times the amplitude of a speech signal for a given time frame (samples) passes through a value of zero, i.e.,

$$\text{ZCR} = \sum_{t=1}^M \frac{1 - \text{sgn}[d(t)]\text{sgn}[d(t+1)]}{2} \quad (9.16)$$

where sgn is the *signum* function. The ZCR is low for voiced speech and high for unvoiced speech. The short-time energy of a speech frame is given by

$$\text{Energy} = \frac{1}{M} \sum_{t=1}^M d^2(t). \quad (9.17)$$

The energy is high for voiced speech and low for unvoiced speech. Voiced speech has a high energy due to the periodicity while unvoiced speech is non-periodic. Examples of the ZCR for selected voiced and unvoiced frames are shown in Figure 9.5. This shows that the ZCR is much lower for voiced speech compared to unvoiced speech. Figure 9.5 also shows a typical characteristic for a given voiced and unvoiced frame. Tracking the non-stationarity of a speech signal should then also reflect a distinction of the voiced and unvoiced frames and should therefore also be included in the PEF.

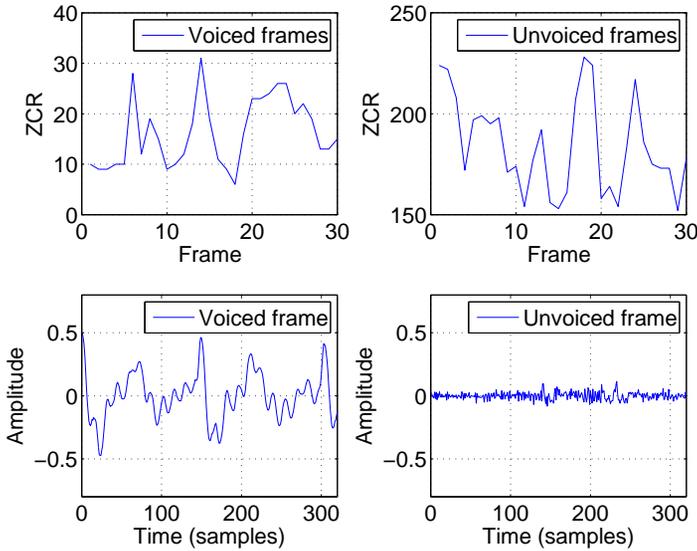


Figure 9.5: Examples of ZCR for selected voiced and unvoiced frames and illustration of a typical voiced and unvoiced frame.

9.3.2 Spectral flatness of the residual

The important issue here is how the voiced-unvoiced frames affects the PEF and hence the PEM-based AFC performance. It is not clear if the cascaded near-end signal model always has an advantage, e.g., when a given frame is unvoiced. The purpose of the PEF is to make the spectrum of the filtered near-end signal as white as possible by modelling the near-end signal. As an attempt to analyze the performance of the PEF the spectral flatness measure (SFM) of the filtered near-end signal (residual) is used [138], i.e.,

$$\text{SFM} = \frac{\exp \left[\frac{1}{L} \sum_{k=0}^{L-1} \ln |E(e^{j\frac{2\pi k}{L}}, \xi)| \right]}{\frac{1}{L} \sum_{k=0}^{L-1} |E(e^{j\frac{2\pi k}{L}}, \xi)|} \quad (9.18)$$

with $E(e^{j\frac{2\pi k}{L}}, \xi)$, $k = 0, \dots, L-1$ the L -point DFT of the residual $e(t, \xi)$. The SFM is expressed on a dB-scale where 0dB corresponds to a flat spectrum.

Figure 9.6 shows the SFM for selected voiced frames and the residual from the different PEFs. A first observation is that the SFM for the voiced frames is rather low, due to the periodicity, which indicates a potential high correlation. The SFM for the LP shows that the LP can increase the SFM however not as good as with the CPZLP (8.10) or with the pitch estimation (9.13). This suggests that for the voiced frames a cascaded near-end signal model is desired. It is not clear how

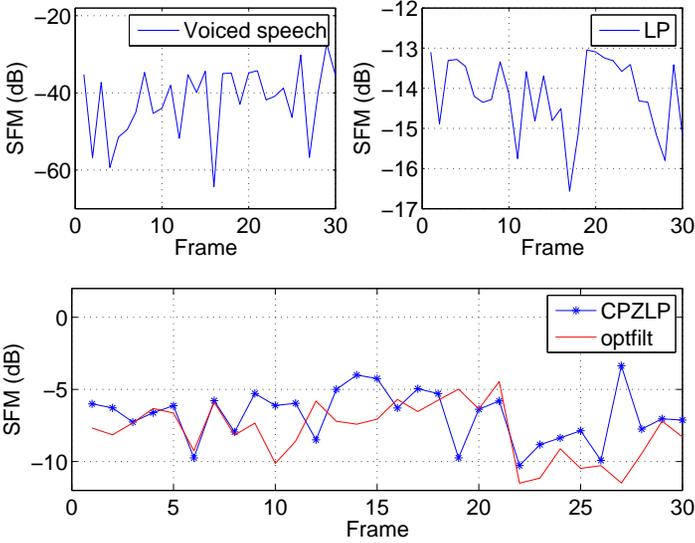


Figure 9.6: SFM for selected voiced frames and for the different PEF.

the SFM directly relates to the PEM-based AFC performance but it does show that the SFM differs significantly from frame to frame and that the PEF therefore should be signal adaptive. On the other hand, the SFM for unvoiced frames is much higher compared to voiced frames and hence a single LP model should be sufficient, which is shown in Figure 9.7. It is also clear that a cascaded near-end signal model does not offer anything additional in terms of SFM for unvoiced frames. Applying a cascaded near-end signal model when an unvoiced frame is detected is undesirable since the spectrum is already flat. Combining the ZCR and the short-time energy for the speech signal used in the simulation resulted in the following classification of voiced-unvoiced frames, out of 3000 frames 751 are classified as unvoiced. This also means that the complexity can also be further reduced compared to always using a cascaded near-end signal model which is a desirable feature in hearing aids.

9.4 Experimental results

In this section, experimental results for the PEM-based AFC with the improved near-end signal model is presented and compared to the results presented in Chapter 8.

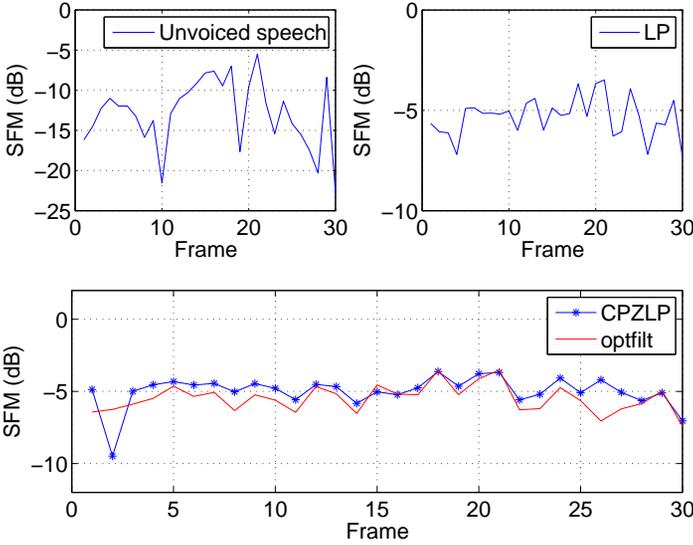


Figure 9.7: SFM for selected unvoiced frames and for the different PEF.

9.4.1 Experimental set-up

The LP model order is fixed to $n_C = 30$. Both near-end signal models are estimated using 50% overlapping data windows of length $M = 320$ samples. The NLMS adaptive filter length is set equal to the acoustic feedback path length, i.e., $n_F = 200$ (measured hearing aid feedback path). The optimal filtering length in (9.2) is set to $N = \frac{M}{4}$. The near-end signal is a 30 second male speech signal sampled at $f_s = 16$ kHz. The speech signal is taken from an interview with two male Dutch-speaking subjects that was digitally broadcasted by the Flemish Radio and Television Network (VRT). The forward path gain $K(t)$ is set 3 dB below the maximum stable gain (MSG) without feedback cancellation.

9.4.2 Results

Different pitch estimation techniques

The instantaneous value of the MSG and the misadjustment for the different pitch estimation techniques with $P = 10$ for different stepsize μ is shown in Figure 9.8. The MSG curves have been smoothed with a one-pole low-pass filter to improve the clarity of the figures. The instantaneous value of the forward path

gain $20 \log_{10} K(t)$ and the MSG without acoustic feedback control (MSG $F(q)$) are also shown.

The AFC-LP is included as a reference since this performance is a baseline for introducing a cascaded near-end signal model [209]. At some point the MSG in the AFC-LP decreases and even gets close to instability. Compared to the AFC-CPZLP, the MSG in this case seems to be more stable with an overall higher MSG compared to the AFC-LP even though the misadjustment is lower for the AFC-LP. The benefit of AFC-CPZLP can be explained by the benefit of using a cascaded near-end signal model. A cascade of near-end signal models removes the coloring and periodicity (due to glottal excitation) in voiced speech segments. On the other hand, a single short-term predictor fails to remove the periodicity, which causes the loudspeaker signal still being correlated with the near-end signal during voiced speech.

The MSG is in general higher using AFC-shiftinv, AFC-orth and AFC-optfilt compared to the existing methods AFC-LP and AFC-CPZLP, which supports the conjecture that an accurate estimation of the near-end signal model results in a better decorrelation and hence an increase in MSG. Using lower stepsize shows a significantly better convergence behavior for AFC-shiftinv, AFC-orth and AFC-optfilt compared to AFC-CPZLP. From these results, it is clear that the frequency estimation methods have a great impact on the AFC performance. On the other hand, it is worth noting that the choice of the stepsize seems to have a great impact on the convergence for AFC-shiftinv, AFC-orth and AFC-optfilt, whereas AFC-CPZLP seems to stabilize faster but at a larger error.

Amplitude and model order

The instantaneous value of the MSG and the corresponding misadjustment for the case where the amplitude and the order estimation is incorporated in the PEF is shown in Figure 9.9(e)-(f) and is compared to using a fixed order as shown in Figure 9.9(a)-(d). For the scenarios with fixed orders the MSG is in general higher for AFC-optfilt compared to AFC-CPZLP and the corresponding misadjustment is also lower for AFC-optfilt. For the AFC-CPZLP a fixed order of 20 seems to be the best choice whereas for AFC-optfilt a fixed order of 10 gives the best result. The fact that AFC-optfilt can achieve a better performance than AFC-CPZLP with a lower order can be explained by using Figures 9.2 and 9.3. The structure of the PEF is more dense towards lower frequencies when the pitch estimation method is used and it is therefore anticipated that the PEF using CPZLP does not sufficiently suppress the tonal components when a lower order is used. Furthermore it is also clear that a fixed order of 30 is too high, which is supported by Figure 9.4, especially when the PEF applies infinite suppression.

Using a variable order compared to a variable amplitude (with a fixed model

order 30) for AFC-optfilt almost results in the same AFC performance, with a small advantage for the variable order performance. The performance when both variable order and variable amplitude are included in the PEF is not shown since the performance is similar to the case when only a variable amplitude is used. This probably happens because at very low amplitude the PEF results in a pole-zero cancellation and no suppression is applied. However the variable amplitude and variable order scenarios slightly outperform the AFC-optfilt with fixed order while the advantage compared to AFC-CPZLP is significant.

Voiced-unvoiced detection

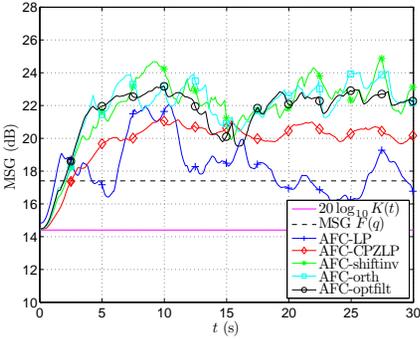
The instantaneous value of the MSG and the corresponding misadjustment for the case where a voiced-unvoiced detector is incorporated in the PEF is shown in Figure 9.10(c)-(d) and is compared to using a fixed order as shown in Figure 9.10(a)-(b). In general the MSG is higher for AFC-optfilt compared to AFC-CPZLP and the corresponding misadjustment is also lower for AFC-optfilt. For both AFC-optfilt and AFC-CPZLP a lower order seems to perform better both in terms of MSG and misadjustment. AFC-optfilt with $P=10$ performs poorly around 15 seconds into the simulation but this problem seems to vanish when the voiced-unvoiced detector is used. The voiced-unvoiced detector improves the AFC-optfilt with up to 2-4dB in MSG. For AFC-CPZLP the performance does not change much using the voiced-unvoiced detector actually the MSG decreases up to 1-2dB in MSG when the voiced-unvoiced detector is used. Again a small advantage is observed for the AFC-optfilt when using the voiced-unvoiced detector but compared to AFC-CPZLP the performance difference is significant.

9.5 Conclusion

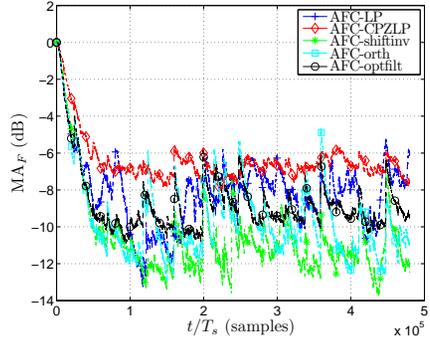
In this chapter we have introduced a PEM-based AFC with an improved PEF that exploits a harmonic sinusoidal near-end signal model. We have shown that various pitch estimation techniques can improve the PEM-based AFC performance. In addition including information such as amplitude, number of harmonics, and a voiced-unvoiced detector can improve the design of the PEF. Furthermore it has been shown why a single near-end signal model based LP fails to provide sufficient decorrelation due to the SFM for voiced frames. This is the motivation behind using a cascaded near-end signal model here.

Experimental results show that in general the PEM-based AFC performance is improved when a cascaded near-end signal model is used. However the inclusion of speech features such as pitch, amplitude, number of harmonics, and a voiced-unvoiced detector further improved the accuracy of the near-end signal model

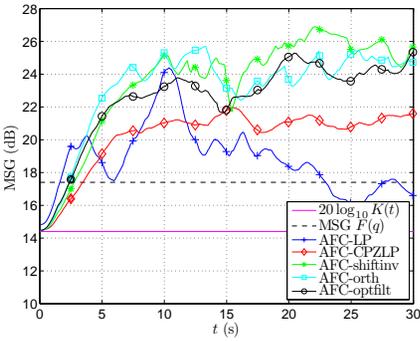
leading to an increased PEM-based AFC performance. The overall message is that exploiting more relevant speech features in the near-end signal model can indeed lead to an increased AFC performance.



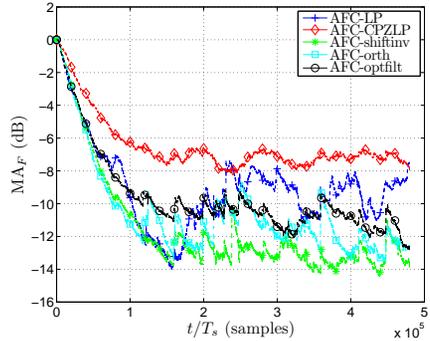
(a) Stepsize $\mu = 0.01$



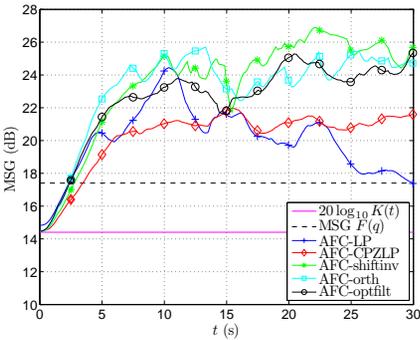
(b) Stepsize $\mu = 0.01$



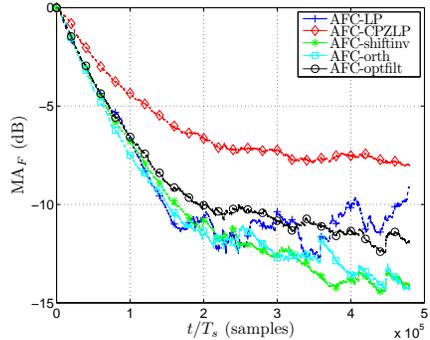
(c) Stepsize $\mu = 0.005$



(d) Stepsize $\mu = 0.005$

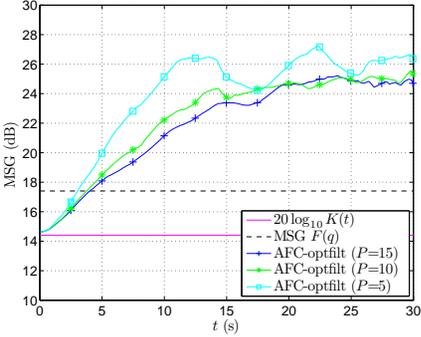


(e) Stepsize $\mu = 0.0025$

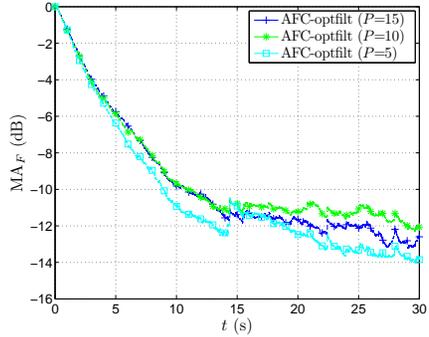


(f) Stepsize $\mu = 0.0025$

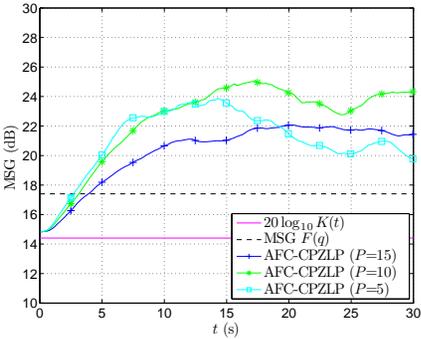
Figure 9.8: MSG and misadjustment for PEM-based AFC using different pitch estimation techniques.



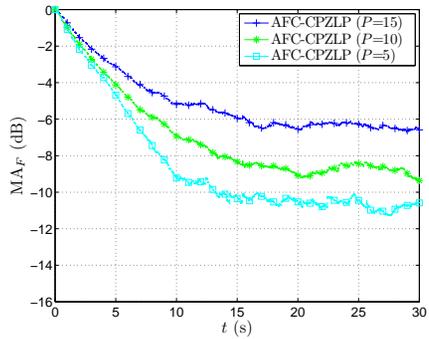
(a) Optfilt using different order



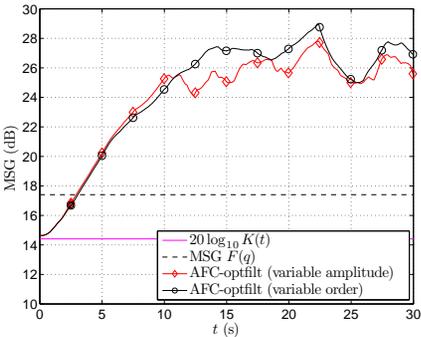
(b) Optfilt using different order



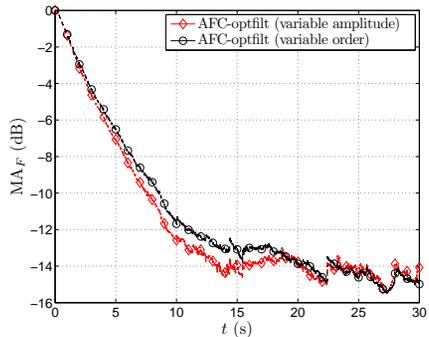
(c) CPZLP using different order



(d) CPZLP using different order

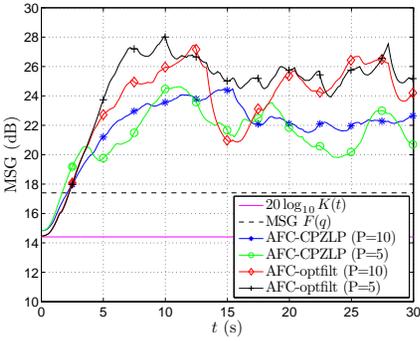


(e) Variable amplitude and order

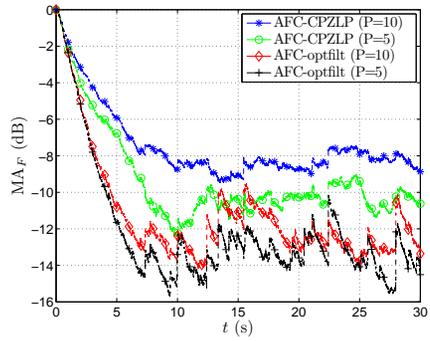


(f) Variable amplitude and order

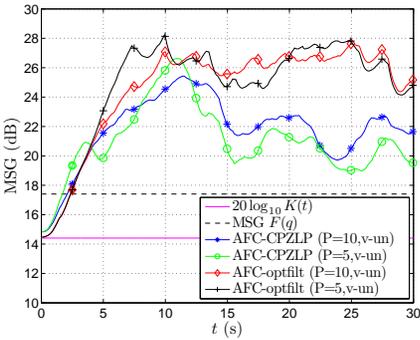
Figure 9.9: MSG and misadjustment for PEM-based AFC when a variable amplitude and order is used compared to a fixed order.



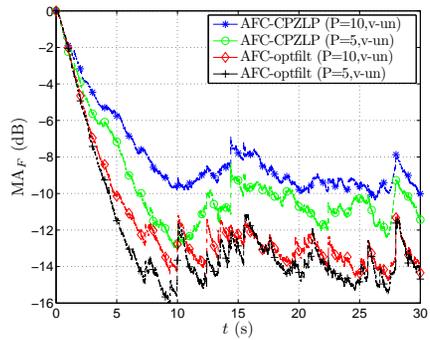
(a)



(b)



(c)



(d)

Figure 9.10: MSG and misadjustment for PEM-based AFC when a voiced-unvoiced detector is used.

Chapter 10

Conclusion and further research

This chapter summarizes the main conclusions presented in this dissertation together with some future research perspectives.

10.1 Conclusion

Hearing impairment can be caused by many factors, e.g., daily exposure to excessive noise in the work environment and listening to loud music which are scenarios that we all can be exposed to in our daily life. Hearing loss can also be age-related which makes the research on hearing aids very important. For hearing aid users background noise and acoustic feedback imposes a major problem in terms of speech understanding and listening comfort. If these problems are not resolved some hearing aid users may even choose not to use their hearing aids. The overall objective of this dissertation is the design of DSP algorithms for hearing aids. The focus is on three main areas such as NR, AFC, and DRC. The DSP algorithms considered here are all adaptive approaches which is important when dealing with time-varying acoustic environments, reverberation, and non-stationary signals such as speech and multi-talker babble.

10.1.1 Noise reduction

The research on the NR problem has been focussed on the SDW-MWF $_{\mu}$ since this approach offers many desirable features, i.e., no a priori information regarding

the desired signal model is required such as microphone configuration, speaker position, and also no calibration procedure is needed. However since the SDW-MWF $_{\mu}$ does not require any a priori information the performance of the SDW-MWF $_{\mu}$ highly depends on the accuracy of the estimated correlation matrices. It is therefore expected that the SDW-MWF $_{\mu}$ is more sensitive to inaccurate estimation of the correlation matrices, e.g., due to non-stationary signals, VAD error, and reverberation. Furthermore the SDW-MWF $_{\mu}$ is based on an assumption that the spatial and the spectral signal characteristic of the noise vary slowly over time typically in the order of seconds which is a problem since both speech and noise are non-stationary signals. Another problem is that the noise-only correlation matrix is kept fixed during speech-plus-noise periods and vice versa. This is a major problem when dealing with non-stationary noise sources since the speech correlation matrix is estimated by subtracting the noise-only correlation matrix from the speech-plus-noise correlation matrix. For this reason a rank-1 SDW-MWF $_{\mu}$ has been proposed which has shown to be less sensitive to errors in the correlation matrices. Furthermore, the SDW-MWF $_{\mu}$ offers a trade-off between NR and speech distortion by increasing a weighting factor μ . The problem is that the SNR improvement comes at a cost of a higher signal distortion. The reason for this is that the weighting factor is typically fixed for each frequency and for each frame.

In **Chapter 2** we have shown that the SDW-MWF $_{\mu}$ is able to improve the SNR, when μ is increased, but the cost is a higher signal distortion which is highly undesirable. However with the rank-1 SDW-MWF $_{\mu}$ the SNR improvement is comparable to SDW-MWF $_{\mu}$ with $\mu=5$ but the signal distortion is much lower. This makes the rank-1 SDW-MWF $_{\mu}$ an interesting approach which is used as a baseline for comparison for the research and development on the SDW-MWF $_{\mu}$. It is also shown how high and low reverberation scenarios negatively affect the SDW-MWF $_{\mu}$ based NR performance especially for high reverberation scenarios and at low input SNR. Also for certain spatial scenarios the SNR improvement is very limited. For this reason we will primarily focus on high reverberation and low input SNR scenarios.

In **Chapter 3** we have analyzed whether a weighting factor that is fixed for each frequency and for each frame is an optimal way to make the trade-off between NR and speech distortion. We have therefore proposed to incorporate the conditional SPP in the SDW-MWF $_{\mu}$, referred to as SDW-MWF $_{\text{SPP}}$, such that speech dominant segments and noise dominant segments are weighted differently. The conditional SPP is estimated and updated for each frequency and for each frame which also improves the tracking of the non-stationarity of the speech and the noise. As opposed to SDW-MWF $_{\text{SPP}}$ another flexible weighting factor is proposed that combines the conditional SPP with a per frame decision in order to make a flexible weighting between NR and speech distortion, which is referred to as SDW-MWF $_{\text{flex}}$. Through experiments it has been demonstrated that the SDW-

MWF_{SPP} and SDW-MWF_{flex} outperform the SDW-MWF _{μ} both in terms of SNR improvement and signal distortion. The same experiments were conducted with the rank-1 formulation which showed that the rank-1 SDW-MWF _{μ} could not be improved with the rank-1 SDW-MWF_{SPP} and the rank-1 SDW-MWF_{flex} without increasing the signal distortion. This suggests that with the rank-1 SDW-MWF _{μ} the optimal performance was immediately achieved.

In **Chapter 4** we have addressed the problem of how the correlation matrices are estimated and updated both in terms of the robustness and tracking capabilities. To improve the SDW-MWF _{μ} based NR we have proposed to estimate the correlation matrices with the use of prior knowledge such that the estimated correlation matrices are guaranteed to have a certain structure which also makes the corresponding filter valid. This prior knowledge is then combined with a continuous updating approach based on the conditional SPP which improves the tracking capabilities since the correlation matrices are now updated during noise-only periods and speech-plus-noise periods. The differences between the proposed method to estimate and update the correlation matrices have also been compared to the traditional method using a perfect VAD. Using the proposed estimation of the correlation matrices with the proposed weighting factors a novel SDW-MWF based NR was achieved that outperformed both the SDW-MWF _{μ} and the rank-1 SDW-MWF _{μ} both in terms of SNR improvement and signal distortion.

In **Chapter 5** we have presented a different multi-channel NR technique based on the well-known SCB also known as the MVDR beamformer. The problem with the SCB differs from the SDW-MWF _{μ} in the sense that the SDW-MWF _{μ} performance relies solely on the correlation matrices. The SCB performance on the other hand depends on the steering vector being accurately estimated. For this reason a RCB was proposed based on using an uncertainty principle such that the steering vector could be properly constrained and estimated. The proposed RCB is focussed on low computational complexity and small arrays which are important factors in hearing aids. The proposed RCB was compared to the SCB both in terms of SNR improvement, signal distortion and computation complexity. Through experimental simulations it has been shown how the RCB was able to outperform the SCB with a very low increase in computational complexity.

10.1.2 Combined noise reduction and dynamic range compression

Combining DSP algorithms for hearing aids is a challenging task and very important since different algorithms can interact in a negative way such that each algorithm that worked independently is now compromised. The problem of having NR and DRC combined is considered to be a major problem since these algorithms serve different purposes. The target of NR is to reduce the noise as

much as possible while preserving the speech. The DRC is basically designed to amplify sounds based on the input level without considering whether the signal contains speech or noise. This can degrade the overall performance especially if the reduced noise level is lower than the preserved speech level and the corresponding amplification is therefore higher for the noise. In this way the NR and the DRC counteract each other.

In **Chapter 6** the design of a DRC algorithm has been introduced and the attention has been focussed toward the scenario where the DRC operates in the presence of background noise. It was shown that the SNR was negatively affected when the DRC was operating in the presence of background noise. During the analysis it was shown how the SNR could be preserved if the DRC gain could be reduced in certain critical bands assuming that knowledge of the speech and the noise was available. This suggested that if the speech dominant segments and the noise dominant segments could be weighted differently the SNR degradation could be avoided. Through experimental simulations it was shown how the presence of background noise negatively affected the DRC and the SNR.

In **Chapter 7** we have presented a combined SDW-MWF based NR and DRC where we have analyzed the undesired interaction effect when the NR and the DRC operate together. It has been shown that the DRC negatively affects any SNR improvement achieved from the NR stage which is highly undesired. This happens because the DRC estimates the gain based solely on the input level such that any low level segments receive higher amplification compared to high level segments. Since the aim of NR is to reduce the noise while preserving the speech then from the DRC point of view the reduced noise is now being considered as a low level segment and is therefore amplified with a higher gain compared to the speech that supposedly should be preserved from the NR point of view. This then leads to the undesired SNR degradation. For this purpose we have proposed to incorporate the conditional SPP into the DRC which was referred to as the dual-DRC. Using the dual-DRC with the conditional SPP the gain is now modified such that the estimation of the DRC now takes speech dominant segments and noise dominant segments into account. Through experimental simulations it has been demonstrated that using the dual-DRC leads to less SNR degradation.

10.1.3 Feedback cancellation

Acoustic feedback is a problem that has a major effect on the performance of a hearing aid. Acoustic feedback is caused by the acoustic coupling between the loudspeaker and the microphone(s) such that amplified sound leaks out and get picked up by the microphone in the hearing aid creating a closed signal loop. Acoustic feedback limits the maximum gain that can be used in the hearing aid which can lead to problem compensating for the hearing loss resulting in

audibility problems. Furthermore, acoustic feedback is audible as a continuous high-frequency tone emanating from the hearing aid leading to a reduced listening comfort and sound quality.

In **Chapter 8** we have shown that the acoustic feedback problem can be dealt with by using a PEM-based AFC approach where LP was used to model the near-end signal. The inverse of the near-end signal model then acts as decorrelation prefilters when applied to the loudspeaker and the microphone signals. However it was shown that using the LP model alone did not provide sufficient decorrelation and therefore this approach did not provide a stable AFC performance. For this reason a cascaded near-end signal model was also proposed using a LP model combined with a PZLP model. Then the LP can model the noise components while the PZLP can model the tonal components. This resulted in an improved AFC performance in terms of MSG and filter misadjustment.

In **Chapter 9** we have improved the cascaded near-end signal model in PEM-based AFC by exploiting the characteristic of speech signals. The idea is to incorporate information such as pitch, amplitude, and the number of harmonics into the PEF used to model the near-end signal. The detection of voiced-unvoiced frames are also included in the cascaded near-end signal model such that in unvoiced frames only a single near-end signal model is used. This is based on the fact that the correlation is highest for voiced frames and therefore a cascaded near-end signal model is used when a voiced frame is detected. Furthermore, a single near-end signal model using LP also has a lower computational complexity compared to using a cascaded near-end signal model. Through experimental simulations it was shown that an accurate modelling of the near-end signal can result in improved PEM-based AFC performance both in terms of MSG and filter misadjustment.

10.2 Suggestions for further research

10.2.1 Noise reduction

Perceptual evaluation. In this dissertation, the experimental results were demonstrated by using objective measures such as intelligibility weighted SNR and signal distortion. However perceptual evaluation using both normal hearing and hearing impaired subjects needs to be validated before the actual benefit of the proposed SDW-MWF can be properly understood. It would in particular be interesting to evaluate if hearing aid users would benefit from the improved tracking with the proposed correlation matrices and the proposed weighting factors which should improve the spectral filtering. For normal hearing subjects the question is whether the inclusion of a non-linear weighting factor that changes

for each frequency and for each frame would be negatively perceived compared to using a fixed weighting factor.

Perceptual motivated weighting factor. In this dissertation, we have proposed to incorporate the conditional SPP in the SDW-MWF such that speech dominant segments and the noise dominant segments are weighted differently. It was shown that in certain scenarios the SNR improvement was achieved without increasing the signal distortion but for certain spatial scenarios the SNR improvement came at the cost of a higher signal distortion. In further research it would therefore be interesting to combine the proposed weighting factor with a perceptual motivated weighting factor. Inspiration can be taken from single-channel NR where masking threshold has been used in the NR process such that the noise is only attenuated when its level exceeds the masking threshold. In this way the signal distortion could be kept low without compromising the SNR improvement. Other information could also be incorporated in the SDW-MWF such as hearing models or hearing loss profile.

Improved spatial filtering in the MWF. In this dissertation, we have used a monaural hearing aid system with two microphones that are located at one single ear. Furthermore we have been focussed on the spectral filtering/tracking part of the SDW-MWF rather than the spatial part. However with the opportunity to exploit a wireless link between the left and the right hearing aid extensive research has been conducted in the area of binaural NR. In future research it would therefore be relevant to evaluate the proposed SDW-MWF in a binaural hearing aid system such that the spatial filtering could be improved. In addition to improving the spatial filter it would also be interesting to evaluate if the proposed SDW-MWF has any effect on the sound localization performance since past work has been focussed on the trade-off between NR and preservation of localization cues using a SDW-MWF setup. It would then be interesting to evaluate if the improved spectral filtering has any negative effect in terms of localization performance.

10.2.2 Combined noise reduction and dynamic range compression

In this dissertation, we have analyzed the undesired interaction effect when NR and DRC are combined. It has been shown that the DRC negatively affects the SNR improvement achieved from the NR stage. It has also been observed that the SNR degradation is worse at high input SNR since the noise in this case is now considered to be a low level segment and therefore a higher gain is applied compared to the speech which is considered to be a high level segment. In future research it would be interesting to fully integrate the NR with the DRC. A first step could be to limit the NR performance with the knowledge of the DRC parameters (e.g. CR, CT, attack and release time) such that a SNR is achieved that leads to

a minimal SNR degradation from the DRC. This can be achieved by exploiting the fact that the proposed weighting factor controls the spectral filtering of the SDW-MWF which means that the spectral filter needs to be integrated in the DRC.

10.2.3 Feedback cancellation

In this dissertation, several ways to estimate and model the near-end signal has been proposed which is used in the PEM-based AFC as a decorrelation process. We have shown that including relevant speech information such as pitch, amplitude, number of harmonics, and the knowledge of voiced-unvoiced frames in a cascaded near-end signal model leads to an improved AFC performance in terms of MSG and filter misadjustment. In future work it would be interesting to be able to directly link the estimation of the accuracy of the near-end signal model with the actual AFC performance. At this point the decorrelation of the loudspeaker and the microphone signal with the inverse near-end signal model is viewed as a separate process compared to the actual the AFC performance.

Combined noise reduction and feedback cancellation is also an area which should be explored since this could potentially change the design of the PEM-based AFC. For example the design of AFC algorithms are primarily performed using a single-microphone whereas NR typically uses a multi-microphone set-up. Future research could be in the direction of extending the PEM-based AFC with a single-channel NR and analyzing the interaction effect. At the same time the PEM-based AFC could also be extended to a multi-microphone scenario which then allows for a more sophisticated set-up with NR and AFC.

Combined NR and DRC with AFC is also an interesting area which again would completely change the design objectives. The development on the PEM-based AFC has been performed with a linear gain but a more realistic compensation is with the use of DRC. The interaction between the NR and DRC has already been shown and therefore a scenario with all components included would potentially reveal the actual performance that can be achieved.

Bibliography

- [1] Acoustical Society of America. ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index. June 1997.
- [2] M. C. Anderson, K. H. Arehart, and J. M. Kates. The acoustic and perceptual effects of series and parallel processing. *EURASIP Journal on Advances in Signal Processing*, 2009, Article ID 619805, 20 pages, 2009. doi:10.1155/2009/619805.
- [3] S. P. Applebaum and D. J. Chapman. Adaptive arrays with main beam constraints. *IEEE Transactions on Antennas and Propagation*, 24(5):650–662, September 1976.
- [4] B. Atal and L. Rabiner. A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(3):201–212, June 1976.
- [5] Auditec. Auditory Tests (Revised), Compact Disc, Auditec, St. Louis, 1997.
- [6] R. G. Bachu, S. Kopparthi, B. Adapa, and B. D. Barkana. Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In *IEEE International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE)*, 2008.
- [7] A. Beck and Y. C. Eldar. Doubly Constrained Robust Capon Beamformer With Ellipsoidal Uncertainty Sets. *IEEE Transactions on Signal Processing*, 55(2):753–758, February 2007.
- [8] R. Bentler, C. Palmer, and H. G. Mueller. Evaluation of a second-order directional microphone hearing aid: I. speech perception outcomes. *Journal of the American Academy of Audiology*, 17(3):179–189, 2006.
- [9] R. A. Bentler. Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence. *Journal of the American Academy of Audiology*, 16(7):473–484, 2005.

- [10] F. Beritelli, S. Casale, G. Rugeri, and S. Serrano. Performance evaluation and comparison of g.729/AMR/fuzzy voice activity detectors. *IEEE Signal Processing Letters*, 9(3):85–88, March 2002.
- [11] M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 208–211, April 1979.
- [12] F. A. Bilsen, W. Soede, and A. J. Berkhout. Development and assessment of two fixed-array microphones for use with hearing aids. *Journal of Rehabilitation Research and Development*, 30(1):73–81, 1993.
- [13] P. J. Blamey, D. S. Macfarlane, and B. R. Steele. An intrinsically digital amplification scheme for hearing aids. *EURASIP Journal on Applied Signal Processing*, 18:3026–3033, 2005.
- [14] P. J. Blamey, L. F. A. Martin, and H. J. Fiket. A digital processing strategy to optimize hearing aid outputs directly. *Journal of the American Academy of Audiology*, 15(10):716–728, 2004.
- [15] S. F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(2):113–120, April 1979.
- [16] C. Boukis, D. P. Mandic, and A. G. Constantinides. Toward bias minimization in acoustic feedback cancellation systems. *Journal of the Acoustical Society of America*, 121(3):1529–1537, March 2006.
- [17] H. Buchner, J. Benesty, and W. Kellermann. Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication. *Signal Processing*, 85(3):549–570, 2005.
- [18] K. Buckley. Broad-band beamforming and the generalized sidelobe canceller. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(5):1322–1323, October 1986.
- [19] D. K. Bustamante and L. D. Braida. Multiband compression limiting for hearing-impaired listeners. *Journal of Rehabilitation Research and Development*, 24(4):149–160, 1987.
- [20] D. K. Bustamante, T. L. Worrall, and M. J. Williamson. Measurement and adaptive suppression of acoustic feedback in hearing aids. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 2017–2020, May 1989.
- [21] D. Byrne. Key issues in hearing aid selection and evaluation. *Journal of the American Academy of Audiology*, 3(2):67–80, March 1992.

- [22] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, August 1969.
- [23] B. D. Carlson. Covariance matrix estimation errors and diagonal loading in adaptive arrays. *IEEE Transactions on Aerospace and Electronic Systems*, 24(4):397–401, July 1988.
- [24] Y. T. Chan, J. M. M. Lavoie, and J. B. Plant. A parameter estimation approach to estimation of frequencies of sinusoids. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(2):214–219, April 1981.
- [25] Y. M. Cheng and D. O’Shaughnessy. Speech enhancement based conceptually on auditory evidence. *IEEE Transactions on Signal Processing*, 39(9):1943–1954, 1991.
- [26] T. Y. C. Ching, H. Dillon, and D. Byrne. Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *Journal of the Acoustical Society of America*, 103(2):1128–1140, 1996.
- [27] M. G. Christensen and A. Jakobsson. *Multi-Pitch Estimation*. Morgan & Claypool, 2009.
- [28] C. Chun-Yang and P. P. Vaidyanathan. Quadratically constrained beamforming robust against direction-of-arrival mismatch. *IEEE Transactions on Signal Processing*, 55(8):4139–4150, Aug 2007.
- [29] K. Chung. Challenges and recent developments in hearing aids part i: Speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends In Amplification*, 8(3):83–124, 2004.
- [30] K. Chung. Effective compression and noise reduction configurations for hearing protectors. *Journal of the Acoustical Society of America*, 121(2):1090–1101, February 2007.
- [31] K. Chung and F. G. Zeng. Using hearing aid adaptive directional microphones to enhance cochlear implant performance. *Hearing Research*, 250(2009):27–37, 2009.
- [32] I. Claesson and S. Nordholm. A spatial filtering approach to robust adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, 40(9):1093–1096, 1992.
- [33] I. Cohen. Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. *IEEE Signal Processing Letters*, 9(4):113–116, April 2002.

- [34] I. Cohen. Analysis of two-channel generalized sidelobe canceller (GSC) with post-filtering. *IEEE Transactions on Speech and Audio Processing*, 11(6):684–699, November 2003.
- [35] I. Cohen. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing*, 11(5):466–475, Sep 2003.
- [36] I. Cohen and B. Berdugo. Speech enhancement for non-stationary noise environments. *Signal Processing*, 81(11):2403–2418, 2001.
- [37] M. T. Cord, R. K. Surr, B. E. Walden, and O. Dyrlund. Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. *Journal of the American Academy of Audiology*, 15(5):353–364, 2004.
- [38] B. Cornelis, M. Moonen, and J. Wouters. Performance analysis of multichannel Wiener filter based noise reduction in hearing aids under second order statistics errors. *Accepted for publication in IEEE Transactions on Audio, Speech and Language Processing*, 2010.
- [39] C. Crandell, T. L. Mills, and R. Gauthier. Knowledge, behaviors, and attitudes about hearing loss and hearing protection among racial/ethnically diverse young adults. *Journal of the National Medical Association*, 96(2):176–186, February 2004.
- [40] R. Curtis and R. Niederjohn. An investigation of several frequency-domain processing methods for enhancing the intelligibility of speech in wideband random noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 602–605, April 1978.
- [41] J. R. Deller, J. H. L. Hansen, and J. G. Proakis. *Discrete-Time Processing of Speech Signals*. IEEE Press., New York, 2 edition, 2000.
- [42] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1):1–38, 1977.
- [43] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters. The effect of multi-microphone noise reduction systems on sound source localization in binaural hearing aids. *Journal of the Acoustical Society of America*, 124(1):484–497, 2008.
- [44] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters. Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids. *Journal of the Acoustical Society of America*, 125(1):360–371, 2009.

- [45] H. Dillon. *Hearing Aids*. Boomerang Press, Thieme, 2001.
- [46] P. S. R. Diniz. *Adaptive Filtering Algorithms and Practical Implementation*. Kluwer Academic Publishers, 1997.
- [47] P. M. Djuric. Asymptotic MAP criteria for model selection. *IEEE Transactions on Signal Processing*, 46(7):2726–2735, October 1998.
- [48] S. Doclo and M. Moonen. GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Transactions on Signal Processing*, 50(9):2230–2244, September 2002.
- [49] S. Doclo, A. Spriet, J. Wouters, and M. Moonen. Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction. *Speech Communication*, 49(2007):636–656, August 2007.
- [50] K. Dong Kook, J. Keun Won, and C. Joon-Hyuk. A new statistical voice activity detection based on UMP test. *IEEE Signal Processing Letters*, 14(11):891–894, November 2007.
- [51] B. W. Edwards. Beyond amplification: Signal processing techniques for improving speech intelligibility in noise with hearing aids. *Seminars In Hearing*, 21(2):137–156, 2000.
- [52] G. Elko. *Superdirectional Microphone Arrays*. Chapter 10 in, *Acoustic Signal Processing for Telecommunications* (Gay S. L. and Benesty J. Eds.), pages 181–237, Kluwer Academic Publishers, Boston, 2000.
- [53] J. C. Ellison, F. P. Harris, and T. Muller. Interactions of hearing aid compression release time and fitting formula: Effects on speech acoustics. *American Academy of Audiology convention, Pennsylvania*, April 2002.
- [54] Y. Ephraim. Statistical-model-based speech enhancement systems. *Proceedings of the IEEE*, 80(10):1526–1555, October 1992.
- [55] Y. Ephraim and D. Malah. Speech enhancement using optimal non-linear spectral amplitude estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1118–1121, April 1983.
- [56] Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32(6):1109–1121, December 1984.
- [57] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(2):443–445, April 1985.

- [58] Y. Ephraim, D. Malah, and B. H. Juang. On the application of hidden markov models for enhancing noisy speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):1846–1856, December 1989.
- [59] Y. Ephraim and H. L. Van Trees. A signal subspace approach for speech enhancement. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 3(4):251–266, July 1995.
- [60] Y. Ephraim and H. L. van Trees. A spectrally-based signal subspace approach for speech enhancement. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 804–807, May 1995.
- [61] H. Fletcher. Auditory patterns. *Review of Modern Physics*, 12:47–65, 1940.
- [62] D. A. Florencio and H. S. Malvar. Multichannel filtering for optimum noise reduction in microphone arrays. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2001.
- [63] U. Forssell and L. Ljung. Closed-loop identification revisited. *Automatica*, 35(7):1215–1241, 1999.
- [64] M. Gabrea, E. Grivel, and M. Najun. A single microphone kalman filter-based noise canceller. *IEEE Signal Processing Letters*, 6(3):55–57, March 1999.
- [65] S. Gannot, D. Burshtein, and E. Weinstein. Iterative and sequential kalman filter-based speech enhancement algorithms. *IEEE Transactions on Speech and Audio Processing*, 6(4):373–385, July 1998.
- [66] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, 49(8):1614–1626, August 2001.
- [67] S. Gazor and Z. Wei. A soft voice activity detector based on a laplacian-gaussian model. *IEEE Transactions on Speech and Audio Processing*, 11(5):498–505, 2003.
- [68] J. D. Gibson, B. Koo, and S. D. Gray. Filtering of colored noise for speech enhancement and coding. *IEEE Transactions on Signal Processing*, 39(8):1732–1742, August 1991.
- [69] B. Gold and N. Morgan. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons, Inc., 1999.
- [70] S. Gordon-Salant. Hearing loss and aging: New research findings and clinical implications. *Journal of Rehabilitation Research and Development*, 2(4):9–24, 2005.

- [71] J. M. Górriz, J. Ramirez, J. C. Segura, and C. G. Puntonet. An effective cluster-based model for robust speech detection and speech recognition in noisy environments. *Journal of the Acoustical Society of America*, 120(1):470–481, 2006.
- [72] J. E. Greenberg. Modified LMS algorithms for speech processing with an adaptive noise canceller. *IEEE Transactions on Speech and Audio Processing*, 6(4):338–351, July 1998.
- [73] J. E. Greenberg, P. M. Peterson, and P. M. Zurek. Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *Journal of the Acoustical Society of America*, 94(5):3009–3010, November 1993.
- [74] L. J. Griffiths and C. W. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, 30(1):27–34, January 1982.
- [75] H. Gustafsson, S. E. Nordholm, and I. Claesson. Spectral subtraction using reduced delay convolution and adaptive averaging. *IEEE Transactions on Speech and Audio Processing*, 9(8):799–807, November 2001.
- [76] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski. New insights into the MVDR beamformer in room acoustics. *IEEE Transactions on Audio, Speech and Language Processing*, 18(1):158–170, January 2010.
- [77] J. A. Haigh and J. S. Mason. Robust voice activity detection using cepstral features. In *Proceedings of TENCON '93. IEEE Region 10 International Conference on Computers, Communications and Automation*, volume 3, pages 321–324, October 1993.
- [78] J. H. L. Hansen and M. A. Clements. Constrained iterative speech enhancement with application to speech recognition. *IEEE Transactions on Signal Processing*, 39(4):795–805, April 1991.
- [79] J. H. L. Hansen, V. Radhakrishnan, and K. H. Arehart. Speech enhancement based on generalized minimum mean square error estimators and masking properties of the auditory system. *IEEE Transactions on Audio, Speech and Language Processing*, 14(6):2049–2063, November 2006.
- [80] W. Hartmann. *Signals, Sound, and Sensation*. Modern Acoustics and Signal Processing AIP Press, 1997.
- [81] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Inc., 4 edition, 2002.
- [82] S. Haykin and K. J. R. Liu. *Handbook on Array Processing and Sensor Networks*. John Wiley & Sons, Inc., 2009.

- [83] J. Hellgren. Analysis of feedback cancellation in hearing aids with filtered-x LMS and the direct method of closed loop identification. *IEEE Transactions on Speech and Audio Processing*, 10(2):119–131, February 2002.
- [84] J. Hellgren, T. Lunner, and S. Arlinger. Variations in the feedback of hearing aids. *Journal of the Acoustical Society of America*, 106(5):2821–2833, November 1999.
- [85] J. Hellgren and F. Urban. Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification. *IEEE Transactions on Speech and Audio Processing*, 9(8):906–913, November 2001.
- [86] R. C. Hendriks, J. Jensen, and R. Heusdens. Noise tracking using DFT domain subspace decompositions. *IEEE Transactions on Audio, Speech and Language Processing*, 16(3):541–553, March 2008.
- [87] T. Herzke and V. Hohmann. Effects of instantaneous multiband dynamic compression on speech intelligibility. *EURASIP Journal on Applied Signal Processing*, 18:3034–3043, 2005.
- [88] O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Transactions on Signal Processing*, 47(10):2677–2684, October 1999.
- [89] C. Hsiang-Feng, X. G. Shawn, D. S. Sigfrid, and A. Abeer. Band-limited feedback cancellation with a modified filtered-X LMS algorithm for hearing aids. *Speech Communication*, 39(1-2):147–161, 2003.
- [90] Y. Hu and P. C. Loizou. A generalized subspace approach for enhancing speech corrupted by colored noise. *IEEE Transactions on Speech and Audio Processing*, 11(4):334–341, 2003.
- [91] Y. Hu and P. C. Loizou. Environment-specific noise suppression for improved speech intelligibility by cochlear implants users. *Journal of the Acoustical Society of America*, 127(6):3689–3695, 2010.
- [92] L. E. Humes, L. Christensen, T. Thomas, F. H. Bees, A. Hedley-Williams, and R. Bentler. A comparison of the aided performance and benefit provided by a linear and a two-channel wide dynamic range compression hearing aid. *Journal of Speech, Language, and Hearing Research*, 42:65–79, 1999.
- [93] O. L. Frost III. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60(8):926–935, August 1972.
- [94] F. Jabloun and B. Champagne. Incorporating the human hearing properties in the signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 11(6):700–708, November 2003.

- [95] L. Jae and A. Oppenheim. All-pole modeling of degraded speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(3):197–210, June 1978.
- [96] J. D. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE Journal on Selected Areas in Communications*, 6(2):314–323, February 1988.
- [97] S. Jongseo, K. Nam Soo, and S. Wonyong. A statistical model-based voice activity detection. *IEEE Signal Processing Letters*, 6(1):1–3, January 1999.
- [98] C. Joon-Hyuk, K. Nam-Soo, and S.K. Mitra. Voice activity detection based on multiple statistical models. *IEEE Transactions on Signal Processing*, 54(6):1965–1976, 2006.
- [99] J. C. Junqua, B. Reaves, and B. Mak. A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognize. In *European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 1371–1374, Genove, Italy, 1991.
- [100] M. Kahrs and K. Brandenburg, editors. *Applications of digital signal processing to audio and acoustics*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.
- [101] S. Kamerling, K. Janse, and F. van der Meulen. A new way of acoustic feedback suppression. In *in preprints AES 105th Convention*, Amsterdam, The Netherlands, May 1998.
- [102] S. D. Kampe and M. K. Wynne. The influence of venting on the occlusion effect. *The Hearing Journal*, 49(4):59–66, 1996.
- [103] J. Kates. Toward a theory of optimal hearing aid processing. *Journal of Rehabilitation Research and Development*, 30:39–48, 1993.
- [104] J. Kates. Room reverberation effects in hearing aid feedback cancellation. *Journal of the Acoustical Society of America*, 109(1):367–378, January 2000.
- [105] J. M. Kates. Feedback cancellation in hearing aids: results from a computer simulation. *IEEE Transactions on Signal Processing*, 39(3):553–562, March 1991.
- [106] J. M. Kates. An evaluation of hearing-aid array processing. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 15–18, October 1995.
- [107] J. M. Kates. Constrained adaptation for feedback cancellation in hearing aids. *Journal of the Acoustical Society of America*, 106(2):1010–1019, 1999.

- [108] J. M. Kates. Principles of digital dynamic-range compression. *Trends In Amplification*, 9(2):45–75, 2005.
- [109] J. M. Kates. *Digital Hearing Aids*. Plural Publishing, 2008.
- [110] J. M. Kates and K. H. Arehart. Multichannel dynamic-range compression using digital frequency warping. *EURASIP Journal on Applied Signal Processing*, 18:3003–3014, 2005.
- [111] G. Keidser and F. Grant. The preferred number of channels (one, two, or four) in NAL-NL1 prescribed wide dynamic range compression (WDRC) devices. *Ear and Hearing*, 22:516–527, 2001.
- [112] Y. L. Ki and J. Souhwan. Time-domain approach using multiple kalman filters and em algorithm to speech enhancement with nonstationary noise. *IEEE Transactions on Speech and Audio Processing*, 8(3):282–291, May 2000.
- [113] M. C. Killion. SNR loss: I can hear what people say, but i can’t understand them. *The Hearing Review*, 4(12):8–14, December 1997.
- [114] S. B. Knebel and R. A. Bentler. Comparison of two digital hearing aids. *Ear and Hearing*, 19(4):280–289, August 1998.
- [115] S. Kochkin. Marketrak v: ”why my hearing aids are in the drawer”: The consumer’s perspective. *The Hearing Journal*, 2000.
- [116] S. Kochkin. Marketrak vi: Consumers rate improvements sought in hearing instruments. *The Hearing Review*, 2002.
- [117] S. Kochkin. Marketrak viii: 25-year trends in the hearing health market. *Hearing review*, 16(11):12–31, 2009.
- [118] B. Kollmeier, J. Peissig, and V. Hohmann. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *Journal of Rehabilitation Research and Development*, 30(1):82–94, 1993.
- [119] F. Kuk, D. Keenan, C-C. Lau, and C. Ludvigsen. Performance of a fully adaptive directional microphone to signals presented from various azimuths. *Journal of the American Academy of Audiology*, 16(6):333–347, 2005.
- [120] F. K. Kuk. Hearing aid design considerations for optimally fitting the youngest patients. *The Hearing Journal*, 52(4), April 1998.
- [121] W. Kyoung-Ho, Y. Tae-Young, P. Kun-Jung, and L. Chungyong. Robust voice activity detection algorithm for estimating noise spectrum. *Electronics Letters*, 36(2):180–181, January 2000.

- [122] L. H. Lei, J. P. Lie, A. B. Gershman, and C. M. S. See. Robust Adaptive Beamforming in Partly Calibrated Sparse Sensor Arrays. *IEEE Transactions on Signal Processing*, 58(3):1661–1667, March 2010.
- [123] E. LePage. Occupational noise-induced hearing loss: Origin, characterisation and prevention. *Acoustics Australia*, 26:57–61, 1998.
- [124] H. Levitt. Noise reduction in hearing aids: a review. *Journal of Rehabilitation Research and Development*, 38(1):111–121, January 2001.
- [125] J. Li, P. Stoica, and Z. Wang. On Robust Capon Beamforming and Diagonal Loading. *IEEE Transactions on Signal Processing*, 51(7):1702–1715, 2003.
- [126] J. Li, P. Stoica, and Z. Wang. Doubly Constrained Robust Capon Beamformer. *IEEE Transactions on Signal Processing*, 52(9):2407–2423, 2004.
- [127] M. Li, H. G. McAllister, N. D. Black, and T. A. De Perez. Wavelet-based nonlinear AGC method for hearing aid loudness compensation. *IEE Proceedings Vision, Image and Signal Processing*, 147(6):502–507, December 2000.
- [128] J. S. Lim and A. V. Oppenheim. Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 67(12):1586–1604, December 1979.
- [129] M. E. Lockwood and D. L. Jones. Beamformer performance with acoustic vector sensors in air. *Journal of the Acoustical Society of America*, 119(1):1–11, January 2006.
- [130] M. E. Lockwood, D. L. Jones, R. C. Bilger, C. R. Lansing, W. D. O’Brien, B. C. Wheeler, and A. S. Feng. Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms. *Journal of the Acoustical Society of America*, 115(1):379–391, January 2004.
- [131] P. Lockwood and J. Boudy. Experiments with a nonlinear spectral subtractor (NSS), hidden markov models and the projection, for robust speech recognition in cars. *Speech Communication*, 11(2-3):215–228, 1992.
- [132] P. Lockwood, J. Boudy, and M. Blanchet. Non-linear spectral subtraction (NSS) and hidden markov models for robust speech recognition in car noise environments. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 265–268, March 1992.
- [133] P. C. Loizou. *Speech Enhancement: Theory and Practice*. CRC Press, Boca Raton, FL, 2007.
- [134] R. G. Lorenz and S. P. Boyd. Robust Minimum Variance Beamforming. *IEEE Transactions on Signal Processing*, 53(5):1684–1696, May 2005.

- [135] T. Lotter and P. Vary. Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model. *EURASIP Journal on Applied Signal Processing*, 7:1110–1126, 2005.
- [136] H. Luts, K. Eneman, J. Wouters, M. Schulte, M. Vormann, M. Buechler, N. Dillier, R. Houben, W. A. Dreschler, M. Froelich, H. Puder, G. Grimm, V. Hohmann, A. Leijon, A. Lombard, D. Mauler, and A. Spriet. Multicenter evaluation of signal enhancement algorithms for hearing aids. *Journal of the Acoustical Society of America*, 127(3):1491–1505, March 2010.
- [137] J-B. Maj, L. Royackers, J. Wouters, and M. Moonen. Comparison of adaptive noise reduction algorithms in dual microphone hearing aids. *Speech Communication*, pages 957–9790, December 2006.
- [138] J. D. Markel and A. H. Gray, Jr. *Linear prediction of speech*. Springer-Verlag, New York, 1976.
- [139] C. Marro, Y. Mahieux, and K. U. Simmer. Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering. *IEEE Transactions on Speech and Audio Processing*, 6(3):240–259, May 1998.
- [140] R. Martin. Spectral subtraction based on minimum statistics. In *Proc. 7th European Signal Processing Conference (EUSIPCO)*, volume 94, pages 1182–1185, 1994.
- [141] R. Martin. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing*, 9(5):504–512, 2001.
- [142] J. A. Maxwell and P. M. Zurek. Reducing acoustic feedback in hearing aids. *IEEE Transactions on Speech and Audio Processing*, 3(4):304–313, July 1995.
- [143] J. J. May. Occupational hearing loss. *American Journal of Industrial Medicine*, 37:112–120, 2000.
- [144] R. McAulay and M. Malpass. Speech enhancement using a soft-decision noise suppression filter. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28(2):137–145, April 1980.
- [145] I. McLoughlin. *Applied Speech and Audio Processing*. Cambridge University Press, 2003.
- [146] U. Mittal and N. Phamdo. Signal/noise KLT based approach for enhancing speech degraded by colored noise. *IEEE Transactions on Speech and Audio Processing*, 8(2):159–167, March 2000.

- [147] B. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, 5th ed edition, 2003.
- [148] B. C. Moore. Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear and Hearing*, 17(2):133–161, 1996.
- [149] B. C. J. Moore. Design and evaluation of a two-channel compression hearing aid. *Journal of Rehabilitation Research and Development*, 24(4):181–192, 1987.
- [150] B. C. J. Moore. Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms. *Speech Communication*, 41(1):81–91, August 2003.
- [151] B. C. J. Moore, R. W. Peters, and M. A. Stone. Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *Journal of the Acoustical Society of America*, 105(1):400–411, January 1999.
- [152] K. Nam Soo and C. Joon-Hyuk. Spectral enhancement based on global soft decision. *IEEE Signal Processing Letters*, 7(5):108–110, May 2000.
- [153] E. Nemer, R. Goubran, and S. Mahmoud. Robust voice activity detection using higher-order statistics in the LPC residual domain. *IEEE Transactions on Speech and Audio Processing*, 9(3):217–231, March 2001.
- [154] A. Neuman, M. Bakke, S. Hellman, and H. Levitt. Effect of compression ratio in a slow-acting compression hearing aid: paired-comparison judgments of quality. *Journal of the Acoustical Society of America*, 96(3):1471–1478, 1994.
- [155] K. Ngo, S. Doclo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen. An integrated approach for noise reduction and dynamic range compression in hearing aids. In *Proc. 16th European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, August 2008.
- [156] K. Ngo, D. L. Jones, and M. Moonen. A low-complexity robust capon beamformer for small arrays. *Submitted for publication*, 2010.
- [157] K. Ngo, M. Moonen, J. Wouters, and S. H. Jensen. The effect of background noise on dynamic range compression in hearing aids. 2011. ESAT-SISTA Technical Report TR 11-71, Katholieke Universiteit Leuven, Belgium.
- [158] K. Ngo, M. Moonen, J. Wouters, and S. H. Jensen. A flexible speech distortion weighted multi-channel Wiener filter for noise reduction in hearing aids. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011.

- [159] K. Ngo, M. Moonen, J. Wouters, and S. H. Jensen. Improved robustness and tracking capabilities for multi-channel Wiener filter based noise reduction in hearing aids. *To be submitted*, 2011.
- [160] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen. Variable speech distortion weighted multichannel Wiener filter based on soft output voice activity detection for noise reduction in hearing aids. In *Proc. 11th International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, 2008.
- [161] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen. Incorporating the conditional speech presence probability in multi-channel Wiener filter based noise reduction in hearing aids. *EURASIP Journal on Advances in Signal Processing*, 2009, Article ID 930625, 11 pages, 2009. doi:10.1155/2009/930625.
- [162] K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen. A combined multi-channel Wiener filter based noise reduction and dynamic range compression in hearing aids. *Submitted for publication*, 2010.
- [163] K. Ngo, T. van Waterschoot, M. G. Christensen, S. H. Jensen M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids using a sinusoidal near-end signal model. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, USA, March 2010.
- [164] K. Ngo, T. van Waterschoot, M. G. Christensen, S. H. Jensen M. Moonen, and J. Wouters. Prediction-error-method-based adaptive feedback cancellation in hearing aids using pitch estimation. In *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, August 2010.
- [165] K. Ngo, T. van Waterschoot, M. G. Christensen, S. H. Jensen M. Moonen, and J. Wouters. Adaptive feedback cancellation with improved prediction error filters in hearing aids. *To be submitted*, 2011.
- [166] J. L. Nielsen and U. P. Svensson. Performance of some linear time-varying systems in control of acoustic feedback. *Journal of the Acoustical Society of America*, 1(106):240–254, 1999.
- [167] M. Nilsson, S. D. Soli, and A. Sullivan. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95(2):1085–1099, February 1994.
- [168] S. Nordebo, I. Claesson, and S. Nordholm. Weighted chebyshev approximation for the design of broadband beamformers using quadratic programming. *IEEE Signal Processing Letters*, 1(7):103–105, 1994.

- [169] S. Nordholm, I. Claesson, and M. Dahl. Adaptive microphone array employing calibration signals: an analytical evaluation. *IEEE Transactions on Speech and Audio Processing*, 7(3):241–252, May 1999.
- [170] J. Oullette. The incredible shrinking microphone. *The industrial physicist*, 5(4):7–9, 1999.
- [171] K. Paliwal and A. Basu. A speech enhancement method based on kalman filtering. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 12, pages 177–180, April 1987.
- [172] T. Petersen and S. Boll. Acoustic noise suppression in the context of a perceptual model. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 6, pages 1086–1088, April 1981.
- [173] P. M. Peterson, N. I. Durlach, W. M. Rabinowitz, and P. M. Zurek. Multimicrophone adaptive beamforming for interference reduction in hearing aids. *Journal of Rehabilitation Research and Development*, 24(4):103–110, 1987.
- [174] A. Pezeshki, B.D. Van Veen, L.L. Scharf, H. Cox, and M.L. Nordenvaad. Eigenvalue beamforming using a multirank MVDR beamformer and subspace selection. *IEEE Transactions on Signal Processing*, 56(5):1954–1967, May 2008.
- [175] R. Plomb. The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *Journal of the Acoustical Society of America*, 83:2322–2327, 1988.
- [176] B. Porat. *A course in digital signal processing*. John Wiley & Sons, Inc., 1997.
- [177] J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles, Algorithms and Applications*. Prentice Hall., 3 edition, 1995.
- [178] L. Rabiner and M. Sambur. Voiced-unvoiced-silence detection using the Itakura LPC distance measure. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 323–326, May 1977.
- [179] L. R. Rabiner and M. R. Sambur. An algorithm for determining the endpoints for isolated utterances. *The Bell System Technical Journal*, 54(2):297–315, February 1975.
- [180] B. Rafaely and S. J. Elliot. A computationally efficient frequency-domain LMS algorithm with constraints on the adaptive filter. *IEEE Transactions on Signal Processing*, 48(6):1649–1655, June 2000.

- [181] B. Rafaely, M. Roccasalva-Firenze, and E. Payne. Feedback path variability modeling for robust hearing aids. *Journal of the Acoustical Society of America*, 107(5):2665–2673, May 2000.
- [182] K. J. Raghunath and V. U. Reddy. Finite data performance analysis of MVDR beamformer with and without spatial smoothing. *IEEE Transactions on Signal Processing*, 40(11):2726–2736, November 1992.
- [183] J. Ramirez, J. M. Gorriz, J. C. Segura, C. G. Puntonet, and A. J. Rubio. Speech/non-speech discrimination based on contextual information integrated bispectrum LRT. *IEEE Signal Processing Letters*, 13(8):497–500, August 2006.
- [184] J. Ramirez, J. C. Segura, C. Benitez, L. Garcia, and A. Rubio. Statistical voice activity detection using a multiple observation likelihood ratio test. *IEEE Signal Processing Letters*, 12(10):689–692, October 2005.
- [185] J. Ramirez, J.C. Segura, C. Benitez, A. de la Torre, and A.J. Rubio. A new kullback-leibler VAD for speech recognition in noise. *IEEE Signal Processing Letters*, 11(2):266–269, February 2004.
- [186] J. Ramirez, P. Yelamos, J. M. Gorriz, and J. C. Segura. SVM-based speech endpoint detection using contextual speech features. *Electronics Letters*, 42(7):426–428, 2006.
- [187] A. Rezayee and S. Gazor. An adaptive KLT approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 9(2):87–95, February 2001.
- [188] C. Ris and S. Dupont. Assessing local noise level estimation methods: Application to noise robust ASR. *Speech Communication*, 34(1-2):141–158, 2001.
- [189] G. Rombouts and M. Moonen. QRD-based unconstrained optimal filtering for acoustic noise reduction. *Signal Processing*, 83(9):1889–1904, 2003.
- [190] G. Rombouts, T. van Waterschoot, K. Struyve, and M. Moonen. Acoustic feedback suppression for long acoustic paths using a nonstationary source model. *IEEE Transactions on Signal Processing*, 54(9):3426–3434, September 2006.
- [191] J. D. Royster and L. H. Royster. Amplified music and its effect on hearing. *Hearing Instruments*, 41:28–29, 1990.
- [192] R. W. Schafer and L. R. Rabiner. Digital representations of speech signals. *Proceedings of the IEEE*, 63(4):662–677, April 1975.

- [193] T. Schneider and R. Brennan. A multichannel compression strategy for a digital hearing aid. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1:411–414, Apr 1997.
- [194] M. R. Schroeder. Improvement of acoustic-feedback stability by frequency shifting. *Journal of the Acoustical Society of America*, 39(9):1718–1724, 1976.
- [195] B. Shield. Evaluation of the social and economic costs of hearing impairment. *Hear-it AISBL*, 2006.
- [196] N. A. Shusina and B. Rafaely. Unbiased adaptive feedback cancellation in hearing aids by closed-loop identification. *IEEE Transactions on Audio, Speech and Language Processing*, 14(2):658–665, march 2006.
- [197] M.G. Siqueira and A. Alwan. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Transactions on Audio, Speech and Language Processing*, 8(4):443–453, July 2000.
- [198] M. Souden, J. Benesty, and S. Affes. On optimal frequency-domain multichannel linear filtering for noise reduction. *IEEE Transactions on Audio, Speech and Language Processing*, 18(2):260–276, February 2010.
- [199] M. Souden, J. Chen, J. Benesty, and S. Affes. An integrated solution for online multichannel noise tracking and reduction. *Accepted for publication in IEEE Transactions on Audio, Speech and Language Processing*, 2011.
- [200] M. Souden, C. Jingdong, J. Benesty, and S. Affes. Gaussian model-based multichannel speech presence probability. *IEEE Transactions on Audio, Speech and Language Processing*, 18(5):1072–1077, 2010.
- [201] P. E. Souza. Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends In Amplification*, 6(4):131–159, 2002.
- [202] P. E. Souza, L. M. Jenstad, and K. T. Boike. Measuring the acoustic effects of compression amplification on speech in noise (1). *Journal of the Acoustical Society of America*, 119(1):41–44, January 2006.
- [203] A. Spanias, T. Painter, and A. Venkatraman. *Audio Signal Processing and Coding*. John Wiley & Sons, Inc., 2007.
- [204] A. Spriet, L. Van Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. Van Dijk, A. Van Wiering, and J. Wouters. Speech understanding in background noise with the two-microphone adaptive beamformer beamTM in the nucleus freedomTM cochlear implant system. *Ear and Hearing*, 28(1):62–72, February 2007.

- [205] A. Spriet, M. Moonen, and J. Wouters. Robustness analysis of GSVD based optimal filtering and generalized sidelobe canceller for hearing aid applications. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 31–34, October 2001.
- [206] A. Spriet, M. Moonen, and J. Wouters. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. *Signal Processing*, 84(12):2367–2387, 2004.
- [207] A. Spriet, M. Moonen, and J. Wouters. Stochastic gradient based implementation of spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction in hearing aids. *IEEE Transactions on Signal Processing*, 53(3):911–925, March 2005.
- [208] A. Spriet, M. Moonen, and J. Wouters. Evaluation of feedback reduction techniques in hearing aids based on physical performance measures. *Journal of the Acoustical Society of America*, 128(3):1245–1261, 2010.
- [209] A. Spriet, I. Proudler, M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal. *IEEE Transactions on Signal Processing*, 53(10):3749–3763, October 2005.
- [210] A. Spriet, G. Rombouts, M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids. *Journal of the Franklin Institute*, 343:545–573, 2006.
- [211] R. W. Stadler and W. M. Rabinowitz. On the potential of fixed arrays for hearing aids. *Journal of the Acoustical Society of America*, 94(3):1332–1342, 1993.
- [212] V. Stahl, A. Fischer, and R. Bippus. Quantile based noise estimation for spectral subtraction and Wiener filtering. In *Proc. 24th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 81, pages 1875–1878, 2000.
- [213] M. R. Stinson and G. A. Daigle. Effect of handset proximity on hearing aid feedback. *Journal of the Acoustical Society of America*, 115(3):1147–1156, March 2003.
- [214] P. Stoica, Z. Wang, and J. Li. Robust Capon Beamforming. *IEEE Signal Processing Letters*, 10(6):172–175, 2003.
- [215] M. A. Stone, B. C. J. Moore, J. I. Alcántara, and B. R. Glasberg. Comparison of different forms of compression using wearable digital hearing aids. *Journal of the Acoustical Society of America*, 106(6):3603–3619, December 1999.
- [216] K. E. Strom. Hearing instrument dispensing in 2000: The HR dispenser survey. *Hearing Reviews*, 8(6):20–42, 2001.

- [217] K. E. Strom. The HR 2006 dispenser survey. *Hearing Reviews*, 13(6):16–38, 2006.
- [218] C. Sydow. Broadband beamforming for a microphone array. *Journal of the Acoustical Society of America*, 96(2):845–849, August 1994.
- [219] R. Tahmasbi and S. Rezaei. A soft voice activity detection using GARCH filter and variance gamma distribution. *IEEE Transactions on Audio, Speech and Language Processing*, 15(4):1129–1134, May 2007.
- [220] K. Takao, H. Fujita, and T. Nishi. An adaptive antenna array under directional constraint. *IEEE Transactions on Antennas and Propagation*, 24(5):662–669, September 1976.
- [221] S. G. Tanyer and H. Ozer. Voice activity detection in nonstationary noise. *IEEE Transactions on Speech and Audio Processing*, 8(4):478–482, July 2000.
- [222] C. Y. Tseng and L. J. Griffiths. A unified approach to the design of linear constraints in minimum variance adaptive beamformers. *IEEE Transactions on Antennas and Propagation*, 40(12):1533–1542, December 1992.
- [223] R. Tucker. Voice activity detection using a periodicity measure. *IEE Proceedings - Part I: Communications, Speech and Vision*, 139(4):377–380, August 1992.
- [224] P. P. Vaidyanathan. *Multirate systems and filter banks*. Prentice Hall., 1993.
- [225] R. A. van Buuren, J. M. Festen, and T. Houtgast. Compression and expansion of the temporal envelope: Evaluations of speech intelligibility and quality. *Journal of the Acoustical Society of America*, 105:2903–2913, 1999.
- [226] T. van Waterschoot and M. Moonen. A pole-zero placement technique for designing second-order IIR parametric equalizer filters. *IEEE Transactions on Audio, Speech and Language Processing*, 15(8):2561–2565, November 2007.
- [227] T. van Waterschoot and M. Moonen. Constrained pole-zero linear prediction: an efficient and near-optimal method for multi-tone frequency estimation. In *Proc. 16th European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, August 2008.
- [228] T. van Waterschoot and M. Moonen. Adaptive feedback cancellation for audio applications. *Signal Processing*, 89(11):2185–2201, November 2009.
- [229] T. van Waterschoot and M. Moonen. Fifty years of acoustic feedback control: state of the art and future challenges. *Proceedings of the IEEE*, 99(2):288–327, February 2011.

- [230] B. D. Van Veen. Beamforming: a versatile approach to spatial filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing Magazine*, 5(2):4–24, April 1988.
- [231] E. Villchur. Signal processing to improve speech intelligibility in perceptive deafness. *Journal of the Acoustical Society of America*, 53(6):1646–1657, 1973.
- [232] N. Virag. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing*, 7(2):126–137, March 1999.
- [233] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo. Robust Adaptive Beamforming Using Worst-Case Performance Optimization: A Solution to the Signal Mismatch Problem. *IEEE Transactions on Signal Processing*, 51(2):312–324, February 2003.
- [234] R. Wang and R. Harjani. Acoustic feedback cancellation in hearing aids. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 137–140, April 1993.
- [235] L. Wei, S. Weiss, and L. Hanzo. A generalized sidelobe canceller employing two-dimensional frequency invariant filters. *IEEE Transactions on Antennas and Propagation*, 53(7):2339–2343, 2005.
- [236] M. W. White. Compression systems for hearing aids and cochlear prostheses. *Journal of Rehabilitation Research and Development*, 23(1):25–39, 1986.
- [237] B. Widrow, Jr. Glover, J.R., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, Jr. Eugene Dong, and R.C. Goodlin. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12):1692–1716, dec 1975.
- [238] B. Widrow and F. L. Luo. Microphone arrays for hearing aids: An overview. *Speech Communication*, 39(2003):139–146, 2003.
- [239] D. B. Williams and V. Madisetti, editors. *Digital Signal Processing Handbook*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 1997.
- [240] K. Wooil, K. Sunmee, and K. Hanseok. Spectral subtraction based on phonetic dependency and masking effects. *IEE Proceedings Vision, Image and Signal Processing*, 147(5):423–427, October 2000.
- [241] S. Q. Wu and J. Y. Zhang. A new robust beamforming method with antennae calibration error. In *IEEE Wireless Communication and Networking Conference*, pages 869–872, New Orleans, USA, 1999.

- [242] K. F. C. Yiu, N. Grbic, T. Kok-Lay, and S. Nordholm. A new design method for broadband microphone arrays for speech input in automobiles. *IEEE Signal Processing Letters*, 9(7):222–224, July 2002.
- [243] N. B. Yoma, F. McInnes, and M. Jack. Robust speech pulse detection using adaptive noise modelling. *Electronics letters*, 1996.
- [244] Z. L. Yu and M. H. Er. A Robust Capon Beamformer against Uncertainty of Nominal Steering Vector. *EURASIP Journal on Applied Signal Processing*, Article ID 62327:1–8, 2006. doi:10.1155/ASP/2006/62327.
- [245] E. W. Yund, C. M. Roup, H. J. Simon, and G. A. Bowman. Acclimatization in wide dynamic range multichannel compression and linear amplification in hearing aids. *Journal of Rehabilitation Research and Development*, 43(4):517–536, 2006.
- [246] G. Zenton, T. Kah-Chye, and B. T. G. Tan. Kalman-filtering speech enhancement method based on a voiced-unvoiced speech model. *IEEE Transactions on Speech and Audio Processing*, 7(5):510–524, September 1999.
- [247] E. Zwicker. Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America*, 33(2):248–249, February 1961.

List of publications

International journal papers

- K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen., “Incorporating the conditional speech presence probability in multi-channel Wiener filter based noise reduction in hearing aids.”, *EURASIP Journal on Advances in Signal Processing*, Article, ID 930625, 11 pages, 2009, doi:10.1155/2009/930625.
- K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen., “A combined multi-channel Wiener filter based noise reduction and dynamic range compression in hearing aids.” *Submitted for publication*, 2010.
- K. Ngo, Douglas L. Jones, and M. Moonen. “A low-complexity robust Capon beamformer for small arrays,” *Submitted for publication*, 2010.
- K. Ngo, M. Moonen, J. Wouters, and S. H. Jensen., “Improved robustness and tracking capabilities for multi-channel Wiener filter based noise reduction in hearing aids.” *To be submitted*, 2011.
- K. Ngo, T. van Waterschoot, M. G. Christensen, S. H. Jensen M. Moonen, and J. Wouters., “Adaptive feedback cancellation with improved prediction error filters in hearing aids.” *To be submitted*, 2011.

International conference papers

- K. Ngo, S. Doclo, A. Spriet, M. Moonen, J. wouters, and S. H. Jensen., “An integrated approach for noise reduction and dynamic range compression in hearing aids.” In *Proc. 16th European Signal Processing Conference*

(*EUSIPCO*), Lausanne, Switzerland, August 2008.

- K. Ngo, A. Spriet, M. Moonen, J. Wouters, and S. H. Jensen., “Variable speech distortion weighted multichannel Wiener filter based on soft output voice activity detection for noise reduction in hearing aids.” In *Proc. 11th International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, 2008.
- K. Ngo, T. van Waterschoot, M. G. Christensen, S. H. Jensen M. Moonen, and J. Wouters., “Adaptive feedback cancellation in hearing aids using a sinusoidal near-end signal model.” In *Proc. 2010 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, Texas, USA, March 2010.
- K. Ngo, T. van Waterschoot, M. G. Christensen, S. H. Jensen M. Moonen, and J. Wouters. “Prediction-error-method-based adaptive feedback cancellation in hearing aids using pitch estimation.” In *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, August 2010.
- K. Ngo, M. Moonen, J. Wouters, and S. H. Jensen., “A flexible speech distortion weighted multi-channel Wiener filter for noise reduction in hearing aids.” In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2011.

Internal reports

- K. Ngo, M. Moonen, J. Wouters, and S. H. Jensen., “The effect of background noise on dynamic range compression in hearing aids.”, ESAT-SISTA Technical Report TR 11-71, Katholieke Universiteit Leuven, Belgium, 2011.

Curriculum vitae



Kim Ngo was born in Esbjerg, Denmark, in 1982. He received his Electrical Engineering degrees, B.Sc. in 2004 and M.Sc. in 2006 from Aalborg University (AAU), Denmark. From 2006 he has been a Ph.D candidate at the Department of Electrical Engineering, Katholieke Universiteit Leuven (K.U.Leuven), under supervision of Prof. Marc Moonen and Prof. Jan Wouters from the Department of Neurosciences, (K.U.Leuven).

From 2006 to 2009, he was involved in the EST-SIGNAL project funded by the EU Marie Curie programme. From 2009 to 2011 he was a Research Assistant with K.U.Leuven. During his graduate studies, he has also been a visiting researcher at Aalborg University (Prof. Søren Holdt Jensen) and University of Illinois at Urbana-Champaign (Prof. Douglas L. Jones). His research interests are in the area of digital signal processing techniques for speech and audio applications such as hearing aids, speech recognition, and audio classification. He is currently a reviewer of the Elsevier Signal Processing Journal, the European Signal Processing Conference, and the IEEE Transactions on Circuits and Systems.

Arenberg Doctoral School of Science, Engineering & Technology

Faculty of Engineering

Department of Electrical Engineering

Research group ESAT-SCD: SISTA/COSIC/DOCARCH

Kasteelpark Arenberg 10, B-3001 Leuven