

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)
In samenwerking met:
FACULTEIT GENEESKUNDE
DEPARTEMENT NEUROWETENSCHAPPEN
Kapucijnenvoer 33, 3000 Leuven

ADAPTIVE FILTERING TECHNIQUES FOR NOISE REDUCTION AND ACOUSTIC FEEDBACK CANCELLATION IN HEARING AIDS

Jury:
Prof. dr. ir. L. Froyen, voorzitter
Prof. dr. ir. M. Moonen, promotor
Prof. dr. J. Wouters, promotor
Prof. dr. ir. A. Leijon (KTH, Sweden)
Prof. dr. ir. I. Proudler (King's College and QinetiQ, UK)
Prof. dr. ir. P. Sommen (TU Eindhoven, the Netherlands)
Prof. dr. ir. D. Van Compernolle
Prof. dr. ir. S. Van Huffel

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Ann SPRIET

U.D.C. 534.77 September 2004

© Katholieke Universiteit Leuven – Faculteit Toegepaste Wetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag 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.

D/2004/7515/77

ISBN 90-5682-538-0

Voorwoord

Dit proefschrift zou niet tot stand zijn gekomen zonder de steun, medewerking en hulp van vele personen. Bij het begin van dit proefschrift zou ik dan ook een aantal personen willen bedanken.

In de eerste plaats gaat mijn oprechte dank uit naar mijn promotoren Prof. Marc Moonen en Prof. Jan Wouters. Ik had het geluk om niet door één maar door twee voortreffelijke promotoren ondersteund en begeleid te worden. Vijf jaar geleden maakten ze me warm voor een boeiende combinatie van twee onderzoeksdomeinen, nl. digitale signaalverwerking voor hoorapparaten en cochleaire implantaten. Dankzij hun enthousiasme, creativiteit en competentie op het vlak van beide onderzoeksdomeinen slaagde ik erin dit doctoraat tot een goed einde te brengen. Ik waardeer de vrijheid die ze me gegeven hebben om dit onderzoek uit te voeren alsook de kansen die ze me geboden hebben om aan conferenties en projecten deel te nemen. Verder wil ik Jan en Marc ook bedanken voor de waardevolle suggesties en hulp bij het schrijven van teksten en het oplossen van wetenschappelijke en niet-wetenschappelijke problemen, de stimulerende werkomgevingen, de jaarlijkse sociale activiteiten (zoals het befaamde Château Hilaire en de Holsbeekse Knabbel&Babbel) en de talrijke – zij het dan ietwat tevergeefse – pogingen om me op mijn West-Vlaamse spraakgebreken te wijzen.

Ik wens de leden van mijn begeleidingscommissie, Prof. Sabine Van Huffel en Prof. Dirk Van Compernolle, te bedanken voor het nalezen van het proefschrift en de waardevolle suggesties en aanvullingen. Ik bedank ook de andere juryleden voor hun onmiddellijke bereidheid om deel uit te maken van de examencommissie. I would like to thank Prof. Arne Leijon for his willingness to come over to Belgium for my Ph.D. defense, despite his busy lecture schedule. My gratitude also goes to Prof. Ian Proudler, not only for being a member of the jury, but also for the fruitful collaboration on acoustic feedback cancellation. Prof. Piet Sommen wens ik te bedanken voor zijn bereidheid om naar Leuven te komen, voor de interesse die hij in de afgelopen vijf jaar voor mijn doctoraatsonderzoek heeft betoond en voor de vlotte samenwerking bij de samenstelling van de Eurasip Newsletters. Verder dank ik Prof. Ludo Froyen

voor het waarnemen van het voorzitterschap van de examencommissie.

De vruchtbare en stimulerende samenwerking met Cochlear Technology Centre Europe heeft mijn doctoraatsonderzoek zeker verrijkt. Het drie-microfoons hoorapparaat dat Cochlear ons bezorgde, maakte het mogelijk om mijn theoretische bevindingen met reële data te valideren. Hiervoor dank ik Bas Van Dijk, Jan Janssen en hun collega's.

Ik ben het F.W.O.-Vlaanderen erkentelijk voor de financiële ondersteuning gedurende de eerste vier jaar van mijn doctoraat.

Tijdens mijn doctoraat kon ik zowel op SISTA als op het Lab Exp. ORL op de steun en hulp rekenen van talrijke collega's. Ik ben hen dankbaar voor de aangename sfeer tijdens en na de werkuren. In het bijzonder zou ik Simon, Geert, Jean-Baptiste, Koen, Ruben, Sharon, Toon en Thomas willen bedanken voor de goede samenwerking, de amusante momenten in de 01.80 (met of zonder wesp) en het lab, de real-time demo's, de talrijke opnames, de praktische tips en de waardevolle discussies.

Vrienden en familie, bedankt voor jullie interesse, vriendschap en het gezellige samenzijn! Een speciaal woordje van dank gaat uit naar Petra en Yves voor het verrassend-uit-de-hoek-komen, de culinaire hoogstandjes en de weetjes over salsa en het beloofde land.

Mama en Papa, Nelly en Dominiek: dank voor de goede zorgen, de warmte en het vertrouwen waarop ik iedere dag kan rekenen.

En, last but not least, bedankt Koen voor ... Je weet wel!

Ann Spriet
September 2004

Abstract

Understanding speech in noise and the occurrence of acoustic feedback belong to the major problems of current hearing aid users. Hence, an urgent demand exists for efficient and well-working digital signal processing algorithms that offer a solution to these issues. In this thesis we develop adaptive filtering techniques for noise reduction and acoustic feedback cancellation. Thanks to the availability of low power digital signal processors, these algorithms can be integrated in a hearing aid.

Because of the ongoing miniaturization in the hearing aid industry and the growing tendency towards multi-microphone hearing aids, robustness against imperfections such as microphone mismatch, has become a major issue in the design of a noise reduction algorithm. In this thesis we propose multi-microphone noise reduction techniques that are based on multi-channel Wiener filtering (MWF). Theoretical and experimental analysis demonstrate that these MWF-based techniques are less sensitive to imperfections than the widely studied Generalized Sidelobe Canceller (GSC). Thanks to their high robustness, they can fully exploit the benefit of an additional third microphone in a behind-the-ear hearing aid and hence reduce more noise. In addition, we show that a subband implementation further improves speech intelligibility. Finally, we establish a generalized noise reduction scheme, called the spatially pre-processed speech distortion weighted MWF, that encompasses the GSC and the MWF as special cases and allows for attractive in-between solutions. Low-cost subband and stochastic gradient implementations are proposed that make the implementation of the developed techniques feasible in future commercial hearing aids.

With the growing use of open fittings and the decreasing distance between the loudspeaker and the microphone of the hearing aid, the demand for effective feedback suppression techniques increases. A promising solution for acoustic feedback is adaptive feedback cancellation. However, the adaptive modelling problem encountered here appears to be highly non-trivial, because of the presence of a closed signal loop that introduces specific signal correlation. As a result, standard adaptive feedback cancellation techniques fail to provide a re-

liable feedback path estimate, when the desired signal is spectrally colored. In this thesis we develop adaptive feedback cancellation techniques that are based on closed-loop identification of the feedback path as well as an autoregressive modelling of the desired signal. Simulations show that the developed techniques provide a better feedback path estimate for spectrally colored and highly time-varying sound signals (such as speech) than the standard adaptive feedback cancellation techniques.

Korte inhoud

Het verstaan van spraak in ruis en het optreden van akoestische terugkoppeling ('feedback') vormen de voornaamste problemen voor hoorapparaatgebruikers. Er is bijgevolg een stijgende vraag naar efficiënte en doeltreffende digitale signaalverwerkingsalgoritmes die voor deze problemen een oplossing bieden. In dit proefschrift ontwikkelen we adaptieve signaalverwerkingstechnieken voor het onderdrukken van ruis en akoestische terugkoppeling in hoorapparaten. Dankzij de komst van DSPs met een laag vermogenverbruik kunnen deze algoritmes in een hoorapparaat ingebouwd worden.

Tengevolge van de toenemende miniaturisatie en het stijgend gebruik van multimicrofoon systemen in de hoorapparatenindustrie is de gevoeligheid van algoritmes voor modelleerfouten, zoals afwijkingen in de microfoonkarakteristieken, een belangrijk criterium bij het ontwerp van ruisonderdrukkingsalgoritmes geworden. In dit proefschrift stellen we multi-microfoon ruisonderdrukkingsalgoritmes voor die gebaseerd zijn op een adaptief meerkanaals Wiener-filter (MWF). Theoretisch en experimenteel onderzoek toont aan dat deze MWF-technieken minder gevoelig zijn voor modelleerfouten dan de 'Generalized Sidelobe Canceller' (GSC), een vaak bestudeerde adaptieve bundelvormingstechniek. Dankzij hun hogere robuustheid kunnen deze MWF-technieken het voordeel van een extra derde microfoon in een achter-het-oor hoorapparaat volledig uitbuiten en bijgevolg meer ruis onderdrukken. Een subbandgebaseerde implementatie resulteert in een verbetering van de spraakverstaanbaarheid. Tenslotte integreren we de klassieke GSC en het MWF in een globaal ruisonderdrukkingsschema, het spraakvervormingsgewogen MWF met spatiale 'preprocessing' genoemd. Afhankelijk van een aantal parameterinstellingen bekomt men één van beide algoritmes of een interessante tussenoplossing. We stellen goedkope stochastische-gradiëntimplementaties van het veralgemeende schema voor die een implementatie in huidige hoorapparaten mogelijk maken. Het performantievoordeel t.o.v. de klassieke GSC, de beperkte berekeningscomplexiteit en de minimale geheugenvereisten maken het een ideale kandidaat voor implementatie in toekomstige hoorapparaten.

De evolutie naar een open aanpassing van het hoorapparaat en de toenemende miniaturisatie verhogen de vraag naar doeltreffende feedbackonderdrukkingstechnieken. Een veelbelovende oplossing voor akoestische terugkoppeling is adaptieve feedbackonderdrukking. De filteradaptatie wordt echter aanzienlijk bemoeilijkt door de (typische) autocorrelatie-structuur van de te verwerken audiosignalen, waardoor er geen betrouwbare schatting van de terugkoppeling wordt verkregen. In dit proefschrift ontwikkelen we adaptieve feedbackonderdrukkingstechnieken die gebaseerd zijn op een gesloten-lus-identificatie van zowel de terugkoppeling als het spectrum van het gewenste audiosignaal. Simulaties tonen aan dat de ontwikkelde algoritmes een nauwkeurigere schatting van de terugkoppeling opleveren voor spectraal gekleurde en tijdsvariërende audiosignalen (zoals spraak) dan de bestaande adaptieve feedbackonderdrukkingstechnieken.

Glossary

Mathematical Notation

a	scalar a
$\mathbf{a}$	vector $\mathbf{a}$
$\mathbf{A}$	matrix $\mathbf{A}$
a^*	complex conjugate of 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 $\mathbf{A}$ or vector $\mathbf{a}$
$x[k]$	discrete-time filter, time-sequence, stochastic process
$X(f)$	Discrete-Time Fourier Transform of $x[k]$
$X_m[k]$	m-th DFT component of the frame $\begin{bmatrix} x[kL-L] & \cdots & x[kL+L-1] \end{bmatrix}^T$
$F(q)$	discrete-time filter with coefficient vector $\mathbf{f}$
$F(q,k)$	time-varying discrete-time filter with coefficient vector $\mathbf{f}[k]$
$G(f)$	power transfer function from the input of the noise reduction algorithm to the output
$G^s(f)$,$G^n(f)$,$G^i(f)$	PTF of speech, noise and spatially white noise (e.g., internal noise) from the algorithm's input to the output $z[k]$
$P_X(f)$	power spectral density of $x[k]$
$P_X^s(f)$, $P_X^n(f)$	PSD of the speech and noise component in $x[k]$, respectively.
$P_{XY}(f)$	cross-power spectral density of $x[k]$ and $y[k]$
$\Gamma_{kl}(f)$	complex coherence between the k-th and l-th microphone signal
$\mathbf{\Gamma}_{kl}^s(f)$,$\mathbf{\Gamma}_{kl}^n(f)$	complex coherence function between the speech/noise component in the k-th and l-th microphone signal
$\mathcal{O}(L)$	order L
$\mathcal{E}\{\cdot\}$	expectation operator
$\mathcal{F}\{\cdot\}$	Discrete-Time Fourier Transform operator

$\|\cdot\|_1$	1-norm
$\|\cdot\|_2$	2-norm
$\|\cdot\|$	absolute value
$\mathbb{N}$	set of natural numbers
$\mathbb{N}_0$	set of natural numbers except 0
$\mathbb{R}$	set of real numbers
$\mathbb{Z}$	set of integer numbers
$\mathbb{C}$	set of complex numbers
$\lceil x \rceil$	smallest integer larger than or equal to $x \in \mathbb{R}$
$\lfloor x \rfloor$	largest integer smaller than or equal to $x \in \mathbb{R}$.
$\otimes$	convolution
$\forall$	for all
$\in$	element of
$\mathrm{mod}(x, y)$	remainder after division of $x \in \mathbb{N}$ by $y \in \mathbb{N}_0$
$a \approx b$	a is approximately equal to b
$a \gg b$	a is much larger than b
$a \ll b$	a is much smaller than b

Fixed Symbols

c	speed of sound in air ($c \approx 340$ m/s)
d	microphone interspacing (uniform array)
d_c	(discrete-time) delay in the cancellation path
d_f	(discrete-time) delay in the forward path $G(q)$
d_t	sum of delay in the forward path and the cancellation path, i.e., $d_t = d_f + d_c$ (feedback)
d_l	distance of the l-th microphone to the center of the array
$e_p[k]$	prediction error
f	frequency-domain variable
f_i^c	center frequency of the ith one-third octave band
f_s	sampling frequency
f_0	pitch frequency of speech
$h_l[k]$	acoustic impulse response between the desired source and the l-th microphone
k, $\bar{k}$	discrete-time index
q	discrete-time delay operator, i.e., $q^{-1}x[k] = x[k-1]$
$r[k]$	external probe signal
$s_d[k]$	clean speech signal
$u[k]$	loudspeaker signal
$w[k]$	excitation of the desired signal model $H(q)$
$x[k]$	desired sound signal (feedback)
$x_l[k]$	l-th microphone signal

$x_l^s[k]$, $x_l^n[k]$	speech/noise component in the l-th microphone signal $x_l[k]$
$y[k]$	microphone signal (including the feedback signal)
$y_o[k]$	speech reference
$y_l[k]$	l-th noise reference
$z[k]$	output of the noise reduction algorithm
A_l	array length
$A(q)$	estimate of the inverse desired signal model $H^{-1}(q)$
$C(q)$	closed-loop transfer function from the desired signal $x[k]$ to the loudspeaker signal $y[k]$
D	downsampling ratio
$D(f,\theta)$	desired spatial directivity pattern of a fixed beamformer
$DI(f)$	directivity index at frequency f
$F(q), \mathbf{f}[k]$	true feedback path
$\hat{F}(q)$, $\hat{\mathbf{f}}[k]$	estimate of the feedback path $F(q)$
$\hat{F}_0(q)$	initial feedback path estimate
$G(q)$	forward path (i.e., signal processing path of the hearing aid)
$G_0(f)$	synthesis prototype filter
$H(q)$	true model of the desired signal $x[k]$
$\hat{H}(q)$	estimate of the desired signal model $H(q)$
$H(f,\theta)$	spatial directivity pattern of a fixed beamformer
$H_0(f)$	analysis prototype filter
$I_i(f)$	band importance function
K	number of subbands in a subband implementation
L_p	filter length of the prototype filter $H_0(f)$ of the analysis filter bank
L	filter length of FIR filters on microphones
M	number of microphones
$T_{60\mathrm{dB}}$	reverberation time
$\mathbf{d}(f,\theta)$	steering vector
$\mathbf{e}_l$	l-th canonical vector
$\mathbf{w}[k]$	stacked filter vector of multi-microphone noise reduction algorithm
$\bar{\mathbf{w}}[k]$	stacked filter vector of the SDW-MWF in the SP-SDW-MWF
$\mathbf{w}_a[k]$	stacked filter vector of the ANC
$\mathbf{w}_l[k]$	L-dimensional filter on the l-th microphone signal
$\bar{\mathbf{w}}_0[k]$	filter on the speech reference in the SP-SDW-MWF
$\mathbf{x}[k]$	stacked data vector
$\mathbf{x}^s[k]$, $\mathbf{x}^n[k]$	stacked speech/noise data vector

$\mathbf{x}_l[k]$	L-dimensional data vector of $x_l[k]$
$\mathbf{B}_y$, $\mathbf{B}_y^n$	circular speech + noise and noise data buffer, respectively
$\mathbf{F}$	DFT matrix, $\mathbf{F}(m,n) = e^{-j\frac{2\pi mn}{L}}$, $0 \leq m,n < L$
$\mathbf{I}_N$	$N \times N$-dimensional identity matrix
$\mathbf{S}_y^{lj}[k]$, $\mathbf{S}_y^{n,lj}[k]$	frequency-domain speech + noise and noise correlation matrix between the reference signals $y_l[k]$ and $y_j[k]$.
$\mathbf{W}[k]$	filter matrix
$\mathbf{X}(f)$	M-dimensional stacked data vector (frequency-domain)
$\mathbf{0}_N$	$N \times N$-dimensional zero matrix
$\alpha(f)$	input noise-to-signal ratio
$\beta(f)$	external-to-internal noise ratio
γ	forgetting factor of the power estimation
δ	speech detection error rate or error in estimated statistics due to non-stationarities in the noise
$\delta[k]$	Dirac impulse
ε_s^2, ε_n^2	speech distortion energy, residual noise energy
ζ^2	constraint value of the noise sensitivity constraint or the QIC
$\eta(f)$	ratio of injected noise power to input noise power of the NSC-GSC
θ	angle
Θ	angular region
κ_i	i-th reflection coefficient
λ	exponential weighting factor
λ_n	exponential weighting factor for the noise correlation matrix
λ_{s+n}	exponential weighting factor for the speech + noise correlation matrix
μ	trade-off parameter between speech distortion and noise reduction
τ	delay
τ_l	relative propagation delay of the lth microphone with respect to the first microphone
ρ	adaptive filter step size
$\bar{\rho}$	normalized adaptive filter step size
$\boldsymbol{\rho}$	step size matrix
Δ	discrete-time delay
ϕ_l	microphone phase mismatch at the l-th microphone w.r.t. the first microphone

v_l	microphone gain mismatch at the l-th microphone w.r.t. the first microphone
ω	angular frequency
Υ_l	microphone mismatch at the l-th microphone w.r.t. the first microphone
$\Psi(f)$	noise sensitivity
Ω	frequency region

Acronyms and Abbreviations

ADC	Analog-to-Digital Converter
ADMA	Adaptive Differential Microphone Array
AFC	Adaptive Feedback Cancellation
AI	Articulation Index
ANC	Adaptive Noise Canceller
AR	autoregressive
ASG	Added Stable Gain
BTE	Behind-The-Ear
CAF	Continuous Adaptation Feedback canceller
cf.	confer: see also
DAC	Digital-to-Analog Converter
dB SPL	Sound Pressure Level referenced to 20 μPa
DFT	Discrete Fourier Transform
DI	Directivity Index
DS	Delay-and-Sum
DSP	Digital Signal Processor
DTFT	Discrete-Time Fourier Transform
e.g.	exempli gratia: for example
etc.	et cetera: and so forth
FD-SG	Frequency-Domain Stochastic Gradient
FD-SG using CM	FD-SG algorithm using correlation matrices
FD-SG with LPF	FD-SG algorithm with Low Pass Filtering
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GSC	Generalized Sidelobe Canceller
GSVD	Generalized Singular Value Decomposition
i.e.	id est: that is
iff	if and only if
IIR	Infinite Impulse Response
ITC	In-The-Canal
ITE	In-The-Ear
IV	Instrumental Variable
LCMV	Linearly Constrained Minimum Variance

LMS	Least Mean Squares
LP	Linear Prediction
LPF	Low Pass Filter
LS	Least Squares
Mflops	Millions of Floating Point operations per Second
MIPS	Millions of Instructions per Second
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MSG	Maximum Stable Gain
MVDR	Minimum Variance Distortionless Response
MWF	Multi-channel Wiener Filter
NLMS	Normalized Least Mean Squares
NSC-GSC	GSC with Noise Sensitivity Constraint
pdf	Probability Density Function
PEM	Prediction Error Method
PEM-AFC	PEM based Adaptive Feedback Canceller
PR	Perfect Reconstruction
PSD	Power Spectral Density
QIC-GSC	GSC with Quadratic Inequality Constraint
QRD	QR Decomposition
RF	Radio-Frequency
RIV	Refined Instrumental Variable
RLS	Recursive Least Squares
$SD_{intellig}$	Intelligibility weighted Speech Distortion
SDR-GSC	Speech Distortion Regularized GSC
SDW-MWF	Speech Distortion Weighted Multi-channel Wiener Filter
SNR	Signal-to-Noise Ratio
$SNR_{intellig}$	Intelligibility weighted SNR
SP-SDW-MWF	Spatially Pre-processed Speech Distortion Weighted Multi-channel Wiener Filter
SRT	Speech Reception Threshold
TD-SG	Time-Domain Stochastic Gradient
TD-SG with LPF	TD-SG algorithm with Low Pass Filtering
TVAR	Time-Varying Autoregressive
WLS	Weighted Least Squares
w.r.t.	with respect to
2-ch AFC	Two-channel Adaptive Feedback Canceller

Contents

Voorwoord **i**

Abstract **iii**

Korte Inhoud **v**

Glossary **vii**

Contents **xiii**

Samenvatting **xxi**

1 Introduction **1**

1.1 Preliminaries 2

1.1.1 The auditory system 2

1.1.2 Hearing impairment 3

1.1.3 Hearing aids and cochlear implants 4

1.2 Noise reduction in hearing aids and cochlear implants 6

1.2.1 Signal processing challenges 6

1.2.2 State-of-the-art noise reduction techniques 9

1.3 Feedback suppression in hearing aids 13

1.3.1 Acoustic feedback in hearing aids 13

1.3.2 State-of-the-art feedback suppression techniques 14
1.4 Commercial hearing aids 17
1.4.1 Noise reduction 18
1.4.2 Feedback suppression 20
1.5 Outline of the thesis and main contributions 21
1.5.1 Design objectives 21
1.5.2 Chapter by chapter outline and contributions 22
1.6 Conclusion 25

2 Multi-microphone noise reduction techniques 27

2.1 Problem statement and notation 28
2.1.1 Signal model 28
2.1.2 Multi-microphone noise reduction 30
2.1.3 Equivalent representation in the frequency-domain . . . 31
2.1.4 Performance measures 36
2.1.5 Acoustical configuration for real-data experiments . . . 39
2.2 Fixed beamforming 41
2.2.1 Spatial directivity pattern and directivity index 42
2.2.2 Filter-and-sum beamforming 42
2.2.3 Delay-and-sum beamforming 43
2.2.4 Differential microphones 44
2.2.5 Robust fixed beamformers 46
2.3 Adaptive beamforming 48
2.3.1 Generalized Sidelobe Canceller 49
2.3.2 Minimum variance distortionless response beamformer . 52
2.3.3 Variants of the standard GSC with improved robustness 53
2.3.4 Adaptive differential microphone array 56
2.4 Multi-channel Wiener filtering 58

2.4.1 Multi-channel Wiener filter (with unknown reference) . 60
2.4.2 Implementation of the MWF 61
2.5 Subband MWF implementation 64
2.5.1 Improved intelligibility at lower computational cost . . . 64
2.5.2 Uniform subband implementation 66
2.5.3 Computational complexity 68
2.5.4 Experimental results 69
2.6 Conclusion 72
2.A Appendix: Adaptive differential microphone array 73

3 Robustness analysis of GSC and MWF for hearing aids 75

3.1 Preliminaries 76
3.1.1 Assumptions of the theoretical analysis 78
3.1.2 Experimental validation: configuration and implementation issues 78
3.2 Theoretical performance analysis 79
3.2.1 GSC with and without noise sensitivity constraint . . . 80
3.2.2 Multi-channel Wiener Filter (MWF) 83
3.3 Analysis of the noise sensitivity 85
3.3.1 Noise sensitivity of GSC and MWF 86
3.3.2 Influence of the noise scenario 87
3.3.3 Influence of the number of microphones 88
3.3.4 Experimental validation: number of microphones 91
3.4 Sensitivity to errors in the assumed signal model 93
3.4.1 Robustness to microphone mismatch 93
3.4.2 Robustness to deviations in microphone positions 95
3.4.3 Robustness to variations in speaker location 97
3.4.4 Experimental validation 98

3.5 Sensitivity to inaccurate estimation of the second order statistics 104
3.5.1 Estimation from finite amount of data 105
3.5.2 Non-stationarities in the noise scenario 106
3.6 Sensitivity to speech detection errors on the performance . . . 109
3.6.1 Multi-channel Wiener Filter 110
3.6.2 Generalized Sidelobe Canceller 111
3.6.3 Comparison between GSC, NSC-GSC and MWF 113
3.6.4 Experimental validation 113
3.7 Conclusions 116
3.A Appendix: Influence of the number of microphones 118

4 Spatially pre-processed speech distortion weighted MWF 119

4.1 Speech distortion weighted MWF 120
4.1.1 Trade-off speech distortion versus noise reduction . . . 120
4.1.2 Equivalent optimization criterion 121
4.1.3 Implementation of SDW-MWF 123
4.2 Spatially pre-processed SDW-MWF 124
4.2.1 Concept 124
4.2.2 Speech distortion regularized GSC 127
4.2.3 SP-SDW-MWF with filter $\bar{\mathbf{w}}_0$ on the speech reference . 128
4.3 Experimental validation 134
4.3.1 Implementation issues 134
4.3.2 SDR-GSC 135
4.3.3 SP-SDW-MWF with $\bar{\mathbf{w}}_0$ 136
4.3.4 Comparison with QIC-GSC 137
4.4 Conclusion 138
4.A Appendix: SP-SDW-MWF with filter $\bar{\mathbf{w}}_0[k]$ 140

5 Stochastic gradient implementation of the SP-SDW-MWF 143

5.1 Towards a stochastic gradient algorithm 144

5.1.1 Optimal stochastic gradient algorithm 145

5.1.2 Approximations of the optimal SG algorithm 146

5.2 Stochastic gradient algorithm 147

5.2.1 Time-domain stochastic gradient algorithm 148

5.2.2 Frequency-domain stochastic gradient algorithm 150

5.3 Improvements of the stochastic gradient algorithms 153

5.3.1 TD-SG algorithm with low pass filtering 154

5.3.2 FD-SG algorithm with low pass filtering 155

5.3.3 FD-SG algorithm using correlation matrices 159

5.4 Computational complexity 162

5.5 Experimental results 164

5.5.1 Implementation issues 165

5.5.2 Comparison of the FD-SG and the FD-SG algorithm with LPF 165

5.5.3 Comparison with FD-NLMS based SPA 170

5.6 Conclusions 172

5.A Appendix: Approximate stochastic gradient algorithm 174

5.A.1 TD-ASG algorithm 174

5.A.2 FD-ASG algorithm 175

5.A.3 Influence of the independence assumptions on the performance 177

6 Adaptive feedback cancellation techniques 181

6.1 Problem statement 182

6.1.1 Acoustic feedback in hearing aids 182

6.1.2 Feedforward suppression 184

6.1.3 Feedback cancellation 185
6.1.4 Performance measures for feedback cancellation 186
6.2 Standard adaptive feedback cancellation 187
6.2.1 Continuous versus non-continuous adaptation 187
6.2.2 Adaptive filtering procedures 189
6.2.3 Bias of the standard continuous adaptation feedback canceller 192
6.3 Solutions to reduce the bias 194
6.3.1 Reducing the correlation between $x[k]$ and $u[k]$ 194
6.3.2 Identification based on a probe signal $r[k]$ 196
6.3.3 Controlling the adaptation of the feedback canceller . . 196
6.4 Closed-loop identification of the feedback path 200
6.4.1 Closed-loop system set-up 201
6.4.2 Indirect and joint input-output method 202
6.4.3 Closed-loop identification with the direct method 203
6.4.4 Feedback cancellation with the direct method and a fixed estimate of the desired signal model 204
6.5 Conclusion 208
6.A Appendix: Continuous-adaptation feedback canceller 209

7 Adaptive feedback cancellation with linear prediction of the desired signal 211

7.1 Introduction 212
7.2 Conditions for identifiability 213
7.2.1 Case I: no noise injection 215
7.2.2 Case II: noise injection 218
7.2.3 Biased feedback path estimate in case of undermodelling 219
7.3 Recursive identification 220

7.3.1 Two-channel adaptive feedback canceller 221
7.3.2 PEM based adaptive feedback canceller 225
7.4 Simulation results 232
7.4.1 Set-up 234
7.4.2 Stationary desired signal $x[k]$ 235
7.4.3 Synthetic time-varying desired signal $x[k]$ 237
7.4.4 Real speech signal 238
7.5 Conclusions 243

8 Conclusions and further research **245**
8.1 Conclusions 245
8.1.1 Noise reduction 245
8.1.2 Feedback cancellation 247
8.2 Suggestions for further research 248
8.2.1 Noise reduction 248
8.2.2 Feedback cancellation 251
8.2.3 Combined solution for acoustic feedback and noise reduction 252

Bibliography **253**

List of publications **277**

Curriculum vitae **281**

Samenvatting

Adaptieve filtertechnieken voor het onderdrukken van ruis en akoestische terugkoppeling in hoorapparaten

Hoofdstuk 1: Inleiding

Gehoorverlies behoort tot de meest voorkomende ziekten en aandoeningen waaraan wereldwijd ongeveer 500 miljoen mensen lijden. De toenemende blootstelling in het dagelijkse leven aan te hoge geluidsniveaus alsook de veroudering van de bevolking zorgen ervoor dat dit cijfer in de toekomst verder zal stijgen. De meeste slechthorenden hebben een perceptueel gehoorverlies. Het grootste probleem bij deze slechthorenden is het verstaan van spraak in de aanwezigheid van omgevingslawaai. Om het verstaan van spraak in ruis te bevorderen, volstaat het niet om alleen het geluidssignaal te versterken, zoals gebruikelijk is in klassieke hoorapparaten, maar moet er bovendien een onderscheid gemaakt worden tussen het gewenste signaal en de ongewenste stoorsignalen.

Recente evoluties in de micro-elektronica maken het mogelijk om een DSP en twee (of meer) microfoons in een hoorapparaat in te bouwen. Uit gebrek aan efficiënte digitale signaalverwerkingstechnieken worden de extra mogelijkheden van digitale signaalverwerking echter nog niet optimaal benut, zodat huidige hoorapparaten vaak ontoereikend zijn om het gehoorverlies van de slechthorende te compenseren. Omgevingsruis wordt niet of onvoldoende onderdrukt.

In het dagelijks gebruik wordt de ruisonderdrukkingsperformantie van de huidige hoorapparaten bovendien vaak aangetast door allerlei imperfecties zoals afwijkingen in de microfoonkarakteristieken. Tenslotte kan het opgenomen geluidssignaal ook maar in geringe mate door een hoorapparaat versterkt worden tengevolge van akoestische terugkoppeling ('feedback'), afkomstig van de akoestische koppeling tussen de luidspreker en de microfoon van het hoorapparaat. De maximum toelaatbare versterking is vaak ontoereikend om het gehoorverlies te compenseren.

Er is bijgevolg een grote behoefte aan doeltreffende signaalverwerkingstechnieken voor het onderdrukken van ruis en akoestische terugkoppeling. In dit proefschrift ontwikkelen we dergelijke adaptieve signaalverwerkingstechnieken. Bij het ontwerp van ruisonderdrukkingsalgoritmes, streven we naar een goede ruisonderdrukking, een lage gevoeligheid aan imperfecties en een lage berekeningscomplexiteit. Op gebied van feedbackonderdrukking, ontwikkelen we algoritmes die een grotere onderdrukking van akoestische terugkoppeling en een betere geluidskwaliteit bieden dan bestaande technieken.

In **Paragraaf 1.1** beschrijven we de werking van het oor en de impact van gehoorverlies op het waarnemen van geluid. Het oor bestaat uit drie delen: het uitwendig oor, het middenoor en het binnenoor met de cochlea en gehoorzenuw (zie Figuur 1.1). Problemen in elk van de drie delen van het oor kunnen tot een gehoorverlies leiden. Het meest voorkomende type van gehoorverlies is *perceptief* gehoorverlies. Dit gehoorverlies is meestal te wijten aan de beschadiging van de haarcellen in de cochlea, i.e. cochleair gehoorverlies. De meest voorkomende oorzaken zijn veroudering en blootstelling aan te hoge geluidsniveaus. Cochleair gehoorverlies resulteert niet alleen in een verhoging van de gehoordrempel maar ook in een verminderd frequentie-oplossend vermogen en een verminderd dynamisch bereik. Ten gevolge van hun verminderd frequentie-oplossend vermogen ondervinden slechthorenden grotere moeilijkheden om spraak te verstaan in achtergrondlawaai dan normaalhorenden. De ernst van het gehoorverlies hangt af van de mate waarin de haarcellen beschadigd zijn. Mild tot ernstig gehoorverlies kan behandeld worden met een hoorapparaat. Bij zeer ernstig gehoorverlies en doofheid is men echter aangewezen op een cochleaire implantaat. Figuur 1.2 beschrijft de algemene werking van een hoorapparaat. De akoestische geluidssignalen worden opgenomen door één of meerdere microfoons. De geluidssignalen bestaan uit het gewenste spraaksignaal, de ruis alsook de signaallek van de luidspreker naar de microfoon(s). De opgenomen signalen worden verwerkt door middel van een signaalverwerkingsalgoritme met als doel het gehoorverlies van de slechthorende te compenseren. Het verwerkte signaal wordt vervolgens uitgestuurd door een luidspreker. De werking van een cochleair implantaat is weergegeven in Figuur 1.2. Een cochleair implantaat zet het geluidssignaal om in elektrische pulsen, die via een 'electrode-array' aangeboden worden aan de gehoorzenuw.

Paragraaf 1.2 definiëert welke signaalverwerking er in een hoorapparaat of een cochleair implantaat vereist is om het gehoorverlies te compenseren. Eén van de grootste problemen van slechthorenden is het verstaan van spraak in ruis. Om dit probleem op te lossen, is er –naast een versterking van het geluidssignaal– nood aan een verbetering van de signaal-tot-ruisverhouding en bijgevolg aan een *ruisonderdrukkingsalgoritme*. In het verleden werden hoorapparaten uitgerust met één enkele microfoon. Eénkanaals ruisonderdrukkingsalgoritmes kunnen echter enkel een onderscheid maken tussen signalen met verschillende temporele en spectrale eigenschappen. Ruissignalen hebben echter vaak een gelijkaardig spectrum als dat van spraak zodat de verbetering in spraakverstaanbaarheid met éénkanaals algoritmes beperkt is. Dankzij de miniaturisatie in de elektronica kunnen er meerdere microfoons in een hoorapparaat ingebouwd worden. Tegenwoordig zijn er reeds hoorapparaten met twee en zelfs drie microfoons commerciëel beschikbaar. Multi-microfoon systemen laten toe om –naast temporele en spectrale informatie– ook spatiale informatie te benutten. Om esthetisch aanvaardbaar te zijn, moet het multi-microfoon systeem echter zo klein mogelijk zijn. Bij kleinschalige systemen gaat ruisonderdrukking echter gepaard met een hoge gevoeligheid aan imperfecties zoals afwijkingen in de microfoonkarakteristieken ('microfoonmismatch') zodat een robuust algoritmisch ontwerp vereist is. Commerciële hoorapparaten bieden in het dagelijks gebruik vaak onvoldoende ruisonderdrukking omwille van een niet-robuust ontwerp.

Naast ruisonderdrukkingstechnieken worden er ook andere signaalverwerkingstechnieken toegepast in hoorapparaten om het gehoorverlies te compenseren. Voorbeelden zijn compressie en spectrale versterking.

Paragraaf 1.2 geeft een beknopt overzicht van de literatuur i.v.m. ruisonderdrukking. Eénkanaals technieken voor *ruisonderdrukking* kunnen ingedeeld worden in enerzijds parametrische technieken zoals Wiener- of Kalman-filtering en anderzijds niet-parametrische technieken zoals spectrale subtractie en deelruimtegebaseerde technieken. Multi-microfoon ruisonderdrukkingstechnieken kunnen ingedeeld worden in enerzijds vaste en adaptieve bundelvorming en anderzijds meer-kanaals Wiener-filtering. Multi-microfoon technieken worden in meer detail besproken in Hoofdstuk 2.

Naast de beperkte ruisonderdrukkingsperformantie vormt akoestische terugkoppeling een ernstig probleem voor hoorapparaatgebruikers. **Paragraaf 1.3** beschrijft het probleem van akoestische terugkoppeling in hoorapparaten. Omwille van akoestische terugkoppeling, kan het opgenomen geluidssignaal maar in geringe mate worden versterkt. Immers, als de versterking opgedreven wordt boven een bepaalde waarde, zal het hoorapparaat beginnen te fluiten en te oscilleren. De maximum toelaatbare versterking is vaak ontoereikend om het gehoorverlies te compenseren. Akoestische terugkoppeling is het gevolg van de akoestische koppeling tussen de luidspreker en de microfoon(s) van het hoorapparaat. Die akoestische koppeling, ook het akoestische feedbackpad genoemd,

is vaak afkomstig van de ontluchting in het oorstukje. Daarnaast wordt het akoestische feedbackpad bepaald door het oorkanaal en de akoestische omgeving. Veranderingen in de akoestische omgeving kunnen aanleiding geven tot grote variaties in het akoestische feedbackpad, zodat *adaptieve feedbackonderdrukkingstechnieken* aangewezen zijn. De miniaturisatie van hoorapparaten en de evolutie naar een open aanpassing van het hoorapparaat zal het risico op akoestische terugkoppeling in de toekomst nog verhogen, zodat er momenteel een grote vraag is naar doeltreffende feedbackonderdrukkingsalgoritmes.

Paragraaf 1.3 geeft een beknopt overzicht van de literatuur i.v.m. feedbackonderdrukking. Signaalverwerkingstechnieken voor *het aanpakken van akoestische terugkoppeling* kunnen ingedeeld worden in 'feedforward'-onderdrukkingstechnieken en feedbackonderdrukkingstechnieken. 'Feedforward'-onderdrukkingstechnieken wijzigen het signaalverwerkingspad van het hoorapparaat zodanig dat akoestische terugkoppeling wordt vermeden. Een gekend voorbeeld is 'notch'-filtering, waarbij een smalbandig frequentiegebied rond de kritische frequenties onderdrukt wordt aan de hand van één of meerdere 'notch'-filters. Een meer belovende techniek is *adaptieve feedbackonderdrukking* (zie Figuur 1.5). Hierbij wordt de ongewenste terugkoppeling tussen luidspreker en microfoon adaptief geïdentificeerd en gecompenseerd. De filteradaptatie wordt echter aanzienlijk bemoeilijkt door de (typische) autocorrelatie-structuur van de te verwerken audiosignalen, zodat er geen betrouwbare schatting van de terugkoppeling wordt verkregen en het gewenste spraaksignaal bijgevolg wordt vervormd. Voor dit probleem is in de literatuur nog geen goede oplossing beschikbaar. Recent werd aangetoond dat een goede schatting verkregen kan worden door middel van een 'filtered-X'-algoritme op voorwaarde dat het spectrum van het gewenste audiosignaal gekend is. In de praktijk is dit spectrum ongekend en bovendien tijdsvariërend. Bovendien blijkt de kwaliteit van de schatting sterk afhankelijk van de nauwkeurigheid van het veronderstelde signaalspectrum.

Paragraaf 1.4 beschrijft de technieken die in commerciële hoorapparaten worden gebruikt om ruis en akoestische terugkoppeling te onderdrukken. Tot voor 1996 waren er enkel analoge of hooguit analoog programmeerbare hoorapparaten beschikbaar. De signaalverwerking in deze hoorapparaten was hoofdzakelijk beperkt tot een (frequentie-afhankelijke) versterking en compressie. Met de komst van digitale hoorapparaten in 1996 werd het mogelijk om meer geavanceerde signaalverwerkingstechnieken zoals ruisonderdrukking en feedbackonderdrukking in te bouwen. Wegens de beperkte rekenkracht zijn de algoritmes die toegepast worden in huidige digitale hoorapparaten echter nog steeds vrij éénvoudig. Op het vlak van *ruisonderdrukking*, worden hoofdzakelijk vaste of adaptieve differentiële microfoons gebruikt. Differentiële microfoons zijn echter zeer gevoelig aan imperfecties zodat hun werkelijke performantie beperkt is. Om de negatieve effecten van *akoestische terugkoppeling* te reduceren, worden verzwakking van de hoge frequenties, 'notch'-filters of –in de meer geavanceerde hoorapparaten– adaptieve feedbackonderdrukkingstechnieken toegepast. Om

vervorming van het gewenste signaal te beperken, wordt de adaptatiesnelheid begrensd of worden er grenzen opgelegd aan het adaptieve filter ten koste van een slechtere 'tracking'-performantie.

Tenslotte wordt in **Paragraaf 1.5** een overzicht gegeven van de verschillende hoofdstukken en worden onze bijdragen toegelicht.

Hoofdstuk 2: Multi-microfoon ruisonderdrukkingstechnieken

Dit hoofdstuk behandelt de basisconcepten van multi-microfoon ruisonderdrukking alsook de voornaamste bestaande multi-microfoon ruisonderdrukkingsalgoritmes. De hoofdbedoeling van een ruisonderdrukkingsalgoritme voor hoorapparaten en cochleaire implantaten is het verbeteren van de spraakverstaanbaarheid. In dit hoofdstuk tonen we aan dat dit doel beter verkregen wordt door de ruisonderdrukkingsalgoritmes in subbanden te implementeren.

Paragraaf 2.1 beschrijft de notatie, het signaalmodel en het algemene concept van een multi-microfoon ruisonderdrukkingsalgoritme, zowel in het tijdsdomein als in het frequentiedomein. Daarnaast definiëren we de objectieve performantiematen en de configuratie van de experimenten die in dit proefschrift gehanteerd worden om de performantie van ruisonderdrukkingsalgoritmes te evalueren. Aangezien de hoofdbedoeling van een ruisonderdrukkingsalgoritme voor hoorapparaten en cochleaire implantaten de verbetering van de spraakverstaanbaarheid is, wordt de kwaliteit van de algoritmes bepaald aan de hand van spraakverstaanbaarheid-gewogen performantiematen. Wegens de kleine afstand tussen de microfoons in een hoorapparaat, moet niet alleen de verbetering in spraakverstaanbaarheid en de berekeningscomplexiteit beschouwd worden maar ook ook de robuustheid tegen imperfecties zoals afwijkingen in de microfoonkarakteristieken.

Multi-microfoon systemen kunnen ingedeeld worden in drie klasses, nl., vaste bundelvormers, adaptieve bundelvormers en meer-kanaals Wiener-filtertechnieken. **Paragraaf 2.2** behandelt *vaste bundelvormers*. Vaste bundelvormers zijn data-onafhankelijk en proberen ruimtelijk in te zoomen op de gewenste spraakbron. Voorbeelden zijn 'delay-and-sum'-bundelvormers, differentiële microfoons, superdirectieve bundelvormers en 'filter-and-sum'-bundelvormers. Omwille van hun eenvoud worden in hoorapparaten vaak eerste-orde (i.e., gebruikmakend van twee omnidirectionele microfoons) differentiële microfoons toegepast. Differentiële microfoons zijn echter zeer gevoelig aan imperfecties, vooral bij de lage frequenties. Die gevoeligheid neemt aanzienlijk toe met het aantal microfoons, zodat hun gebruik beperkt is tot een eerste-orde ontwerp of hooguit een tweede-orde ontwerp bij de hoge frequenties.

In tegenstelling tot vaste bundelvormers, maken adaptieve bundelvormers en meer-kanaals Wiener-filtertechnieken gebruik van data-afhankelijke filters die aangepast worden aan de (tijdsvariërende) omgeving, waardoor ze vaak resulteren in een betere ruisonderdrukking. **Paragraaf 2.3** bespreekt *adaptieve bundelvormers.* Meer-kanaals Wiener-filtering komt aan bod in Paragraaf 2.4. Een klassieke adaptieve bundelvormer is de 'Generalized Sidelobe Canceller' (GSC). De adaptieve differentiële microfoon die gebruikt wordt in de recentste digitale hoorapparaten is hiervan een speciaal geval. De GSC is opgebouwd uit een vaste bundelvormer die een spraakreferentie genereert, een 'blocking'-matrix die ruisreferenties genereert en een 'adaptive noise cancellation'-trap (ANC) die gebruik maakt van een meer-kanaals adaptief filter. De GSC veronderstelt de positie van de gewenste spraakbron vast en gekend, de microfoonconfiguratie gekend, de gebruikte microfoons perfect gekalibreerd en reverberatie afwezig. In de praktijk is zelden aan deze veronderstellingen voldaan zodat er signaallek optreedt in de ruisreferenties. Ten gevolge van die signaallek wordt het gewenste spraaksignaal vervormd. Verschillende varianten op de GSC worden voorgesteld die de spraakvervorming proberen te beperken. Vaak gebruikte technieken zijn het beperken van de adaptatie tijdens periodes van ruis en het begrenzen van het adaptief gedeelte, b.v. d.m.v. een 'norm-constraint'. De GSC met adaptatie tijdens ruis en de GSC met 'norm-constraint' worden gebruikt als referentie voor de ruisonderdrukkingstechnieken die in dit proefschrift worden ontwikkeld.

Een tweede categorie van adaptieve multi-microfoon ruisonderdrukkingstechnieken, die besproken worden in **Paragraaf 2.4**, zijn meer-kanaals Wiener-filtertechnieken. In dit proefschrift concentreren we ons op een meer-kanaals Wiener-filter (MWF) waarbij het referentiesignaal ongekend is. Dit MWF maakt een optimale schatting (in termen van MMSE) van de spraakcomponent in één van de microfoonsignalen. Hierbij wordt de som van ruis en spraakvervorming geminimaliseerd. In tegenstelling tot vaste en adaptieve bundelvormers en andere meer-kanaals Wiener-filtertechnieken vereist dit MWF noch a priori informatie over het gewenste signaalmodel noch kalibratie waardoor het een veelbelovende techniek lijkt wat betreft robuustheid tegen imperfecties. Tot nog toe zijn er echter enkel matrixgebaseerde implementaties van dit MWF beschikbaar zodat de berekeningscomplexiteit van het MWF te hoog is voor gebruik in hoorapparaten.

In **Paragraaf 2.5** ontwikkelen we een subbandgebaseerde implementatie van het MWF. Naast een significant lagere berekeningscomplexiteit t.o.v. de 'fullband'-gebaseerde implementaties biedt de subbandgebaseerde implementatie een verbetering in spraakverstaanbaarheid. Immers, spraakverstaanbaarheidgewogen performantiematen geven aan dat een minimalisatie van de ruis in afzonderlijke frequentiebanden of subbanden de spraakverstaanbaarheid meer optimaliseert dan een minimalisatie van de ruis over het volledige frequentiebereik. De significante reductie in berekeningscomplexiteit brengt een im-

plementatie van het MWF in een commercieel digitaal hoorapparaat binnen handbereik.

Hoofdstuk 3: Analyse van de robuustheid van GSC en MWF voor toepassing in hoorapparaten

Bij kleinschalige multi-microfoon systemen zoals deze in hoorapparaten gaat ruisonderdrukking gepaard met een hoge gevoeligheid aan imperfecties of modelleerfouten zoals afwijkingen in de microfoonkarakteristieken. In dit hoofdstuk, evalueren we zowel analytisch als experimenteel de gevoeligheid van de GSC, de GSC met 'norm-constraint' en het MWF voor verschillende modelleerfouten. In de analyse concentreren we ons op kleinschalige microfoonsystemen zoals gebruikt in hoorapparaten.

In **Paragraaf 3.2** leiden we formules af die de optimaal haalbare performantie (i.e. met oneindige filterlengtes) van de verschillende algoritmes voorspellen, gegeven een bepaalde microfoonconfiguratie en een bepaald ruisscenario. We tonen aan dat –onder zwakke voorwaarden– het MWF overeenkomt met een serieschakeling van een optimale 'Minimum Variance Distortionless Response'-bundelvormer die een schatting maakt van de spraakcomponent in één van de microfoons, en een éénkanaals Wiener-filter.

De gevoeligheid van een ruisonderdrukkingsalgoritme voor fouten in de veronderstellingen over het gewenste signaalmodel wordt in de literatuur gekwantificeerd aan de hand van de ruisgevoeligheid: hoe hoger de ruisgevoeligheid, hoe gevoeliger het algoritme voor fouten in de veronderstellingen. In **Paragraaf 3.3**, toont een theoretische analyse van de ruisgevoeligheid aan dat een robuust ontwerp vooral cruciaal is in complexere ruissituaties zoals b.v. meerdere ruisbronnen en bij een lage signaal-tot-ruisverhouding (SNR). De gevoeligheid aan modelleerfouten in het gewenste signaalmodel neemt ook toe met een toenemend aantal microfoons per oppervlakte. In dit proefschrift tonen we aan dat –ten gevolge van de hoge gevoeligheid– het aantal microfoons dat efficiënt benut kan worden in een achter-het-oor hoorapparaat beperkt is tot drie op voorwaarde dat het ruisonderdrukkingsalgoritme voldoende robuust is. Een extra derde microfoon biedt vooral een groot voordeel bij complexere ruisscenario's zoals meerdere ruisbronnen en in omgevingen met veel reverberatie.

Paragraaf 3.4 bestudeert de gevoeligheid van de GSC en het MWF voor verschillende soorten modelleerfouten in het gewenste signaalmodel, i.e. afwijkingen in de microfoonkarakteristieken, de microfoonposities en de positie van de gewenste spreker. Onder de verschillende modelleerfouten, zijn voornamelijk afwijkingen in de microfoonkarakteristieken schadelijk voor de performantie van de GSC, zelfs als de adaptatie van de ANC-trap beperkt wordt tot ruis-

periodes. Omwille van de hoge gevoeligheid van de GSC, is een begrenzing van het adaptief gedeelte, b.v. d.m.v. een 'norm-constraint', noodzakelijk. De performantie van het MWF wordt daarentegen niet aangetast door modelleerfouten. In vergelijking tot de GSC met 'norm-constraint' onderdrukt het MWF meer ruis en kan het MWF het voordeel van een extra derde microfoon in een hoorapparaat ten volle benutten.

Zowel de GSC als het MWF maken gebruik van schattingen van de tweede-orde statistieken van de ruis en/of de spraak. Aangezien het MWF geen voorkennis vereist over het gewenste signaalmodel, wordt verwacht dat het MWF gevoeliger is voor een onnauwkeurige schatting van de tweede-orde statistieken. Die onnauwkeurige schatting kan het gevolg zijn van zowel niet-stationariteiten in de ruis als van detectiefouten van het gebruikte spraakdetectie-mechanisme. In **Paragraaf 3.5** tonen we aan dat het MWF zijn voordeel t.o.v. de GSC en de GSC met 'norm-constraint' behoudt op voorwaarde dat de spatiale en de lange-termijn gemiddelde spectrale eigenschappen van de ruis traag variëren in de tijd. **Paragraaf 3.6** onderzoekt het effect van een foutieve spraakdetectie op de performantie van de GSC en het MWF. We tonen aan dat –in tegenstelling tot de GSC en de GSC met 'norm-constraint'– de spraakverstaanbaarheidgewogen SNR-verbetering van het MWF niet aangetast wordt door spraakdetectiefouten, noch wanneer spraak foutief als ruis gedetecteerd wordt, noch wanneer ruis foutief als spraak gedetecteerd wordt. Wanneer spraak foutief als ruis gedetecteerd wordt, zal de spraakvervorming echter wel sterk toenemen met het percentage foutief gedetecteerde samples. Als het percentage spraakdetectiefouten lager is dan 20%, blijft de spraakvervorming door het MWF echter beperkt, terwijl er meer ruis wordt onderdrukt dan door de GSC met 'norm-constraint'.

De theoretische analyse in dit hoofdstuk is gebaseerd op een aantal veronderstellingen zoals b.v. oneindige filterlengtes en de afwezigheid van reverberatie. Experimenten met reële data verkregen met een achter-het-oor hoorapparaat bevestigen de theoretische resultaten. De in dit hoofdstuk ontwikkelde theoretische analyse is bijgevolg een handige methode om het effect van een specifieke modelleerfout op de performantie van een ruisonderdrukkingsalgoritme te voorspellen.

Hoofdstuk 4: Spraakvervormingsgewogen MWF met spatiale 'pre-processing' (SP-SDW-MWF)

In dit hoofdstuk tonen we aan dat de GSC en het MWF geïntegreerd kunnen worden in één globaal ruisonderdrukkingsschema, het spraakvervormingsgewogen MWF met spatiale 'pre-processing' (SP-SDW-MWF) genoemd. Afhankelijk van een aantal parameterinstellingen bekomt met één van beide algoritmes

of een interessante tussenoplossing. Het globale schema brengt –in tegenstelling tot de GSC– spraakvervorming in rekening en reduceert meer ruis dan de GSC met 'norm-constraint' voor eenzelfde maximale hoeveelheid spraakvervorming.

In **Paragraaf 4.1** veralgemenen we het ontwerpcriterium van het MWF zodat het mogelijk is om spraakvervorming en ruisonderdrukking tegenover elkaar af te wegen. Het veralgemeende MWF wordt spraakvervormingsgewogen MWF (SDW-MWF) genoemd. Bovendien formuleren we een equivalent ontwerpcriterium voor het SDW-MWF dat meer gelijkenis vertoont met het criterium van de ANC-trap in de GSC. Tenslotte tonen we aan dat de matrixgebaseerde recursieve implementaties van het MWF uitgebreid kunnen worden zodat ze toepasbaar zijn op het SDW-MWF.

In **Paragraaf 4.2** integreren we de GSC en het MWF in één globaal ruisonderdrukkingsschema, het SP-SDW-MWF genoemd. Het SP-SDW-MWF, dat beschreven is in Figuur 4.2, is opgebouwd uit een vaste bundelvormer die een spraakreferentie genereert, een 'blocking'-matrix die ruisreferenties genereert en een SDW-MWF. Het globale schema laat toe de verschillen tussen de GSC en het MWF beter te begrijpen. Afhankelijk van een aantal parameterinstellingen, nl. de instelling van een parameter die afweegt tussen spraakvervorming en ruisonderdrukking en de aan- of afwezigheid van een filter op de spraakreferentie, bekomt met één van beide algoritmes. Bovendien biedt het globale schema ook nieuwe tussenoplossingen, zoals de 'Speech Distortion Regularized GSC' (SDR-GSC).

- Zonder filter op de spraakreferentie verkrijgen we de *'Speech Distortion Regularized GSC'*. De SDR-GSC voegt robuustheid toe aan de GSC door spraakvervorming in rekening te brengen in het ontwerpcriterium van het adaptief gedeelte. Afhankelijk van de instelling van de 'tradeoff'-parameter wordt meer of minder aandacht besteed aan spraakvervorming. De GSC is een speciaal geval van de SDR-GSC waarbij geen rekening wordt gehouden met spraakvervorming. De SDR-GSC is dus een alternatieve methode om de gevoeligheid van de GSC voor modelleerfouten te verminderen. In vergelijking tot de bestaande GSC met 'norm-constraint', behaalt de SDR-GSC een betere performantie bij kleine modelleerfouten, terwijl het dezelfde robuustheid garandeert bij grote modelleerfouten.

- Aangezien het SP-SDW-MWF spraakvervorming expliciet in rekening brengt in zijn ontwerpcriterium, kan er een filteroperatie op de spraakreferentie toegevoegd worden. We tonen mathematisch aan dat *het SP-SDW-MWF met extra filter* in de afwezigheid van modelleerfouten overeenkomt met een serieschakeling van een GSC (of een SDR-GSC) en een spraakvervormingsgewogen één-kanaals Wiener-filter. In de aanwezigheid van modelleerfouten compenseert de extra filteroperatie op de spraakreferentie het performantieverlies van de SDR-GSC ten gevolge van model-

leerfouten. In tegenstelling tot de SDR-GSC en de GSC verslechtert de performantie niet ten gevolge van afwijkingen in de microfoonkarakteristieken.

Paragraaf 4.3 illustreert de theoretische resultaten aan de hand van experimenten met een achter-het-oor hoorapparaat. Het SP-SDW-MWF verhoogt inderdaad de robuustheid van de GSC tegen modelleerfouten. In vergelijking met de GSC met 'norm-constraint' onderdrukt het SP-SDW-MWF meer ruis voor eenzelfde maximale hoeveelheid spraakvervorming.

Hoofdstuk 5: Stochastische-gradiënt-implementatie van het SP-SDW-MWF

De recursieve matrixgebaseerde implementaties voor het MWF kunnen toegepast worden op het SP-SDW-MWF. In tegenstelling tot de GSC met 'norm-constraint', zijn er echter nog geen goedkope stochastische-gradiënt-algoritmes voor het SP-SDW-MWF beschikbaar. Alhoewel een subbandgebaseerde implementatie reeds een aanzienlijke complexiteitsreductie met zich mee brengt, lijkt omwille van de beperkte rekencapaciteit in huidige hoorapparaten alsook omwille van het vermogenverbruik verder onderzoek naar goedkopere realisaties nuttig. In dit hoofdstuk worden een aantal stochastische-gradiënt-implementaties van het SP-SDW-MWF ontwikkeld, zowel in het frequentie- als in het tijdsdomein.

Op basis van het ontwerpcriterium van het SP-SDW-MWF leiden we in **Paragraaf 5.1** een stochastische-gradiënt-implementatie af. Dit algoritme vereist echter een instantane schatting van de spraakcorrelatiematrix. Aangezien het zuiver spraaksignaal in de praktijk niet beschikbaar is, kan die schatting niet gemaakt worden. Bijgevolg moeten er benaderingen worden ingevoerd. In Paragraaf 5.1 geven we een beknopte samenvatting van de mogelijke benaderingen. De voorgestelde oplossingen worden uitvoerig besproken in Paragraaf 5.2 en Paragraaf 5.3.

In **Paragraaf 5.2** stellen we voor om de instantane schatting van de spraakcorrelatiematrix te benaderen aan de hand van instantane schattingen van de correlatiematrix van de spraak + ruis en de correlatiematrix van de ruis. De correlatiematrix van de spraak + ruis wordt op een ander tijdstip geschat dan de correlatiematrix van de ruis. De berekening van de schattingen gebeurt aan de hand van databuffers. Om de convergentiesnelheid te verbeteren en de berekeningscomplexiteit te verminderen, voeren we het stochastische-gradiënt-algoritme uit in het frequentiedomein. In het geval van niet-stationaire ruis resulteert de niet-continue instantane schatting van de correlatiematrix van de

spraak + ruis en de ruis echter in een onnauwkeurige schatting van de spraakcorrelatiematrix wat aanleiding geeft tot een grote residuele 'excess'-fout.

In de praktijk zijn het korte-termijn gemiddelde spectrum van de spraak en de ruis vaak sterk niet-stationair terwijl de spatiale eigenschappen (b.v. positie van de ruisbron) en het lange-termijn gemiddelde spectrum vaak trager variëren in de tijd. In dit geval volstaat het om een schatting van de lange-termijn spraakcorrelatiematrix te gebruiken in het stochastische-gradiënt-algoritme. Een vrij nauwkeurige schatting van de lange-termijn spraakcorrelatiematrix kan dan verkregen worden door de instantane schattingen van de correlatiematrix van de ruis en de spraak + ruis in de tijd uit te middelen. Dit vereist echter rekenintensieve matrixbewerkingen. In **Paragraaf 5.3** stellen we twee oplossingen voor om rekenintensieve matixbewerkingen te vermijden.

Een eerste methode bestaat erin om de term die spraakvervorming in rekening brengt in het stochastische-gradiënt-algoritme (ook regularisatieterm genoemd) te benaderen met behulp van een laagdoorlaatfilter. Het laagdoorlaatfilter vermindert de fluctuaties in de filtercoefficiënten die afkomstig zijn van een onnauwkeurige instantane schatting van de spraakcorrelatiematrix terwijl het nog steeds een goede 'tracking' biedt van een verandering in ruisscenario. Het algoritme met laagdoorlaatfilter vereist echter nog steeds opslag van data in databuffers en bijgevolg te veel geheugenruimte voor hoorapparaten. In het geval van een frequentiedomein-implementatie kan ook een alternatieve benadering gemaakt worden die minder geheugen vergt. In deze methode wordt de regularisatie-term benaderd met behulp van diagonale frequentiedomein-correlatiematrices i.p.v. met data-vectors uit databuffers.

In **Paragraaf 5.4** en **Paragraaf 5.5** vergelijken we de berekeningscomplexiteit en de performantie van de ontwikkelde stochastische-gradënt-implementaties van het SP-SDW-MWF met een NLMS-gebaseerde implementatie van de GSC met 'norm-constraint'. Experimenten met een drie-microfoon hoorapparaat tonen aan dat het laagdoorlaatfilter de performantie van het stochastische-gradiënt-algoritme in spectraal niet-stationaire ruis aanzienlijk verbetert. Uit initiële experimenten blijkt dat dit ook geldt voor het frequentiedomein-stochastische-gradiënt-algoritme dat gebruik maakt van diagonale correlatiematrices. Bovendien behouden de voorgestelde stochastische-gradiënt-algoritmes het performantie-voordeel van het SP-SDW-MWF t.o.v. de GSC met 'norm-constraint', terwijl de berekeningscomplexiteit gelijkaardig is als die van de NLMS-gebaseerde implementatie van de GSC met 'norm-constraint'. De beperkte berekeningscomplexiteit, de minimale geheugenvereiste alsook het performantie-voordeel t.o.v. GSC met 'norm-constraint' maken het stochastisch-gradënt-algoritme met correlatiematrices een ideale kandidaat voor implementatie in toekomstige hoorapparaten.

Hoofdstuk 6: Adaptieve feedbackonderdrukkingstechnieken

Akoestische terugkoppeling ('feedback') beperkt de maximale versterking die in een hoorapparaat gebruikt kan worden. Die maximum toelaatbare versterking is vaak onvoldoende om het gehoorverlies van de slechthorende te compenseren, zodat er nood is aan feedbackonderdrukkingstechnieken. Dit hoofdstuk geeft een overzicht van de bestaande technieken voor het onderdrukken van akoestische terugkoppeling.

Paragraaf 6.1 geeft een wiskundige formulering van akoestische terugkoppeling in hoorapparaten en een overzicht van de mogelijke oplossingsmethodes. Ten gevolge van de akoestische koppeling tussen de luidspreker en de microfoon van het hoorapparaat wordt –naast het geluidssignaal uit de omgeving– ook het uitgestuurde luidsprekersignaal opgenomen door de microfoon, wat resulteert in een gesloten-lus-systeem. Bij een te hoge versterking van het hoorapparaat wordt het gesloten-lus-systeem onstabiel. In de literatuur zijn er twee mogelijke oplossingsmethodes voorgesteld om de negatieve effecten van akoestische terugkoppeling te reduceren, i.e. 'feedforward'-onderdrukkingstechnieken en feedbackonderdrukkingstechnieken. *'Feedforward'-onderdrukkingstechnieken* wijzigen het signaalverwerkingspad van het hoorapparaat zodanig dat het gesloten-lus-systeem stabiel is. Voorbeelden zijn het verzwakken van de kritische frequenties met behulp van een 'notch'-filter, het toevoegen van tijdsvariërende elementen aan het signaalverwerkingspad en het egaliseren van de fase van de open-lus-responsie. De toename in versterking die met behulp van 'feedforward'-onderdrukkingstechnieken bekomen wordt, is echter beperkt. Bovendien wordt de geluidskwaliteit vaak aangetast. Een meer belovende techniek is *feedback-onderdrukking*. Hierbij wordt de ongewenste terugkoppeling tussen luidspreker en microfoon geïdentificeerd en gecompenseerd. Het akoestische feedbackpad tussen de luidspreker en de microfoon kan aanzienlijk variëren in de tijd zodat adaptieve feedbackonderdrukkingstechnieken aangewezen zijn. De performantie van een feedbackonderdrukkingsalgoritme wordt geëvalueerd aan de hand van de afwijking tussen het werkelijke en het geïdentificeerde feedbackpad en de toename in de maximum toelaatbare versterking.

Paragraaf 6.2 en **Paragraaf 6.3** geven een gedetailleerde bespreking van de standaard adaptieve feedbackonderdrukkingstechnieken. We tonen aan dat –omwille van de aanwezigheid van een gesloten-lus– de standaard adaptieve feedbackonderdrukkingstechnieken geen betrouwbare schatting van het akoestische feedbackpad kunnen maken wanneer het gewenste signaal spectraal gekleurd is zoals b.v. spraak. De modelleerfout of bias in het geschatte feedbackpad resulteert in vervorming van het gewenste geluidssignaal. In de literatuur zijn er verschillende oplossingen voorgesteld om de modelleerfout of de bias te verminderen. Klassieke technieken zijn het verminderen van de correlatie tussen de ingang van het adaptieve filter en het gewenste geluidssignaal door het

toevoegen van een vertraging of niet-lineariteit aan het signaalverwerkingspad, het toevoegen van een extern ruissignaal aan het luidsprekersignaal of het controleren van de adaptatie van het adaptief filter. Bij elk van deze technieken is er een afweging tussen de modelleerfout of bias enerzijds en de geluidskwaliteit en 'tracking'-performantie anderzijds, zodat er typisch een aanzienlijke bias overblijft.

Een aantrekkelijkere oplossing om de bias te reduceren, bestaat erin het akoestische feedbackpad te beschouwen als het voorwaartse pad van een gesloten-lus-systeem en de systeemidentificatie-methodes voor gesloten-lus-systemen toe te passen. **Paragraaf 6.4** bespreekt de drie bestaande methodes voor gesloten-lus-systeemidentificatie in de context van feedbackonderdrukking, nl. de indirecte methode, de directe methode en de 'joint-input-output'-methode. Onder de drie methodes, is vooral de directe methode interessant voor feedbackonderdrukking. In tegenstelling tot de andere technieken, vergt de directe methode noch voorkennis over het signaalverwerkingspad van het hoorapparaat noch het gebruik van een extern ruissignaal. Recent werd een adaptieve feedonderdrukkingstechniek op basis van de directe methode voorgesteld die gebruik maakt van het model van het gewenste geluidssignaal. Er werd aangetoond dat een goede schatting van het feedbackpad verkregen kan worden door middel van een 'filtered-X'-algoritme op voorwaarde dat het model van het audiosignaal gekend is. In de praktijk is dit signaalmodel ongekend en bovendien tijdsvariërend. Bovendien blijkt de kwaliteit van de schatting sterk afhankelijk van de nauwkeurigheid van het veronderstelde signaalmodel. Bijgevolg wensen we een adaptieve schatting te maken van het signaalmodel. In het volgende hoofdstuk ontwikkelen we adaptieve feedbackonderdrukkingstechnieken die gebaseerd zijn op identificatie van zowel het feedbackpad als het gewenste signaalmodel met als doel ook voor ongekende en tijdsvariërende gekleurde signalen een betrouwbare schatting van het feedbackpad te verkrijgen.

Hoofdstuk 7: Adaptieve feedbackonderdrukkingstechnieken met lineaire predictie van het gewenste signaal

In dit hoofdstuk ontwikkelen we adaptieve feedbackonderdrukkingsalgoritmes die –ook voor spectraal gekleurde en tijdsvariërende audiosignalen– een betrouwbare schatting van het akoestische feedbackpad opleveren. De ontwikkelde algoritmes zijn gebaseerd op een gesloten-lus-identificatie van zowel het akoestisch feedbackpad als het model van het gewenste audiosignaal.

In het algemeen blijkt het niet mogelijk om in een gesloten-lus-systeem het feedbackpad en het model van het gewenste audiosignaal tegelijk te schatten. In **Paragraaf 7.2** tonen we aan dat onder bepaalde realistische voorwaarden

identificatie van beide modellen toch mogelijk is. Zo volstaat het b.v. om een korte vertragingstijd in het signaalverwerkingspad van het hoorapparaat in te lassen. Andere mogelijkheden zijn het niet-lineair of tijdsvariant maken van het signaalverwerkingspad of het injecteren van een ruissignaal aan de ingang van de luidspreker, zolang hierdoor de kwaliteit van het gewenste signaal niet aangetast wordt. In de afleiding van de voorwaarden voor identificeerbaarheid veronderstellen we dat het gewenste signaal gegenereerd is door een 'all-pole'-filter te exciteren met witte ruis. We tonen aan dat de identificatie van het akoestische feedbackpad bemoeilijkt wordt wanneer het excitatiesignaal periodisch is, zoals b.v. bij stemhebbende klanken of bij periodische audiosignalen.

In **Paragraaf 7.3** leiden we twee klasses van adaptieve filtertechnieken af om het feedbackpad en het signaalmodel te identificeren, namelijk een twee-kanaals-identificatie-gebaseerde procedure en een predictie-fout-methode. Aangezien de meeste audiosignalen benaderd kunnen worden als een auto-regressief random proces, veronderstellen we in de identificatieprocedures dat het inverse-signaalmodel gemodelleerd kan worden met een FIR filter. De *twee-kanaals-identificatie-gebaseerde procedure* schat enerzijds het inverse-signaalmodel en anderzijds de convolutie van het feedbackpad met dit inverse-signaalmodel, wat resulteert in een lineair ontwerpcriterium. Een schatting van het feedbackpad wordt verkregen door convolutie van het laatstgenoemde filter met het inverse van het eerstgenoemde filter. De *predictie-fout-methode* maakt een rechtstreekse schatting van het inverse-signaalmodel en het akoestische feedbackpad en laat bijgevolg toe om een andere adaptatiesnelheid of een ander adaptatie-algoritme te gebruiken bij de schatting van het inverse-signaalmodel en de schatting van het feedbackpad. Dit is wenselijk aangezien het audiosignaal in de praktijk vaak sterk tijdsvariërend is (b.v. spraak), terwijl het akoestische feedbackpad trager variëert in de tijd. De schatting van het feedbackpad gebeurt met behulp van een 'filtered-X'-algoritme, dat ook geïnterpreteerd kan worden als een 'instrumental variable'-methode met filtering van de ingangsdata. Voor de schatting van het inverse-spraakmodel wordt een lineaire predictiemethode uit de spraakcodering gebruikt, aangezien deze methode convergentie naar een stabiel spraakmodel garandeert.

Paragraaf 7.4 vergelijkt de performantie van de twee-kanaals-identificatie-gebaseerde procedure en de predictie-fout-methode met de standaard adaptieve feedbackonderdrukkingstechniek en het 'filtered-X'-algoritme dat gebruik maakt van een vaste schatting van het lange-termijn gemiddelde signaalmodel. In de simulaties concentreren we ons vooral op de invloed van een vertraging in het signaalverwerkingspad op de performantie van de verschillende algoritmes. Simulaties tonen aan dat wanneer de voorwaarden vervuld zijn, de voorgestelde adaptieve schema's de performantie van de standaard adaptieve feedbackonderdrukkingstechniek overtreffen. Voor sterk tijdsvariërende audiosignalen zoals reële spraak levert de predictie-fout-methode een betere schatting op dan de twee-kanaals-identificatie-gebaseerde procedure en het 'filtered-X'-

algoritme dat gebruik maakt van een vast signaalmodel. Bovendien verbetert de performantie van de predictie-fout-methode aanzienlijk bij het toevoegen van een extern laagvermogen ruissignaal, terwijl het voordeel van een laagvermogen ruissignaal voor de standaard adaptieve feedbackonderdrukkingstechniek en het 'filtered-X'-algoritme dat een vaste schatting van het lange-termijn gemiddelde signaalmodel gebruikt, verwaarloosbaar is.

Hoofdstuk 8: Besluit en suggesties voor verder onderzoek

Paragraaf 8.1 geeft een samenvatting van de resultaten en **Paragraaf 8.2** formuleert een aantal suggesties voor verder onderzoek.

Verbeteringen aan het SP-SDW-MWF

In Hoofdstuk 4 werd een veralgemeend ruisonderdrukkingsschema voorgesteld, nl. het spraakvervormingsgewogen MWF met spatiale 'pre-processing' (SP-SDW-MWF). Aan dit schema kunnen nog een aantal verbeteringen aangebracht worden.

Het SP-SDW-MWF vereist een schatting van de correlatiematrix van de spraak in de berekening van de regularisatieterm. De nauwkeurigheid van die schatting is bepalend voor de performantie van het SP-SDW-MWF. In Hoofdstuk 3 bemerkten we dat die schatting bemoeilijkt wordt door niet-stationariteiten in de ruis alsook door fouten in het spraakdetectie-algoritme. Verder onderzoek kan verricht worden naar het verbeteren van de nauwkeurigheid van de geschatte correlatiematrix. Op het vlak van éénkanaals ruisonderdrukking werd een minimum-statistiek-methode voorgesteld die het mogelijk maakt om het vermogenspectrum van de ruis ook te schatten tijdens periodes van spraak en bijgevolg een nauwkeurigere schatting oplevert van het vermogenspectrum van de spraak. Het zou interessant zijn om die minimum-statistiek-methode uit te breiden naar een multi-microfoon procedure, waarbij niet alleen het vermogenspectrum maar ook het kruisvermogenspectrum tussen de verschillende microfoonsignalen berekend wordt.

Bij zeer lage signaal-tot-ruisverhouding (SNR) kan er geen betrouwbare schatting van de correlatiematrix van de spraak gemaakt worden, zodat de adaptieve filters naar een ongewenste oplossing kunnen convergeren. In die situaties van lage SNR is het gebruik van een vaste bundelvormer aangewezen. Het zou bijgevolg interessant zijn om een een controle-algoritme in het ruisonderdrukkingsschema in te bouwen die deze overgang naar een vaste bundelvormer tot stand brengt. Het SP-SDW-MWF leent zich hier gemakkelijk toe. Door de adaptieve filters uit te schakelen wordt de uitgang van de vaste bundelvormer in de spatiale 'pre-processor' bekomen. Een elegantere methode wordt verkre-

gen door een 'norm-constraint' op te leggen op de adaptieve filters of door in het ontwerpcriterium meer aandacht te besteden aan spraakvervorming. Dit laat een beperkte adaptatie van het adaptief filter toe zolang het resulterend schema niet te ver afwijkt van de vaste bundelvormer.

De spatiale 'pre-processing'-stap in het SP-SDW-MWF kan verder geoptimaliseerd worden. Als vaste bundelvormer werd in dit proefschrift b.v. een 'delay-and-sum'-bundelvormer gebruikt. Deze bundelvormer is –in het geval van kleinschalige systemen– zeer robuust, maar de hoeveelheid ruisonderdrukking is beperkt. Verder onderzoek gaat uit naar meer geoptimaliseerde vaste bundelvormers, die meer ruis onderdrukken en toch nog voldoende robuust zijn. Binnen de onderzoeksgroep werden recent een aantal wiskundige technieken ontwikkeld die, gegeven statistische informatie over de mogelijke modelleerfouten, een robuuste vaste bundelvormer ontwerpen. Deze technieken kunnen toegepast worden bij het ontwerp van de vaste bundelvormer en de 'blocking'-matrix in het SP-SDW-MWF. Een eventuele aanpassing is noodzakelijk om het hoofdeffect in rekening te brengen.

Validatie aan de hand van perceptieve testen

Tot nu toe werden de bestudeerde en ontwikkelde algoritmes geëvalueerd aan de hand van objectieve spraakverstaanbaarheid-gewogen performantiematen. Om het werkelijk voordeel van de ontwikkelde algoritmes voor de slechthorende te kunnen inschatten, is een validatie aan de hand van perceptieve testen met normaal- en slechthorenden noodzakelijk. Uit de perceptieve testen zal ook blijken of de aanwezigheid van een prominent eigen-stemsignaal de werking van het algoritme verstoort.

Perceptueel gemotiveerde ruisonderdrukking

In Hoofdstuk 2 hebben we aangetoond dat een subbandgebaseerde implementatie van het MWF betere spraakverstaanbaarheid-gewogen performantiematen oplevert dan een 'fullband'-gebaseerde implementatie. Het MMSE-criterium waarop het MWF gebaseerd is, is echter vanuit een psycho-akoestisch oogpunt niet zo relevant. Op het vlak van éénkanaals ruisonderdrukkingstechnieken werden er reeds spectrale wegingsfuncties voorgesteld die gebaseerd zijn op modellen van maskering. Het zou interessant zijn om te onderzoeken of modellen van maskering of modellen van het gehoor ook succesvol toegepast kunnen worden bij het ontwerp van multi-microfoon ruisonderdrukkingssystemen.

Binaurale ruisonderdrukking

In dit proefschrift hebben we ons geconcentreerd op monaurale microfoonsystemen. Met de komst van een draadloze link tussen linker- en rechteroor worden ook binaurale systemen mogelijk in commerciële hoorapparaten. Het zou daarom interessant zijn om te onderzoeken in welke mate de in dit proefschrift

ontwikkelde technieken ook voordeel bieden voor binaurale systemen. Naast het onderdrukken van ruis proberen binaurale technieken de natuurlijke cues voor localisatie (b.v. de interaurale tijdsvertraging en de interaurale niveauverschillen) te bewaren om het richtinghoren te verbeteren. Het MWF schat het spraaksignaal zoals het ontvangen wordt aan de microfoon, wat gunstig lijkt voor binauraal horen. Vaak is er echter een afweging tussen het onderdrukken van ruis enerzijds en het bewaren van de localisatie-cues anderzijds zodat bijkomende voorwaarden op het MWF nodig kunnen zijn.

Feedbackonderdrukking

In Hoofdstuk 7 werden adaptieve-filtertechnieken ontwikkeld die –onder bepaalde voorwaarden– een betrouwbare schatting van het feedbackpad mogelijk maken. In dit proefschrift werd vooral het inlassen van een korte vertragingstijd onderzocht, omdat die methode het gewenste signaal zeker niet aantast. Het injecteren van een extern ruissignaal of het toevoegen van niet-lineariteiten kunnen echter een nuttig alternatief zijn. Uit het proefschrift blijkt b.v. dat het feedbackpad niet bepaald kan worden in geval van periodische signalen, zoals b.v. een perfecte sinus of een stemhebbende klank. Initiële experimenten tonen aan dat het injecteren van ruis de performantie voor een reëel spraaksignaal aanzienlijk verbetert. Bovendien laat de aanwezigheid van een extra ruissignaal het gebruik van alternatieve gesloten-lus-identificatieprocedures voor feedbackonderdrukking toe. Het is bijgevolg aan te bevelen om in verder onderzoek de potentiële voordelen van het injecteren van ruis of het toevoegen van niet-lineariteiten te bestuderen. Een belangrijke vereiste bij het gebruik van deze technieken is het behoud van de geluidskwaliteit. Additieve ruis is niet hoorbaar zolang het niveau lager is dan de gehoordrempel van de slechthorende of zolang de ruis gemaskeerd wordt door het geluidssignaal. Het zou daarom interessant zijn om de bestaande modellen voor luidheid en maskering te vertalen naar een rekenkundig efficiënt model voor de generatie van minimaal hoorbare signalen en niet-lineariteiten.

In Hoofdstuk 7 werden twee verschillende adaptieve feedbackonderdrukkingsalgoritmes voorgesteld, nl. een twee-kanaals-identificatie-gebaseerde procedure en een predictie-fout-methode. Bijkomend onderzoek kan verricht worden naar het ontwikkelen van alternatieve identificatieprocedures. Zo kan b.v. het gebruik van een tijdsvariërend autoregressief model of een Kalman filtering gebaseerde techniek de identificatie van het zeer snel tijdsvariërende spraakmodel en bijgevolg de kwaliteit van het geschatte feedbackpad verbeteren.

Gecombineerde oplossing voor het onderdrukken van ruis en akoestische terugkoppeling

Tenslotte suggereren we om de ontwikkelde feedbackonderdrukkings- en ruisonderdrukkingsalgoritmes te combineren in één signaalverwerkingsschema (cf.

de combinatie van ruisonderdrukking met het onderdrukken van akoestische echo). Een voor de hand liggende oplossing bestaat erin om het ruisonderdrukkingsalgoritme en het feedbackonderdrukkingsalgoritme onafhankelijk van elkaar in serie uit te voeren, ofwel door eerst de akoestische terugkoppeling te onderdrukken in alle microfoonsignalen en vervolgens het ruisonderdrukkingsalgoritme uit te voeren, ofwel omgekeerd, door eerst het multi-microfoon ruisonderdrukkingsalgoritme uit te voeren en daarna een éénkanaals feedbackonderdrukkingsalgoritme. Alhoewel de laatste oplossing de voorkeur geniet op het vlak van berekeningscomplexiteit, denken we dat de tweede oplossing verkieslijk is wat performantie betreft. Zo kan de aanwezigheid van het ruisonderdrukkingsalgoritme de identificeerbaarheid van het signaalmodel en het feedbackpad bevorderen omwille van zijn tijdsvariërend karakter. Daarenboven loont het de moeite om na te gaan in hoeverre de informatie van extra microfoonsignalen de voorwaarden voor identificeerbaarheid van het feedbackpad al dan niet versoepelt. Tenslotte zou het interessant zijn om de ruis- en feedbackonderdrukking te combineren in één multi-microfoons algoritme, waarbij –aan de hand van de verschillende microfoonsignalen en de ingang van de luidspreker– het gewenste geluidssignaal, i.e. zonder terugkoppeling en ruis, rechtstreeks wordt geschat. Dit probleem vertoont analogie met de gezamenlijke onderdrukking van ruis en echo, waarvoor recent binnen de onderzoeksgroep algoritmes werden ontwikkeld. Echter, de aanwezigheid van het signaalverwerkingspad vergt specifieke oplossingen.

Chapter 1

Introduction

HEARING loss is one of the most prevalent chronic health conditions, affecting about 500 million people world-wide. According to many surveys, one out of ten people suffers from hearing loss and would benefit from using hearing aids [107]. Because of the increased exposure to noise in daily life and the aging of the world's population, this number is expected to further increase in the future.

The most common type of hearing impairment is sensorineural hearing loss. People with this kind of hearing loss do not only suffer from an increased hearing threshold but also from a reduced speech discrimination ability, making it difficult for them to communicate in noisy environments. Compensation of sensorineural hearing loss does not only require amplification of the perceived sound signal, as is done in traditional hearing aids, but also reduction of the background noise with respect to the desired speech signal.

Recent advances in electronics miniaturization made it possible to integrate a digital signal processor and two (or more) microphones in commercial hearing instruments. However, due to the lack of effective digital signal processing algorithms, the additional possibilities of digital signal processing are not yet fully exploited, so that the benefit of commercial hearing aids for hearing impaired people is still limited. Commercial hearing aids do not sufficiently discriminate between speech and noise. Furthermore, in real-life use, they often suffer from an additional performance loss due to imperfections such as microphone mismatch, so that speech intelligibility in noise is hardly or insufficiently improved. In addition, as hearing aids become smaller and smaller, acoustic feedback, i.e., the acoustical coupling between the loudspeaker and the microphone(s) of the hearing aid, poses a major problem to hearing aid users. Acoustic feedback results in severe distortion of the desired signal and howling if the hearing aid

gain is increased. As a result, the maximum amplification that can be used in a commercial hearing aid is often too small to compensate for the hearing loss in a patient. Therefore, an urgent demand exists for efficient and well-working signal processing algorithms for noise reduction and acoustic feedback suppression.

In this thesis, we develop *several noise reduction algorithms and acoustic feedback suppression techniques for hearing aids*. Chapter 2 to Chapter 5 concentrate on multi-microphone noise reduction: in the algorithmic design, we aim for a good noise reduction performance, a low sensitivity to imperfections such as microphone mismatch and a low computational cost. Chapter 6 to Chapter 7 discuss adaptive feedback cancellation techniques that provide a better feedback suppression and sound quality than existing techniques.

1.1 Preliminaries

This section briefly describes the auditory system as well as the impact of hearing impairment on the ability to perceive sound.

1.1.1 The auditory system

The auditory system, depicted in **Figure 1.1**, consists of three major parts: the outer ear, the middle ear and the inner ear. Each part of the system serves a vital function in our ability to effectively hear sound [174, 175]. The outer ear consists of the pinna and the ear canal. The pinna captures the sound energy and directs it through the ear canal to the ear drum. The middle ear converts the sound waves into mechanical vibrations, which are transmitted to the fluid in the inner ear. The inner ear contains the cochlea and the auditory nerve. The cochlea converts the mechanical signals into neural signals. The auditory nerve transmits these neural pulses to the brain, where they are translated into specific sounds. The auditory signal processing in the cochlea is largely responsible for the hearing-specific properties such as the frequency-specific sensitivity and the low hearing threshold.

The cochlea is divided along its length by the basilar membrane: sound waves result in pressure differences between one side of the basilar membrane and the other and hence, in movement of the basilar membrane. Hair cells are aligned into three to five outer rows (outer hair cells) and one inner row (inner hair cells) that extend along the membrane. These outer and inner hair cells have stereocilia or "hairs" that stick out and are in contact with a second membrane, called the tectorial membrane. When the basilar membrane moves up and down, the stereocilia at the tops of the hair cells bend back and forth. The mechanical properties of the basilar membrane vary progressively along its length so that each region of hair cells on the basilar membrane responds

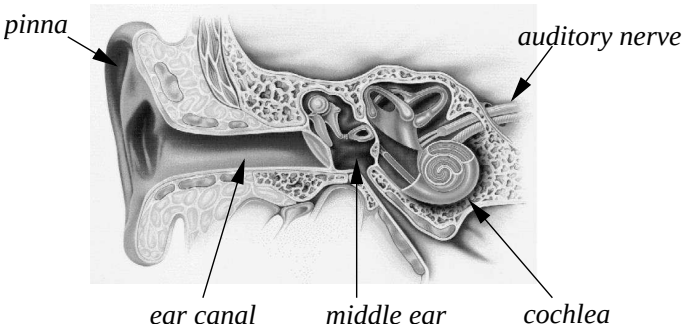


Figure 1.1: The auditory system.

best to a specific characteristic frequency: hair cells in the base of the cochlea respond to high frequencies, while hair cells situated in the apex of the cochlea react to low frequencies. This explains the *frequency-specific sensitivity* of the ear, which can be compared with a filterbank.

The movement of the stereocilia of the inner hair cells leads to the generation of neural pulses in the auditory nerve. Hence, the inner hair cells convert mechanical motion into neural activity. The outer hair cells have a different function in the cochlea: they selectively amplify the vibrations of the basilar membrane in a highly nonlinear way. The outer hair cells are responsible for the *high frequency resolution* and the *high sensitivity to weak sounds (i.e., the low hearing thresholds)* of the auditory system. In addition, they produce nonlinear, compressive input-output functions on the basilar membrane for frequencies close to the characteristic frequency, which allows acoustic information over a wide dynamic range to be transferred to the brain.

1.1.2 Hearing impairment

According to the part of the auditory system that is affected, hearing loss is classified into conductive, sensorineural or mixed hearing loss. *Conductive hearing loss* is caused by problems in the outer and middle ear that interfere with the transmission of sound to the inner ear. This type of hearing loss can be corrected by medical intervention or by (frequency-dependent) amplification of sound. *Sensorineural hearing loss* refers to problems in the inner ear, i.e., either the cochlea (i.e., cochlear hearing loss) or the auditory nerve (i.e., retrocochlear hearing loss). In about 90 % of the cases, hearing loss belongs to this last category.

The majority of cases of sensorineural hearing loss are *cochlear hearing loss* and are caused by damage to hair cells in the cochlea so that the conversion from mechanical movement to neural activity is affected [174, 175, 178]. Since

cochlear hair cells, once destroyed, do not regenerate, this type of hearing loss is permanent. The most common causes for the destruction of hair cells are aging and exposure to loud sound. The hair cells in the base of the basilar membrane are usually first damaged so that hearing loss first occurs at the high frequencies.

Usually, the outer hair cells are more susceptible to damage than the inner hair cells. *Damage to the outer hair cells* produces several changes in the perception of sound: an increased hearing threshold (i.e., loss of sensitivity to weak sounds), a reduced frequency selectivity and a reduced dynamic range. The most obvious sign of a hearing problem is an *increased hearing threshold*. If the outer hair cells are destroyed, the amplitude of the basilar membrane vibrations is reduced, leading to a loss of sensitivity to weak sounds. A second change is in frequency selectivity, which refers to the ability to separate or resolve frequency components in a sound. Damage of the outer hair cells leads to reduced sharpness of tuning on the basilar membrane, which can be interpreted as a widening of the filters in the auditory filter bank, and hence, *reduced frequency selectivity*. As a result, hearing impaired people are very susceptible to masking produced by background noise, which explains the great difficulty they experience in understanding speech in noisy environments. In addition, because of the reduced frequency selectivity, speech cues are blurred, making it harder to understand speech, even when no background noise is present. A third consequence of outer hair cell damage is *loudness recruitment and a reduced dynamic range*, which mainly result from the loss of the normal compression on the basilar membrane. Loudness recruitment refers to an abnormal growth of loudness level to a slight increase in intensity of an acoustic signal.

In addition to outer hair cell damage, also *damage to the inner hair cells* gives rise to an elevated hearing threshold. Inner hair cell damage results in a less efficient transduction of mechanical vibrations into neural activity, so that the amount of basilar membrane vibration needed to reach the threshold is larger than normal. Other consequences of inner hair cell damage are a reduced flow of auditory information in the auditory nerve, and, in extreme cases, no transfer of activity at some regions of the basilar membrane, so-called dead regions.

Depending on the amount of inner and outer hair cell damage, the hearing loss varies from mild (i.e., between 20 dB to 40 dB), moderate (i.e., between 40 dB and 70 dB), severe (i.e., between 70 dB and 90 dB) to profound or "deafness" (i.e., larger than 90 dB).

1.1.3 Hearing aids and cochlear implants

Mild to severe hearing loss may be (partially) compensated for by hearing aids. The general configuration of a hearing aid is depicted in **Figure 1.2**. The *hearing aid* records the sound signal(s) in the acoustical environment through

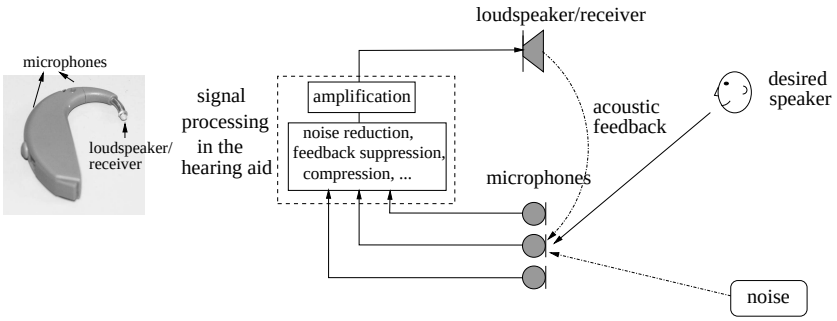


Figure 1.2: Configuration of a hearing aid.

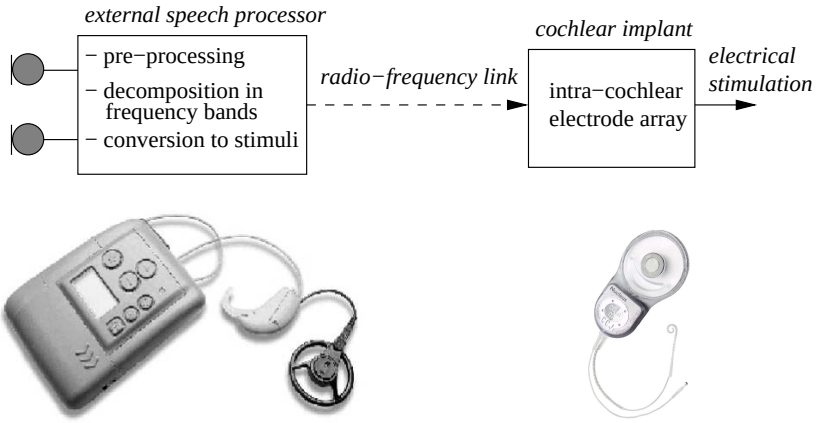


Figure 1.3: Configuration of a cochlear implant.

one or more microphones. The acoustic sound signal(s) consist of the speech signal uttered by the speaker to whom the hearing aid user is listening, one or more interfering background noises (e.g., computer fans, other speakers present in the room, traffic noise, ...) as well as possible signal leakage from the hearing aid loudspeaker to the microphone, also referred to as acoustic feedback (cf. Section 1.3). The recorded microphone signal(s) are processed by some signal processing algorithms (cf. Section 1.2 and Section 1.3) with the aim of compensating for the hearing impairment of the hearing aid user. The processed sound is then applied to the ear canal through a loudspeaker (also known as receiver). Hence, hearing aids assume that the inner hair cell function is still in tact.

People with a profound hearing loss have little inner and outer hair cells left, so that no benefit may be obtained with a hearing aid. If the auditory nerve is still in tact, a *cochlear implant* may restore the perception of sound in these

patients: a cochlear implant bypasses damaged hair cells and directly stimulates the auditory nerve by means of an intra-cochlear electrode array, which is inserted in the cochlea of the patient by a surgeon. **Figure 1.3** depicts the general configuration of a cochlear implant [153]. The acoustic sound signal(s) are recorded through one or more microphones, usually located in a behind-the-ear hearing aid. An externally worn speech processor applies some pre-processing to the microphone signals (e.g., compression, noise reduction), decomposes the sound signal in different frequency bands for delivery to the different electrodes and converts the processed frequency-domain signals into electric current stimuli. These electric stimuli are transmitted by means of a radio-frequency (RF) link to the intra-cochlear electrodes of the cochlear implant, which stimulate the auditory nerve at different places in the cochlea.

1.2 Noise reduction in hearing aids and cochlear implants

The task of a hearing aid or cochlear implant is to process the perceived sound in such a way that the normal hearing function is restored. Section 1.2.1 identifies the major signal processing algorithms that are required in a hearing aid or cochlear implant to compensate for sensorineural hearing loss. People with a sensorineural hearing loss have more difficulty understanding speech in noise than people with normal hearing. An effective way to compensate for this is the use of a noise reduction algorithm. An overview of the literature on noise reduction is given in Section 1.2.2.

1.2.1 Signal processing challenges

Compensation of sensorineural hearing loss does not only require amplification of the perceived sound signal, as is done in traditional hearing aids, but also additional, more complex signal processing by the hearing aid (or cochlear implant) such as noise reduction and dynamic range compression.

Noise reduction

In daily life, people are often exposed to background noise. This noise e.g. arises from audio equipment, computer fans, traffic, other speakers present in the room (i.e., cocktail party noise). One of the major complaints of people with a cochlear hearing loss is the difficulty understanding speech when background noise is present [45, 86, 135, 175] (cf. Figure 1.2). This disability not only arises from the reduced audibility, but also from the reduced frequency selectivity, the loudness recruitment and other abnormalities of perception (e.g., a reduced ability to localize sound) [6, 55, 175, 178]. As a result, it can not be resolved by a pure (frequency-dependent) amplification of the incoming sound: hearing

aid users with ordinary hearing aids often report that they hear people talking but do not understand what is said.

The capability of an individual listener to understand speech in noise is often measured by the speech-reception threshold (SRT) in noise. The SRT is defined as the signal-to-noise ratio (SNR) where 50 % of the speech is understood. While people with normal hearing still succeed to understand 50 % of the speech in a noisy environment with an SNR as low as -5 dB, people with a mild to severe cochlear hearing loss may require an SNR that is up to 15 dB (on average 5 dB) higher [45, 200, 201, 178]. The deficit in SRT relative to normal hearing is particularly large, when the background noise is fluctuating and when the desired and interfering source are spatially separated [178].

A very effective way of compensating for the overall effect of cochlear hearing loss on the ability to understand speech in noise, is to increase the SNR by means of a *noise reduction algorithm* [176]. In order to be beneficial for a majority of hearing aid users, the algorithm should reduce the background noise by at least 5 dB [44]. For cochlear implant users, the required amount of noise suppression is even larger [134]: the SRT of cochlear implant users lies between +5 dB and 15 dB [95, 197, 278]. A small improvement in SNR, though, may result in a large improvement in intelligibility [178, 200, 201, 225].

In the past, hearing aids were only equipped with one single omnidirectional microphone. However, *single-microphone* noise reduction techniques can only differentiate between signals that have different temporal and spectral characteristics. As a result, they cannot improve the local SNR in each frequency band, but only the global SNR. Speech intelligibility is then only improved if the frequency bands that are attenuated contain so much noise that they mask out the useful speech information in other bands. However, the frequency spectra of everyday noises are often similar to the spectrum of speech, which makes it difficult for single-channel noise reduction schemes to effectively eliminate most of the noise without reducing speech intelligibility at the same time. An example is the well-known cocktail party situation, where speech has to be understood in the presence of multiple talkers.

The limited benefit of single-microphone techniques for speech intelligibility [136, 178, 180, 257, 267] is manifested by the growing tendency in the development of commercial hearing aids towards the use of directional microphone(s) and/or multi-microphone techniques. These techniques allow to exploit the spatial diversity of the speech and the noise signals in addition to the spectral and temporal differences.

Most studies on *multi-microphone* hearing aid applications [74, 72, 97, 123, 131, 140, 198, 214, 225, 224, 245] make use of microphone arrays with dimensions that exceed those of commercially behind-the-ear (BTE) hearing aids. However, in order to be aesthetically acceptable, the size of the hearing aid should

be as small as possible. Thanks to the ongoing miniaturization in electronics, more microphones and more processing power can be built-in in current and future hearing aids. At present, commercial BTE hearing aids with two (e.g., Phonak Claro, GN ReSound Canta 7) and even three (e.g., Siemens Triano 3) omnidirectional microphones are already commercially available. Because of the small microphone interspacing, the noise reduction performance is, however, sensitive to imperfections encountered in real-life applications such as array misalignment, imperfections in the mounting and microphone mismatch. In [114], e.g., gain and phase differences between microphone characteristics of up to 6 dB and 10°, respectively, have been reported. Calibrating each hearing aid [19] individually is cumbersome and expensive. Moreover, even after calibration, the microphone characteristics may drift over time, e.g., due to aging or the presence of dirt in the acoustic pathway [114, 251]. Consequently, commercial hearing aids often suffer from a performance loss in real-life use. Hence, in addition to the noise reduction performance, robustness against imperfections is a major issue in the development of multi-microphone noise reduction algorithms for hearing aids.

Remark 1.1 With the future availability of a high-rate wireless link between left and right ear, also *binaural hearing aid systems* (i.e., with two audio outputs, one at each ear) become feasible for commercial application. Binaural signal processing techniques try to provide the natural benefits of sound localization and speech intelligibility that originate from binaural listening by combining information from microphones at both ears [32, 81, 100, 170, 186, 271, 274]. In this thesis, we will, however, concentrate on small-sized monaural hearing aid systems (i.e., with one audio output) located at one single ear. □

Other signal processing techniques

In addition to noise reduction, also other signal processing techniques are applied or considered for application in hearing aids, attempting to improve the speech intelligibility and/or the listening comfort. These signal processing features include dynamic range compression, spectral contrast enhancement and sound classification.

Due to the destruction of the outer hair cells, people with cochlear hearing loss have lost the normal compression function on the basilar membrane, resulting in a reduced dynamic range and loudness recruitment. Dynamic range compression [101, 102, 119, 151, 177, 178, 248, 249] tries to compensate for the loudness recruitment and the reduced dynamic range by restoring the audibility of weak sounds without loud, intense sounds being over-amplified and hence, becoming uncomfortably loud. With linear amplification this is not possible. The main benefit of compression is that it makes speech audible over a wide range of sound levels without adjustment of the volume control [178].

To compensate for the effect of reduced frequency selectivity, spectral contrast enhancement techniques [7, 61, 279] have been studied and commercially applied (e.g., GN ReSound Canta 7). Spectral contrast enhancement increases the amplitude difference between peaks and valleys in the sound spectrum in order to compensate for the blurring effect that is caused by the reduced frequency resolution. However, the improvement of spectral contrast enhancement on the ability to understand speech is questionable [7, 61, 178].

Hearing aids and cochlear implants are used in different acoustic environments (e.g., speech in quiet, speech in noise, music). Depending on the listening situation, different signal processing features (e.g., a different amplification scheme [133, 137], the use of a noise reduction algorithm) may be required to ensure optimum listening comfort, sound quality and intelligibility. E.g., the use of a noise reduction algorithm is essential during a conversation in the presence of background noise, but may be disadvantageous in other listening conditions such as traffic situations or when listening to music [3, 193]. Therefore, some of the latest hearing aids are provided with different hearing programs. The different programs can then be activated by the hearing aid user by means of a switch button. The manual switching between programs may, however, be annoying or even difficult for some users. To enable an automatic switching by the hearing aid, sound classification algorithms are being developed [3, 124, 193] and implemented in hearing aids (e.g., the AutoSelect^{TM} system in Phonak Claro). Sound classification algorithms distinguish between different listening environments and background noises based on sound features such as the modulation characteristics of the signal (e.g., the level of amplitude modulation).

Dynamic compression, spectral contrast enhancement and sound classification techniques are not studied in this thesis.

1.2.2 State-of-the-art noise reduction techniques

This section summarizes the literature on noise reduction. In Section 1.4, we indicate which noise reduction techniques are currently adopted in commercial hearing aids. A detailed mathematical explanation of the noise reduction techniques that are important for this thesis, can be found in Chapter 2.

Among the different noise reduction algorithms proposed in the literature, we can distinguish between single-channel and multi-microphone techniques.

Single-channel noise reduction

Single-channel noise reduction techniques only exploit spectral and temporal differences between the speech and the noise signal to suppress the noise. Basically, these techniques compute estimates of the short-term spectral characteristics of the speech and the noise. These estimates are then combined according to a certain optimization criterion to produce an enhanced speech

signal. Single-channel noise reduction algorithms can be broadly classified in parametric or model-based approaches and non-parametric approaches. *Parametric techniques* [31, 63, 64, 82, 244] model the speech and sometimes also the noise [66] as a stochastic auto-regressive model (AR). Based on the estimates of the AR-parameters, a Kalman filter (time-domain) [63, 66] or Wiener filter (frequency-domain) [31, 82, 244] is computed, which is then applied to the noisy speech signal. *Non-parametric techniques* do not estimate the speech parameters, but rather exploit an estimate of the noise statistics to produce an enhanced speech signal. Spectral subtraction and single-channel subspace-based techniques belong to this category. The estimation and speech enhancement takes place in a transform-domain: spectral subtraction algorithms [16, 53, 52, 79, 156, 164, 165, 168, 265] operate in the frequency-domain (e.g., the Discrete-Fourier Transform (DFT) domain or the wavelet transform domain [78]), while subspace based techniques transform the signals by means of a generalized singular value decomposition, which is also referred to as the Karhunen-Loève Transform (KLT) [34, 54, 83, 105, 110, 115, 148, 172, 208]. Generally, single-channel speech enhancement techniques assume that the noise statistics are more stationary than the statistics of the speech so that they can be estimated during periods of noise-only. Recently, a minimum statistics approach has been proposed that also allows to update the noise power spectral density during speech activity, by tracking the spectral minima of the input signal [163].

Although an increase in listening comfort has been reported in many cases, single-channel noise reduction techniques have so far produced no or limited benefit for speech intelligibility [4, 178], indicating that an increase in SNR does not automatically yield an increase in intelligibility. This is most likely due to the fact that single-channel techniques cannot improve the local SNR at each frequency (cf. Section 1.2.1) as well as to the artefacts introduced by the nonlinear filtering process and an inaccurate estimation of the noise characteristics. In real-life communication situations, the speech and the noise signal considerably overlap in the time- and frequency-domain, which makes it extremely difficult for single-channel techniques to substantially suppress the noise without introducing speech distortion and so-called musical noise [22]. Especially in low SNR and spectrally highly non-stationary noise (such as babble noise), which are typical ingredients of a cocktail-party situation, they suffer from a bad performance.

Multi-microphone noise reduction

In most scenarios, the desired speaker and the disturbing sources are located at different positions in space. Multi-microphone noise reduction techniques are able to exploit this spatial diversity in addition to spectral and temporal differences: a beam can be steered towards the desired speaker and/or zeroes can be steered in the directions of the noise sources. Multi-microphone techniques can be classified into fixed beamforming techniques and adaptive techniques.

The latter include adaptive beamforming and multi-channel Wiener filtering techniques.

Fixed beamforming

In *fixed beamforming* techniques, the filters that are applied to the microphone signals are fixed and hence, data-independent. The filter coefficients are optimized so that a focussed beam is steered to the target direction, i.e., typically the forward-look direction in hearing aids, while reducing the background noise coming from other directions as much as possible. Fixed beamformers that have been considered for hearing aid applications include delay-and-sum beamformers [170, chapter 4], differential microphone arrays [49, 254] and superdirective beamformers [9, 28, 32, 123, 131, 214, 224, 225, 245]. A more general approach to design broadband beamformers is filter-and-sum beamforming [37, 39, 120, 142, 188, 191, 280], where the filter coefficients are optimized so that the spatial directivity pattern optimally fits a given desired directivity pattern with respect to some cost function. The design of fixed beamformers relies on assumptions about the signal model (e.g., identical and perfectly omnidirectional microphones, known array geometry and speaker position, ...). To guarantee their performance in real-life use, special attention should therefore be paid to the robustness against imperfections (e.g., microphone mismatch, misteer), especially when used in combination with small-sized arrays such as hearing aids [28, 245]. Robustness against imperfections can be improved by performing a (time-consuming) calibration procedure [19, 250, 280] with each hearing aid, by imposing a sensitivity constraint on the filter coefficients [28, 245] or by incorporating statistical knowledge of the possible model errors in the design [37, 38]. Obviously, calibrating each hearing aid individually is time-consuming and expensive, so that especially the last two techniques are of interest for hearing aid applications.

Adaptive beamforming

The second class of multi-microphone techniques are adaptive techniques. In contrast to fixed beamformers, these signal processing techniques make use of data-dependent filter coefficients that are adapted to respond to time-varying environments. Adaptive techniques generally have a better noise reduction performance than fixed beamforming techniques, particularly if the number of interferers is small (i.e., smaller than the number of microphones) and in acoustic environments with little reverberation, but are more sensitive to distortion or cancellation of the desired speech signal. Two categories of adaptive techniques can be distinguished: adaptive beamforming and multi-channel Wiener filtering techniques.

Adaptive beamforming techniques (e.g., the Frost beamformer [62, 261], the Generalized Sidelobe Canceller [77]) typically solve a linearly constrained minimum variance (LCMV) optimization criterion [62, 117, 260, 261], minimizing

the output power or output noise power subject to the constraint that signals coming from a certain region or direction (i.e., ideally the direction of the desired speech source) are preserved. The LCMV beamformer has been widely studied for application in hearing aids [72, 74, 75, 98, 96, 97, 131, 140, 141, 139, 157, 160, 198, 271]. An efficient realization of the LCMV is the Generalized Sidelobe Canceller (GSC) [20, 77]. The GSC transforms the constrained LCMV optimization criterion into an unconstrained criterion through a combination of a fixed spatial pre-processor, i.e., a fixed beamformer and a blocking matrix, and an adaptive noise canceller (ANC). The ANC then minimizes the output power while the blocking matrix is designed to avoid so-called speech leakage into the noise references. However, because of room reverberation, microphone mismatch, microphone placement uncertainty and misteer, the speech signal may leak into the noise references, resulting in speech distortion and speech cancellation by the ANC. Therefore, different variants of the standard GSC have been proposed. Speech leakage is reduced by incorporating multiple linear constraints in the design of the fixed spatial pre-processor [96], by using a spatial filter designed blocking matrix [187] or an adaptive blocking matrix [92, 94, 103]. To limit speech distortion caused by the ANC, adaptation is only performed during periods of noise-only [74, 92, 258] or/and a quadratic inequality constraint is applied to the ANC [29, 92, 96, 103, 108, 157] (cf. Chapter 2).

Multi-channel Wiener filtering

A second category of adaptive multi-microphone noise reduction techniques is multi-channel Wiener filtering (MWF) [2, 35, 36, 57, 109, 162, 189, 190, 212, 213, 220, 235]. These techniques provide a minimum mean square error (MMSE) estimate of either the clean *speech source* signal, the (reverberant) speech component in one of the microphone signals or a given reference signal. In contrast to adaptive beamforming techniques, MWF techniques exploit both spectral and spatial differences between the speech and the noise sources, so that inevitably some speech distortion will be introduced. Among the different MWF techniques, we can distinguish between three classes, i.e., the GSC with single-channel post-filter [56, 162, 220], the MWF using calibration sequences [189, 190] and the MWF with unknown reference [2, 35, 36, 212, 213, 235].

In this thesis, we concentrate on the third class of MWF techniques, i.e., the MWF with unknown reference. In contrast to the GSC and the other two MWF techniques, this MWF technique requires neither a priori information about the desired signal model nor calibration, making it appealing from a robustness point of view. Because of its high computational complexity, it has, however, not been considered yet for application in hearing aids. In this thesis, we compare the performance and robustness of the MWF with unknown reference and the GSC for small-sized arrays as used in commercial hearing aids. In addition, low-cost subband and stochastic gradient implementations of the MWF are derived, making it feasible for commercial hearing aids.

1.3 Feedback suppression in hearing aids

In addition to the difficulty understanding speech in noise, the occurrence of acoustic feedback poses a major problem to hearing aid users. Acoustic feedback, which refers to the acoustical coupling between the loudspeaker and the microphone(s) of the hearing aid (cf. Figure 1.2), introduces serious artefacts that compromise the potential benefit of a hearing aid. As a result, also feedback suppression techniques (cf. Section 1.3) are called for in hearing aids. The acoustic feedback problem is explained in Section 1.3.1. Section 1.3.2 gives an overview of the literature on feedback suppression.

1.3.1 Acoustic feedback in hearing aids

Acoustic feedback refers to the acoustical coupling between the loudspeaker and the microphone(s) of the hearing aid (cf. Figure 1.2). Because of this coupling, the amplified sound sent through the loudspeaker is fed back into the microphone(s). As a result, the hearing aid produces severe distortion of the desired signal and an annoying howling sound when the gain is increased. The use of an amplification that brings the system close to instability indicates that the hearing aid user desires more amplification from the hearing aid than can actually be provided. As a direct result of this gain limitation, low-energy signals fall below the hearing threshold so that the hearing aid does not compensate for the hearing loss in the patient [194]. To increase the maximum gain that can be used in a hearing aid without making it unstable, *feedback suppression* algorithms are called for, particularly in hearing aids for the severely hearing impaired.

In most cases, the acoustic feedback stems from the vent, i.e., the hole in the earmold of the hearing aid [90, 129, 205]: depending on the size of the vent, venting establishes an acoustic feedback path that limits the maximum stable gain in a hearing aid to 40 dB and often even less [129]. Venting is, however, essential to reduce the occlusion effect. This effect [91, 121, 207, 272] refers to the increase in loudness of the own voice and the low frequency boost that hearing aid users experience when the ear canal is completely blocked with an earmold. The unnatural perception of their own voice while talking is disturbing to most hearing aid users and is often a reason to stop wearing their hearing aids. Hence, eliminating the vent or considerably reducing its size is not an acceptable solution for the acoustic feedback problem. On the contrary, in the hearing industry, there is even a growing tendency towards hearing aids with an open fitting (i.e., without an earmold) to improve listening comfort and binaural hearing (e.g., ReSound Air^{TM} of GN ReSound).

In addition to the vent size, also the geometric configuration of the hearing aid, the ear canal and the acoustics outside the ear determine the feedback path [91]. Because of the shorter distance between loudspeaker and micro-

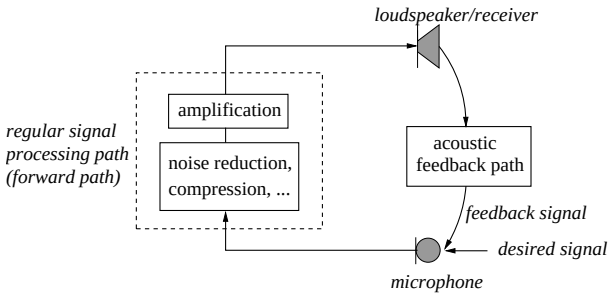


Figure 1.4: Feedforward suppression algorithm.

phone(s), the attenuation of the feedback path is smaller for in-the-ear (ITE) and in-the-canal (ITC) hearing aids than for behind-the-ear (BTE) hearing aids. Since the ear canal shape differs among hearing aid users, the feedback path is user-dependent. In addition, the acoustic feedback path can vary significantly depending on the acoustics outside the ear (e.g., room reverberation, jaw movements, etc.) [91, 128, 129, 205]: reflecting surfaces close to the ear, such as a telephone handset or the palm of a hand, can temporarily reduce the feedback path attenuation by 10 dB to 20 dB and hence, cause instability [91, 205, 246].

People with a moderate hearing loss suffer from an increase in hearing threshold of 40 dB to 70 dB (cf. Section 1.1.2), while the maximum stable gain in a hearing aid is limited to at most 40 dB and often even less. As a result, there is a strong need for efficient feedback suppression techniques. Because of the time-variability of the feedback path, adaptive solutions that can deal with variations in the feedback path, are called for.

1.3.2 State-of-the-art feedback suppression techniques

To reduce the negative effects introduced by acoustic feedback (i.e., howling and the limited maximum possible amplification), several techniques have been proposed in the literature. A brief overview of these techniques is provided in this section. For a detailed mathematical explanation, we refer to Chapter 6. In Section 1.4, we indicate which feedback suppression techniques are currently implemented in commercial hearing aids.

Feedback suppression techniques can be broadly classified into feedforward suppression and feedback cancellation techniques.

Feedforward suppression

In feedforward suppression techniques, the regular signal processing path of the hearing aid, i.e., the so-called forward path (cf. **Figure 1.4**), is modified in such

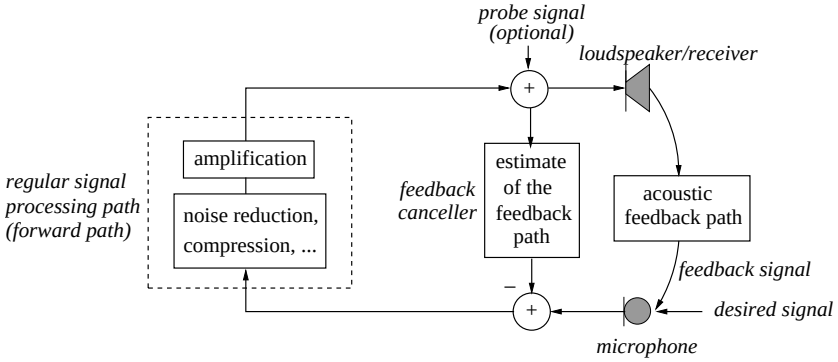


Figure 1.5: Feedback cancellation algorithm.

a way that it is stable in conjunction with the feedback path. The most common technique is the use of a notch filter. In a notch filter, the gain is reduced in a narrow frequency band around the critical frequencies, whenever feedback occurs [21, 99, 166, 202]. Other examples include equalizing the phase of the open-loop response [84, 268] and using time-varying elements (such as frequency shifting, delay and phase modulation) in the forward path [21, 184, 216]. The increase in maximum stable gain with feedforward suppression techniques has generally been found limited. In addition, feedforward suppression techniques all compromise the basic frequency response of the hearing aid, and hence, may seriously affect the sound quality.

Feedback cancellation

A more promising solution for acoustic feedback is the use of a feedback cancellation algorithm, which is depicted in **Figure 1.5**. The feedback canceller estimates the feedback signal and subtracts this estimate from the microphone signal, so that ideally, only the desired signal is preserved at the input of the forward path. Since the acoustic path between the loudspeaker and the microphone can vary significantly depending on the acoustical environment [91, 128, 205, 246], the feedback canceller must be adaptive.

Currently available adaptive feedback cancellers can be divided into two classes: algorithms with a continuous adaptation and algorithms with a *non-continuous adaptation* [76, 122, 166]. The latter adapt the coefficients of the feedback canceller only when instability is detected or when the input signal level is low [122, 126, 127, 196, 253]. Due to the reactive, rather than proactive, adaptation, these systems may be objectionable. A *continuous adaptation* feedback canceller (CAF) continuously adapts the coefficients of the feedback path estimate in such a way that the energy of the feedback compensated signal is minimized. However, the adaptive modelling problem encountered here ap-

pears to be highly non-trivial, because of the presence of a closed signal loop (i.e., the forward path) that introduces specific signal correlation. As a result, the standard continuous adaptation feedback cancellers fail to provide a reliable ('unbiased') feedback path estimate, when the desired signal is spectrally colored[1], e.g., a speech or music signal. The CAF then tends to cancel the desired signal instead of the feedback signal, so that the desired signal gets distorted.

Solutions for the 'bias'-problem

In the literature, different solutions have been proposed to reduce the bias or model error of the standard CAF. Classical approaches include introducing signal decorrelating operations in the forward path (such as delays or non-linearities) [112, 118, 184, 221], adding a probe signal to the loudspeaker input [51, 222] and controlling the adaptation of feedback canceller, e.g., by means of constrained adaptation [24, 67, 125, 129, 132, 147, 182, 183, 226]. Generally, these techniques involve a trade-off between desired signal cancellation on the one hand and tracking performance or sound quality on the other hand, so that typically a significant residual bias remains.

In [88, 89, 217], alternative solutions for the bias in the feedback path estimate have been presented that are based on closed-loop system identification [59, 60]. Depending on the assumptions that are made about the forward path, closed-loop identification methods can be classified as *direct*, *indirect* or *joint input-output* methods [59]. In contrast to the indirect and the joint input-output approach, the direct method requires neither assumptions about the forward path (e.g., linear) nor the presence of an external probe signal, making it an appealing approach for feedback cancellation. In [87, 88, 89], an adaptive feedback cancellation technique based on the direct method of closed-loop identification and a *fixed* estimate of the desired signal spectrum has been proposed: it is shown that the bias of the feedback path estimate can be avoided completely provided that the desired signal is non-tonal and its spectrum is known. In practice, the desired signal spectrum is unknown and highly time-varying, so that it is desirable to estimate this model adaptively. In general, however, it has been proven that the feedback path and the desired signal spectrum cannot both be identified in a closed-loop setup, unless a probe signal is inserted or non-linearities are introduced in the forward path [87, 89].

In this thesis, we revise the conditions for identifiability of the feedback path and the desired signal spectrum with the direct closed-loop identification approach: we show that not only inserting non-linearities or a probe signal, but also adding a small delay in the forward path or in the cancellation path (i.e., at the input of the feedback canceller) can render the system identifiable. In

[1] Spectral coloring refers to the fact that certain spectral components are stronger than other spectral components. A spectrally colored signal may be a broadband (e.g., a speech signal) as well as a narrowband signal (e.g., a sinusoid or alarm signal).

addition, adaptive feedback cancellation techniques based on direct closed-loop identification of both the feedback path and the desired signal spectrum will be developed, with the aim of providing an unbiased feedback path estimate for unknown and time-varying desired signals.

1.4 Commercial hearing aids

Although digital signal processing techniques for feedback suppression and noise reduction have been studied for a while, it is only since a few years that they have found application in commercial hearing aids.

In the past, only analog or at most analog programmable hearing aid systems were available. The signal processing in these instruments was mainly limited to frequency-dependent amplification and compression [276]. Recent advances in the field of digital circuit design and manufacturing made the use of digital signal processing in commercial hearing instruments possible. In 1996, the first fully-digital commercial hearing aids appeared on the market (e.g., Widex Senso). Despite their high cost (i.e., on average 2000 euro per hearing aid) compared to analog instruments, digital hearing aids are nowadays the most growing segment of the two billion dollar whole-sale hearing aid market. Considering the fact that only about 5 % of all people who can benefit from a hearing aid actually purchases and uses one, there is still a huge additional market potential left. Furthermore, the hearing aid market is expected to grow substantially over the next decades due to aging of the world's population.

Digital hearing aids allow for more advanced signal processing features such as noise reduction and feedback suppression[276]. However, because of power restrictions and size limitations, the processing power available in current digital hearing aids is limited. The latest developed digital BTE hearing aids e.g., have a processing power of 2 to 4 millions of instructions per second (MIPS) and a program and data memory of about 2 to 4 kByte each [129, 215], but require that batteries are replaced every week. Consequently, the signal processing schemes that are applied in currently available commercial hearing aids are rather simple. Below, we indicate which noise reduction and feedback suppression techniques are typically adopted in commercial hearing aids.

Remark 1.2 In current cochlear implants, the available processing capacities are larger (e.g., $20 - 100$ MIPS) because of the larger power consumption required for the RF data link and the resulting larger speech processor size. However, the processing power has to be shared with the basic cochlear implant processing such as the conversion to electrical stimuli. In addition, the ongoing miniaturization of cochlear implants (e.g., the use of a behind-the-ear instead of a body-worn speech processor) motivate to keep the computational cost of the noise reduction algorithms as low as possible in order to limit power consumption. □

1.4.1 Noise reduction

The first commercial hearing aids consist of one single microphone, which is either omnidirectional or directional (e.g., the analog Phonak Classica C-D). Directional microphones have a stronger sensitivity to the front direction than to the back so that signals coming from behind the listener are attenuated. However, in some scenarios, the use of directivity was found to be disadvantageous (e.g., safety in traffic, listening to music). Therefore, most single-microphone hearing aids are equipped with an omnidirectional microphone.

Single-microphone hearing aids reduce the noise based on differences in frequency spectra or differences in modulation. The simplest technique consists of a high pass filter [149]. Provided that the background noise has most of its power at the low frequencies so that it masks the low frequency content of the speech signal, a high pass filter reduces the loudness of the noise, while not or little affecting speech intelligibility. More advanced single-microphone noise reduction schemes decompose the incoming signal into different frequency bins and attenuate the different frequency bins based on the degree of amplitude modulation: frequency components with a high modulation are supposed to be speech and are passed through whereas frequency components with a low modulation are supposed to be noise and hence, attenuated [30]. These and other related single-channel noise reduction techniques have been applied in the first digital hearing aids (e.g., Oticon Digifocus, GN Danavox[2] Danalogic, Widex Senso, Siemens Prisma) and are still being used in the latest digital hearing aids (e.g., GN ReSound Canta 7, Phonak Claro, Widex Senso Diva). However, single-channel noise reduction algorithms have not been found to produce systematic improvements in speech intelligibility, but only improved listening comfort [136, 178, 180, 257, 267].

Thanks to the miniaturization of microphones [195], two and even three microphones (e.g., Siemens Triano 3) can now be integrated in commercial BTE and ITE hearing aids. Because of space limitations, multi-microphone noise reduction systems in ITE hearing aids consist at most of a combination of a directional and omnidirectional microphone, where a switch allows the user to choose between the directional or omnidirectional behavior (e.g., Siemens VoiceMic^{TM}, Etymotic D-mic). The directional microphone is either realized in hardware or in software using differential microphone array techniques [49, 254]. The actual improvement in speech intelligibility of these fixed microphone algorithms is, however, limited to 2 a 3 dB [277].

In 2001, the first commercial BTE hearing aids with adaptive multi-microphone noise reduction techniques appeared on the market (e.g., Phonak Claro [138], GN ReSound Canta7 [157], Widex Senso Diva [113]). In these systems, an adaptive differential microphone is efficiently realized by combining the output signals of two identical omni-directional microphones in endfire combination

[2]Now, GN ReSound.

(cf. Chapter 2): the noise reduction scheme adaptively places a null in the direction of the strongest interferer [157, 252]. In the presence of a single noise source, this two-microphone adaptive directionality system may offer a significant improvement in intelligibility with respect to fixed directional patterns [209]. In more complicated scenarios, such as multiple noise sources and reverberation, no significant benefit has been obtained.

To increase the noise reduction performance in complicated noise scenarios, hearing aid systems with three microphones have recently been introduced on the market, e.g., the array Link-iT of Etymotic Research and the BTE Siemens Triano 3. The array Link-iT of Etymotic Research is an analog head-worn assistive listening device that can be used in combination with a BTE. The array combines the outputs of three hardware directional microphones to produce a fixed, highly directive polar pattern [158, 224, 225]. Its considerable size (i.e., a length of about 5 cm), may, however, make people reluctant to wear it. In 2002, the first three-microphone BTE (with a microphone interspacing of 0.8 cm) has been commercialized by Siemens (i.e, Siemens Triano 3) [203]. This three-microphone BTE combines fixed first and second-order differential microphone arrays. At low frequencies, where the sensitivity to imperfections is the highest, a first-order differential microphone is created by combining the outer two microphones; at high frequencies, the three microphone signals are used to realize a second-order differential microphone.

Because of the ongoing miniaturization in the hearing aid industry, robustness against imperfections (such as microphone mismatch, array misalignment, the head shadow effect) has become a major issue in the design of noise reduction techniques for hearing aids. Indeed, currently available multi-microphone hearing aids often suffer from a performance degradation in real-life use because of a non-robust algorithmic design [19, 251]. Some hearing aid manufacturers try to compensate for this by purchasing expensive pairs of matched microphones, by performing a calibration procedure with a reference signal [19] or by incorporating some kind of online microphone matching procedure (e.g., GN ReSound Canta 7). Although these measures reduce the model errors to some extent, they do not completely avoid them (e.g., online microphone matching) and/or are too expensive and time-consuming to be of real practical use (e.g., calibration with a reference signal). E.g., online microphone matching techniques typically assume that the sound is coming from the front and only try to compensate for a gain deviation at one particular frequency. Consequently, an urgent demand exists in the hearing aid industry for robust and efficient noise reduction algorithms.

Finally, it is noted that while current commercial hearing aids already provide some kind of noise reduction, the application of noise reduction techniques in commercial cochlear implants is still non-existing. However, thanks to the development of low-power DSPs, integration of multi-microphone noise reduction techniques is nowadays also a hot topic in the cochlear implant industry.

1.4.2 Feedback suppression

Electronic solutions for acoustic feedback were not really feasible until the advent of digital or at least analog programmable (e.g., Danavox DFS Genius) hearing aids. In the past, feedback was controlled mechanically, e.g., by reducing the vent size and ensuring a tight fit of the earmold, at the expense of listening comfort (e.g., the occlusion effect).

The first and probably still the most common electronic solutions for acoustic feedback in commercial hearing aids are the reduction of the high frequency gain and the use of notch filters [194, 269]. Typically, the feedback path provides less attenuation at high frequencies [90], where most hearing aid users have the largest hearing loss. As a result, acoustic feedback often first occurs in the high frequency range. Attenuating the high frequencies reduces the risk of acoustic feedback, but also compromises the audibility of the high frequencies. To minimize the detrimental effect of gain limitation, some hearing aid manufacturers use one or several notch filters, so that the gain is only reduced in narrow bands around the critical frequencies [194, 269].

The first application of adaptive feedback cancellation techniques in commercial hearing aids appeared in the Danavox BTE DFS Genius [46, 47] followed by the digital Danalogic hearing instruments of GN Danavox [222]: in the first design, the acoustic feedback path was identified based on an external probe signal that was inserted in the loudspeaker. With this feedback suppression system, the maximum stable gain before instability was found to be increased by about 10 dB. Nowadays, adaptive feedback cancellation techniques are still only encountered in the most advanced digital hearing aids (e.g., Oticon Adapto, Widex Senso Diva, GN ReSound Canta 4 and Canta 7). To preserve signal quality, the use of an external probe signal is limited or avoided as much as possible [194]. It is, e.g., only generated during fitting [143] or start-up [126, 127, 128, 129] in order to initialize filter settings or it is only used in a limited frequency band, e.g., at the low frequencies [183].

Most sound signals in every-day-life are spectrally colored, e.g., speech, music. To reduce the bias of the feedback canceller and the resulting distortion of the desired signal, the adaptation of the feedback canceller is controlled. Often, the adaptation speed is kept small [127, 132, 182, 183]. When feedback or a sudden change in feedback path is detected, the adaptation speed is temporarily increased (e.g., Widex Senso Diva [132, 143], Oticon Adapto [182, 183]). In order to further limit signal distortion, the adaptation of the feedback canceller is sometimes restricted to the high frequencies (e.g., Oticon Adapto) [183]. In other hearing aid systems, e.g., GN ReSound Canta 4 and Canta 7, the bias is reduced by means of constrained adaptation [125, 126, 127, 130, 194].

Constrained adaptation or a slow adaptation speed, however, only limit the bias of the feedback canceller to some extent, while considerably compromising

the tracking of changes in the feedback path. As a result, acoustic feedback still often occurs, e.g., when placing a telephone set close to the ear or when taking off the hearing aid during operation. In addition, the maximum amplification that is available with commercial hearing aids is still too limited to compensate for the hearing loss in each individual patient. Furthermore, to improve listening comfort and binaural hearing, there is a growing tendency in the hearing aid industry towards hearing aids with open fittings (e.g., ReSoundAir^{TM} of GN ReSound, OpenEar Acoustics^{TM} of Oticon), making the risk for acoustic feedback even higher [58]. Consequently, there is a high demand for adaptive feedback cancellation techniques that provide a reliable feedback path estimate without compromising the tracking performance or the signal quality.

1.5 Outline of the thesis and main contributions

This section starts with defining the objectives of the algorithms that are developed in this thesis. In addition, we provide a chapter by chapter overview of the thesis and summarize its contributions. References will be given to the publications that have been produced in the frame of this work.

1.5.1 Design objectives

In this thesis, we present noise reduction algorithms and feedback cancellation techniques for hearing aids. Since the noise reduction and the feedback cancellation techniques should both be able to deal with time-varying acoustic environments and/or non-stationary signals, we focus on **adaptive** solutions. In the algorithm design, special attention is paid to the specific requirements of hearing aids (such as low processing delay, the small-size, low computational complexity). E.g., in order not to degrade intelligibility due to asynchronism with lip reading [247], the delay in the signal processing path should be smaller than 20 ms.

Noise reduction

In the field of noise reduction, we concentrate on **multi-microphone** algorithms. In contrast to single-channel techniques, multi-microphone techniques are able to improve the local SNR at each frequency by exploiting the spatial characteristics of the speech and the noise signal and hence, have a larger potential benefit for speech intelligibility. In the evaluation and, where possible, also in the design of the algorithms, we will focus on improvement in **speech intelligibility** instead of improvement in global SNR.

In order to be aesthetically and commercially acceptable, the dimensions of the multi-microphone noise reduction systems will be kept small. Generally, we will consider multi-microphone systems with the dimensions of a commercial BTE. As already mentioned, robustness against imperfections becomes a crucial issue

when dealing with such **small-sized** multi-microphone systems. Therefore, we aim at algorithms that are **robust against imperfections** such as microphone mismatch. The sensitivity of several multi-microphone algorithms against different types of imperfections will be assessed in detail. In addition, the impact of the noise scenario as well as the number of microphones in a BTE on the robustness will be investigated.

Last, but not least, the computational complexity of the developed noise reduction algorithms will be reduced such that they are **feasible for implementation in commercial hearing aids and cochlear implants**.

Feedback cancellation

Standard continuous adaptation feedback cancellers suffer from a serious bias, when the input signal is spectrally colored. Classical solutions for reducing this bias trade off between desired signal cancellation on the one hand and tracking performance or sound quality on the other hand, so that typically a significant residual bias remains. An attractive alternative is to view the feedback path as a part of a closed-loop system and to apply closed-loop system identification theory. However, the solutions that have been proposed so far in this context, either require perfect knowledge of the (unknown and time-varying) desired signal spectrum [89, 88] or the use of a high-level probe signal [217].

In this thesis, we therefore aim at adaptive feedback cancellation techniques that provide **a reliable feedback path estimate for spectrally colored and highly time-varying sound signals** such as speech, **without compromising the sound quality and the tracking performance**. The use of non-linearities and/or a (high-level) external probe signal is avoided as much as possible. If a probe signal is inserted, its level is kept low with respect to the ordinary output to reduce audibility. In addition, in order to preserve intelligibility [247] and sound quality [88, 145], we allow only a small delay in the regular signal processing path.

In the design of the feedback cancellation algorithms, we concentrate on a single-microphone set-up. Since the first concern is to develop well-working adaptive feedback cancellers, the performance of the different feedback cancellers will be assessed using (sometimes expensive) recursive least squares type algorithms. The corresponding cheaper stochastic gradient implementations that are feasible for commercial hearing aids, will also be described.

1.5.2 Chapter by chapter outline and contributions

The thesis can be divided into two parts. The first part, i.e., from Chapter 2 to Chapter 5, concentrates on multi-microphone noise reduction, while the second part, i.e., Chapter 6 and Chapter 7, deals with feedback cancellation.

Multi-microphone noise reduction

Chapter 2 starts with an overview of several multi-microphone noise reduction techniques. Special attention is paid to the algorithms that are encountered in commercial hearing aids. Multi-microphone noise reduction techniques can be broadly classified in fixed beamforming, adaptive beamforming (such as the widely used Generalized Sidelobe Canceller) and multi-channel Wiener filtering. In particular, we will concentrate on an MWF technique [36] that, in contrast to fixed and adaptive beamformers, requires neither a priori knowledge about the desired signal model (such as the microphone characteristics and the position of the desired speaker) nor a calibration procedure. Although this MWF technique seems promising from a robustness point of view, it has not been considered yet for implementation in hearing aids because of its high computational cost. We show that a subband implementation results in a better intelligibility at a significantly lower computational cost compared to a fullband realization, enabling the use of the MWF technique in future hearing aids.

Publications related to this chapter are [231, 232, 235].

In **Chapter 3**, the MWF technique and the GSC are compared in terms of performance and robustness for small-sized microphone systems as used in commercial hearing aids. We present a theoretical framework to predict the impact of different types of model errors on the performance of the algorithms. It is shown that because of the high sensitivity of small-sized systems, the number of useful microphones in a commercial BTE is limited to three, provided that the noise reduction algorithm is sufficiently robust. Theoretical and experimental analysis reveal that robustness of the GSC is especially crucial in complicated noise scenarios and that microphone mismatch is particularly harmful to the GSC, even if the ANC is adapted during noise-only periods. As a consequence, a noise sensitivity constraint (e.g., a quadratic inequality constraint on the ANC) is required. The MWF on the other hand, is not degraded by errors in the signal model and reduces more noise than a robust GSC with a noise sensitivity constraint (NSC-GSC). However, the MWF is expected to be more sensitive to inaccurate estimation of the second order statistics caused by e.g., non-stationarities in the noise or speech detection errors. Experiments demonstrate that the MWF preserves its benefit over the GSC and the NSC-GSC if the spatial and *long-term* spectral characteristics of the noise vary relatively slowly in time (on the order of seconds) and the speech detection error rate is smaller than 20 %.

Publications related to this chapter are [230, 233, 234, 238, 236, 242].

In **Chapter 4**, we show that the GSC and the MWF can be integrated into one generalized noise reduction scheme, referred to as *spatially pre-processed speech distortion weighted MWF,* which consists of a fixed, spatial pre-processor and

an adaptive speech distortion weighted MWF. The new scheme encompasses the GSC and the MWF as extreme cases. In addition, it allows for in-between solutions such as the speech distortion regularized GSC. The SDR-GSC or more general the SP-SDW-MWF adds robustness to the GSC by taking speech distortion explicitly into account in the design criterion of the adaptive stage. Compared to the widely studied GSC with quadratic inequality constraint (QIC-GSC), the SP-SDW-MWF achieves a better noise reduction performance, for a given maximum speech distortion level.

Publications related to this chapter are [237, 239].

In **Chapter 5**, we develop a low-cost, stochastic gradient implementation of the SP-SDW-MWF that is feasible for implementation in commercial hearing aids. To speed up convergence and reduce computational complexity, the algorithm is implemented in the frequency-domain. Experimental results with a three-microphone BTE show that the proposed frequency-domain stochastic gradient algorithm preserves the benefit of the exact SP-SDW-MWF over the QIC-GSC, while its computational cost is comparable to the least mean squares (LMS) based scaled projection algorithm [29] for implementing the QIC-GSC.

Publications related to this chapter are [41, 42, 43, 240, 241].

Adaptive feedback cancellation

Chapter 6 gives an overview of the literature on feedback suppression and, in particular, adaptive feedback cancellation. We show that because of the presence of a closed signal loop, the standard adaptive filtering techniques for open-loop systems (as, e.g., used in echo cancellation) fail to provide a reliable feedback path estimate if the desired signal is spectrally colored. A promising solution to avoid this model error or bias is to apply the identification procedures for closed-loop systems [59, 60]. Chapter 6 discusses the basic concepts of closed-loop system identification as well as some recent applications to feedback cancellation. However, the solutions that have been proposed so far in this context either require perfect knowledge of the (unknown and time-varying) desired signal spectrum [88, 89] or the use of a high-level probe signal [217] in order to obtain an unbiased feedback path estimate.

In **Chapter 7**, we propose adaptive feedback cancellation techniques that are based on direct closed-loop identification of the feedback path as well as the (autoregressive) modelling of the desired signal, with the aim of providing an unbiased feedback path estimate for unknown and time-varying desired signals. In general, both models are not simultaneously identifiable in the closed-loop system at hand [89]. We show that under certain conditions, e.g., if a delay is inserted in the forward path, identification of both models is indeed possible. Simulation results demonstrate that the proposed techniques outperform the standard continuous adaptation feedback cancellation algorithm if the conditions for identifiability are satisfied.

Publications related to this chapter are [227, 228, 229, 243].

Finally, **Chapter 8** provides the overall conclusions of the thesis as well as some suggestions for further research.

1.6 Conclusion

Due to the aging of the world's population and the increased exposure to noise in daily life, the number of people suffering from a cochlear hearing loss is expected to further increase in the future. With the advent of low-power DSPs, an urgent demand exists for efficient and effective algorithms that provide benefits for the individual patient.

In Section 1.1, the implications of cochlear hearing loss on the ability to perceive sound have been discussed. The major signal processing algorithms that are required in a hearing aid or cochlear implant to compensate for the resulting negative effects have been identified in Section 1.2. People with a cochlear hearing loss do not only suffer from an increased hearing threshold but also from an increased difficulty understanding speech in noise. As a result, in addition to amplifying the acoustic sound signal, the hearing aid should perform more complex signal processing such as noise reduction. An overview of the literature on noise reduction algorithm has been given in Section 1.2.

The potential benefit of a hearing aid may, however, be seriously compromised due to acoustic feedback, so that also feedback suppression or feedback cancellation techniques are called for in a hearing aid. In Section 1.3, we have described the acoustic feedback problem in hearing aids and we have summarized the literature on feedback suppression.

In Section 1.4, we have indicated which noise reduction and feedback suppression techniques are currently being adopted in commercial hearing aids. As for noise reduction, there is a high need for low-cost multi-microphone algorithms that are robust to imperfections encountered in real-life use. In the field of acoustic feedback suppression, adaptive feedback cancellation techniques are called for that provide a reliable feedback path estimate for spectrally colored input signals without compromising the tracking performance or the signal quality.

Finally, the outline of the thesis and its main contributions have been presented in Section 1.5.

Chapter 2

Multi-microphone noise reduction techniques

NOISE reduction algorithms are crucial in hearing aids to improve the intelligibility of speech in background noise. Multi-microphone systems exploit spatial in addition to temporal and spectral information of the desired signal and the noise signal, and are thus preferred to single-channel procedures (such as spectral subtraction).

This chapter provides the basic concepts and algorithms for multi-microphone noise reduction necessary to understand the material presented in the forthcoming chapters. Several multi-microphone noise reduction techniques, both fixed and adaptive, are discussed. Special attention will also be paid to the algorithms that are encountered in commercial hearing aids. In addition, the set-up and appropriate performance measures for evaluating noise reduction algorithms for hearing aid applications are described. The major goal of a noise reduction algorithm in a hearing aid application is to improve speech intelligibility and not fullband SNR. In this chapter, we show that this aim is better achieved by performing the noise reduction algorithms in subbands instead of in fullband. Furthermore, a subband implementation results in a significant reduction in computational complexity, enabling the implementation of expensive matrix-based algorithms in future hearing aids.

This chapter is organized as follows. In Section 2.1, we describe the notation, the signal model and the general set-up of a multi-microphone noise reduction algorithm, both in the time domain and the frequency domain. In addition, we define the performance measures and the acoustical configuration of the real-data experiments that are used in this thesis to assess the performance of a noise reduction algorithm for hearing aid applications. Given the small size of hearing

aids, not only the improvement in speech intelligibility and the computational complexity should be considered, but also the sensitivity or robustness of the algorithm to signal model errors such as microphone mismatch.

Multi-microphone noise reduction algorithms can be classified into fixed and adaptive techniques. Section 2.2 describes several fixed beamforming techniques, while Section 2.3, Section 2.4 and Section 2.5 concentrate on adaptive multi-microphone techniques. The latter can be classified in adaptive beamforming techniques (Section 2.3), which are often implemented using a Generalized Sidelobe Canceller structure, and multi-channel Wiener filtering (MWF) techniques (Section 2.4). In particular, we will concentrate on an MWF technique [36] that does neither require a priori knowledge about the desired signal model nor calibration. Although this MWF technique seems promising from a robustness point of view, it has not been considered yet for implementation in hearing aids because of its high computational cost. In Section 2.5, we propose a subband implementation of this MWF that achieves a better intelligibility at a significantly lower computational cost than a fullband implementation, making it suitable for future hearing aids.

2.1 Problem statement and notation

This section presents a mathematical formulation of multi-microphone noise reduction both in the time-domain and the frequency-domain. In addition, we define the notation and performance measures that are used throughout this thesis to evaluate a noise reduction algorithm and describe the set-up of the acoustical configuration that has been used for the real-data experiments.

2.1.1 Signal model

In this section, we introduce the signal model that is considered in the development of the noise reduction algorithms. For the time being, we assume that no acoustic feedback is present. In Chapter 6, the signal model will be extended to also include acoustic feedback.

Consider a hearing aid with M microphones in a noisy acoustic environment, as depicted in Figure 1.2 and Figure 2.1. Suppose that the acoustic environment consists of one or more background noise sources and suppose that $s_d[k]$ is the clean speech signal, uttered by the person to whom the hearing aid user is listening. The acoustic waves, coming from the different signal sources, propagate through the air and are reflected by the walls, the ceiling, the floor and objects present in the room, before reaching the microphone(s) (cf. Figure 2.1). The shortest propagation path between the signal source and the microphone is referred to as the *direct path*, and the combined effect of multiple reflections

is called *reverberation*. The entire acoustic propagation from a source signal to a microphone is often characterized by a linear impulse response, called the *acoustic impulse response*.

Each microphone signal $x_l[k], l = 1, ..., M$, at time k, can be decomposed into a speech component $x_l^s[k]$ and an additive noise component $x_l^n[k]$ as:

$$\boxed{x_l[k] = h_l[k] \otimes s_d[k] + x_l^n[k] = x_l^s[k] + x_l^n[k],} \tag{2.1}$$

where $h_l[k]$ represents the acoustic impulse response from the desired speaker to the l-th microphone and where $\otimes$ denotes the convolution operator. The acoustic impulse response $h_l[k]$ is generally modelled as an FIR filter $\mathbf{h}_l[k]$ of length L_H

$$\mathbf{h}_l[k] = \begin{bmatrix} h_{l,0}[k] & h_{l,1}[k] & \cdots & h_{l,L_H-1} \end{bmatrix}^T, \tag{2.2}$$

hence,

$$h_l[k] \otimes s_d[k] = \sum_{j=0}^{L_H-1} h_{l,j}[k] s_d[k-j]. \tag{2.3}$$

During periods of speech + noise, the microphone signals $x_l[k]$, $l = 1, \cdots, M$ contain speech + noise contributions. During periods of noise-only, the microphone signals $x_l[k]$ only consist of a noise component $x_l^n[k]$, i.e., $x_l[k] = x_l^n[k]$.

The additive noise component $x_l^n[k]$ is assumed to be uncorrelated with the clean speech signal $s_d[k]$ and hence, with the speech component $x_l^s[k]$. The noise component $x_l^n[k]$ at the l-th microphone consists of two parts, i.e., external noise $x_l^e[k]$ and internal noise $x_l^i[k]$:

$$x_l^n[k] = x_l^e[k] + x_l^i[k] \tag{2.4}$$

- The external noise $x_l^e[k]$ originates from the background noise sources. These noise sources can be *localized* (e.g., computer fan, car noise, other people) or *diffuse*, i.e., coming from all directions (e.g., wind). In case of localized noise sources, $x_l^e[k]$ equals

  $$x_l^e[k] = \sum_j h_{jl}[k] \otimes n_j^e[k],$$

  where $h_{jl}[k]$ represents the acoustic impulse response from the j-th external noise source n_j^e to the l-th microphone. Some external noise sources are spectrally stationary or have slowly time-varying spectral properties (e.g., car noise), while others are highly spectrally non-stationary (e.g., multi-talker babble, music). Often, the external noise is spectrally colored (e.g., car noise, speech, music).

- The internal noise $x_l^i[k]$ consists of sensor noise as well as quantization noise of the analog-to-digital converter (ADC). Since the internal noise

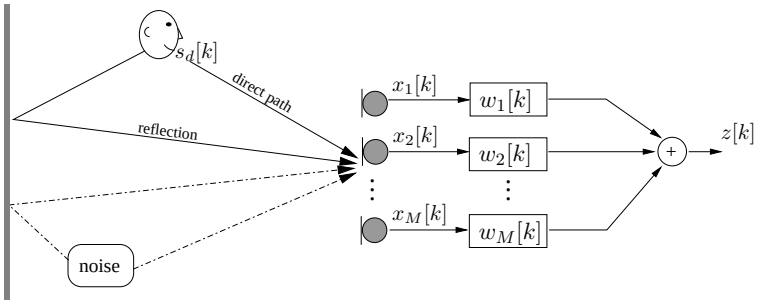


Figure 2.1: Multi-microphone noise reduction algorithm.

is generally uncorrelated over the different microphones, it is modelled as spatially white noise. The internal noise level depends on the quality of the microphones and the accuracy of the quantization. For omnidirectional hearing aid microphones, typical sensor noise levels equal about 30 dB SPL.

Remark 2.1 The signal model (2.1) assumes that the microphones are perfectly omnidirectional (i.e., have the same sensitivity for all directions in space) and have a flat frequency response equal to 1. If this is not the case, the microphone characteristics (directivity pattern, spectral characteristics) can be incorporated into the acoustic impulse responses of the speech and the noise sources. □

2.1.2 Multi-microphone noise reduction

Figure 2.1 depicts the general set-up of a multi-microphone noise reduction algorithm. The microphone signals $x_l[k]$ are filtered by (adaptive or fixed) filters $w_l[k]$ and combined in order to obtain an enhanced speech signal $z[k]$:

$$z[k] = \sum_{l=1}^{M} w_l[k] \otimes x_l[l]. \tag{2.5}$$

In general, the filters $w_l[k]$ are FIR filters $\mathbf{w}_l[k]$ with length L, i.e.,

$$\mathbf{w}_l[k] = \begin{bmatrix} w_{l,0}[k] & w_{l,1}[k] & \cdots & w_{l,L-1}[k] \end{bmatrix}^T. \tag{2.6}$$

Let $\mathbf{w}[k] \in \mathbb{R}^{ML\times 1}$ be defined as the stacked filter vector

$$\mathbf{w}[k] = \begin{bmatrix} \mathbf{w}_1^T[k] & \mathbf{w}_2^T[k] & \cdots & \mathbf{w}_M^T[k] \end{bmatrix}^T. \tag{2.7}$$

Similarly, define the ML-dimensional stacked data vector $\mathbf{x}[k]$ as,

$$\mathbf{x}[k] = \left[\begin{array}{cccc} \mathbf{x}_1^T[k] & \mathbf{x}_2^T[k] & \cdots & \mathbf{x}_M^T[k] \end{array} \right]^T, \tag{2.8}$$

$$\mathbf{x}_l[k] = \left[\begin{array}{cccc} x_l[k] & x_l[k-1] & \cdots & x_l[k-L+1] \end{array} \right]^T. \tag{2.9}$$

Then, the output signal $z[k]$ at time k can be expressed as

$$\boxed{z[k] = \sum_{l=1}^{M} \mathbf{w}_l^T[k]\mathbf{x}_l[k] = \mathbf{w}^T[k]\mathbf{x}[k].} \tag{2.10}$$

Using (2.1),(2.5), $z[k]$ can be decomposed into a speech component $z^s[k]$ and a noise component $z^n[k]$:

$$z[k] = z^s[k] + z^n[k] \tag{2.11}$$

$$z^s[k] = \sum_{l=1}^{M} w_l[k] \otimes x_l^s[k] = \left\{ \sum_{l=1}^{M} w_l[k] \otimes h_l[k] \right\} \otimes s_d[k] \tag{2.12}$$

$$z^n[k] = \sum_{l=1}^{M} w_l[k] \otimes x_l^n[k] \tag{2.13}$$

The goal of a noise reduction algorithm is to minimize the energy of the residual noise $z^n[k]$ as much as possible without severely distorting the clean speech signal $s_d[k]$ or the speech component $x_l^s[k]$ in the l-th microphone signal.

2.1.3 Equivalent representation in the frequency-domain

Multi-microphone noise reduction techniques are often designed and analyzed in the frequency-domain, assuming stationary signals and infinite filter lengths.

Notation

The frequency-domain representation $X(f)$ of $x[k]$ is obtained by the discrete-time Fourier transform (DTFT) as

$$X(f) = \mathcal{F}\{x[k]\} = \sum_{k=-\infty}^{k=\infty} x[k] e^{-j2\pi \frac{f}{f_s} k}, \tag{2.14}$$

with f_s the sampling frequency and $-\frac{f_s}{2} < f \leq \frac{f_s}{2}$. The power spectral density (PSD) of the signal $x[k]$ is defined as

$$P_X(f) = \mathcal{E}\{|X(f)|^2\}, \tag{2.15}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator. The cross-PSD between signals $x[k]$ and $y[k]$ equals:

$$P_{XY}(f) = \mathcal{E}\{X(f)Y^*(f)\}. \tag{2.16}$$

Where needed, the superscripts s and n are used to refer to the contribution of the speech and noise signal only: e.g., the PSD of the speech component $x_l^s[k]$ is denoted as $P_{X_l}^s(f)$, the PSD of the noise component $x_l^n(f)$ as $P_{X_l}^n(f)$. The superscripts e and i are used to refer to the external and internal noise contribution, respectively. We assume that the internal-to-external noise ratio $P_{X_l}^i(f)/P_{X_l}^e(f)$ is the same at all microphones l and equals $\beta(f) = P_X^i(f)/P_X^e(f)$. For external noise levels between 50 dB SPL and 70 dB SPL, typical values for β are between -20 dB and -40 dB [14].

Frequency-domain representation

Using (2.1), the microphone signals $X_l(f)$, $l = 1, \cdots, M$ can be written as:

$$X_l(f) = X_l^s(f) + X_l^n(f) = H_l(f)S_d(f) + X_l^n(f), \tag{2.17}$$

where $H_l(f)$ is the acoustic transfer function between the speech source $S_d(f)$ and the l-th microphone.

Let $\mathbf{X}(f) \in \mathbb{C}^{M\times 1}$ be defined as the stacked vector

$$\mathbf{X}(f) = \begin{bmatrix} X_1(f) & X_2(f) & \cdots & X_M(f) \end{bmatrix}^T, \tag{2.18}$$

and define $\mathbf{H}(f)$ as

$$\mathbf{H}(f) = \begin{bmatrix} H_1(f) & H_2(f) & \cdots & H_M(f) \end{bmatrix}^T. \tag{2.19}$$

Then, (2.17) can be written as

$$\mathbf{X}(f) = \mathbf{X}^s(f) + \mathbf{X}^n(f) = \mathbf{H}(f)S_d(f) + \mathbf{X}^n(f), \tag{2.20}$$

with $\mathbf{X}^s(f)$ and $\mathbf{X}^n(f)$ defined similarly as in (2.18).

To simplify the formulation of the noise reduction algorithms, we define

$$\mathbf{W}(f) = \begin{bmatrix} W_1(f) & W_2(f) & \cdots & W_M(f) \end{bmatrix}^H, \tag{2.21}$$

with $W_l(f) = \sum_{m=0}^{L-1} w_{l,m} e^{-j2\pi \frac{f}{f_s} m}$. The output $Z(f)$ of the multi-channel noise reduction algorithm can then be expressed as

$$Z(f) \;=\; \mathbf{W}^H(f)\mathbf{X}(f) = \underbrace{\mathbf{W}^H(f)\,\mathbf{H}(f)S_d(f)}_{Z^s(f)} + \underbrace{\mathbf{W}^H(f)\mathbf{X}^n(f)}_{Z^n(f)}. \tag{2.22}$$

Complex coherence function

The filter coefficients $\mathbf{W}(f)$ and the performance of a noise reduction algorithm will often be described as a function of the complex coherence functions between the different microphone signals.

The *complex coherence* $\Gamma_{kl}(f)$ between the k-th and l-th microphone signals with PSD respectively $P_{X_k}(f)$ and $P_{X_l}(f)$ and cross-PSD $P_{X_kX_l}(f)$, is defined as:

$$\Gamma_{kl}(f) = \frac{P_{X_kX_l}(f)}{\sqrt{P_{X_k}(f)P_{X_l}(f)}}. \tag{2.23}$$

For a *homogeneous sound field*, i.e., $P_{X_l}(f) = P_X(f)$, $l = 1, \cdots, M$, the correlation matrix $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}$ of the microphone signals equals[1]

$$\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} = P_X(f)\mathbf{\Gamma}(f), \tag{2.24}$$

where $\mathbf{\Gamma}(f)$ is the *coherence matrix* defined as

$$\mathbf{\Gamma}(f) = \begin{bmatrix} 1 & \Gamma_{12}(f) & \cdots & \Gamma_{1M}(f) \\ \Gamma_{21}(f) & 1 & \cdots & \Gamma_{2M}(f) \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma_{M1}(f) & \Gamma_{M2}(f) & \cdots & 1 \end{bmatrix}. \tag{2.25}$$

Similarly, we define the noise and speech correlation matrix as

$$\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} = P_X^n(f)\mathbf{\Gamma}^n(f), \tag{2.26}$$

$$\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} = P_X^s(f)\mathbf{\Gamma}^s(f). \tag{2.27}$$

When the microphone configuration and the sound field are known, $\Gamma_{kl}(f)$ can be computed mathematically (cf. (2.35) and (2.36) below). For a *single speech source*,

$$\mathbf{\Gamma}^s(f) = \frac{P_{S_d}(f)}{P_X^s(f)}\mathbf{H}(f)\mathbf{H}^H(f). \tag{2.28}$$

The coherence of *spatially uncorrelated noise* (e.g., internal noise) equals

$$\Gamma_{kl} = \delta[k-l], \tag{2.29}$$

where $\delta[k]$ is the Dirac impulse. The coherence function $\Gamma_{kl}(f)$ for a set of L *uncorrelated sound sources* with individual coherence functions $\Gamma_{kl}^j(f)$ and PSD $P_X^j(f)$ at the M microphones, $j = 1, \cdots, L$ is given by

$$\Gamma_{kl}(f) = \frac{\sum_{j=1}^{L} \alpha_j(f)\Gamma_{kl}^j(f)}{\sum_{j=1}^{L} \alpha_j(f)}, \tag{2.30}$$

[1] Spectral differences between the microphone signals can be included in (2.24) by modifying the coherence matrix $\mathbf{\Gamma}$(f) in (2.25), as will be shown in Section 3.4.1.

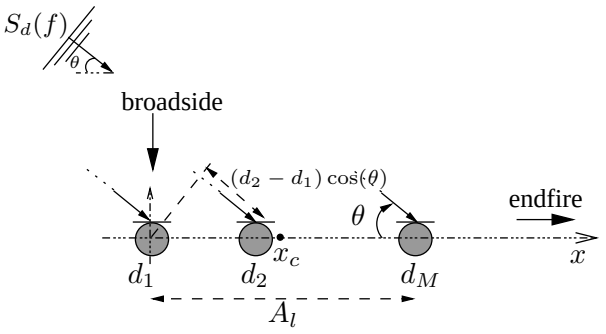


Figure 2.2: Linear microphone array in free field.

with $\alpha_j(f) = \frac{P_X^j(f)}{P_X^1(f)}$. E.g., in the presence of sensor noise, the noise coherence $\Gamma_{kl}^n(f)$ is given by [14]

$$\Gamma_{kl}^n(f) = \frac{\Gamma_{kl}^e(f)}{1+\beta(f)} \qquad \text{for } k \neq l, \tag{2.31}$$

since $\Gamma_{kl}^i(f) = \delta[k-l]$, with $\Gamma_{kl}^e(f)$ the coherence function of the external noise and with $\beta(f)$ the internal-to-external noise ratio at the different microphones.

Special case: linear microphone array in the free field

For simplicity, we often consider a linear M-microphone array in the *free field* as depicted in **Figure 2.2**, in the mathematical algorithmic descriptions. The array length is denoted by A_l and the distance between the l-th microphone and the center x_c of the array by d_l. In this thesis, we concentrate on small-sized microphone systems (i.e., with array lengths A_l of at most 6 cm) as used in commercial hearing aids. For such small-sized arrays, the distance of the sources to the array is generally larger than the minimum distance $\frac{A_l^2 f_s}{c}$ m (with c the speed of sound in air (i.e., $c \approx 340$ m/s) and $f_s = 16$ kHz the sampling frequency) for which far-field propagation may be assumed [159]. Hence, we may assume that the speech and noise sources are in the *far-field* region of the microphone array, which implies planar wave propagation.

Under these *far and free field assumptions*, mathematical expressions can be derived for the acoustic transfer functions $H_l(f)$ of the desired speech source $S_d(f)$ as well as for the coherence matrix $\mathbf{\Gamma}^n(f)$ of some particular noise fields. Suppose that the desired speech source $S_d(f)$ is located at an angle θ with respect to the microphone array. If the angle θ between the desired source and the microphone array equals $\theta = 90°$ or $\theta = 270°$, the array is called *broadside*, whereas if the angle $\theta = 0°$ or $\theta = 180°$, we have an *endfire* array. In this thesis,

we will consider an endfire configuration with $\theta = 0°$, as this approximates the configuration of a commercial BTE with the desired speaker positioned in front of the hearing aid user.

Assuming *identical, perfect omnidirectional microphones*, the acoustic transfer function $H_l(f)$ from $S_d(f)$ to the l-th microphone signal equals

$$\boxed{H_l(f) = \underbrace{e^{-j2\pi f\tau_1}}_{H_1(f)} e^{-j2\pi f \frac{(d_l - d_1)}{c}\cos\theta},} \tag{2.32}$$

where τ_1 denotes the propagation time to the first microphone. Hence, the different microphone signals $X_l(f)$ are delayed versions of each other. The vector of relative propagation delays for a sound incident from a particular direction θ is called the *steering vector* $\mathbf{d}(f,\theta)$:

$$\mathbf{d}(f,\theta) = \left[\begin{array}{cccc} 1 & e^{-j2\pi f \frac{d_2 - d_1}{c}\cos\theta} & \dots & e^{-j2\pi f \frac{d_M - d_1}{c}\cos\theta} \end{array}\right]. \tag{2.33}$$

Using (2.22), (2.32) and (2.33), the output $Z(f)$ of the multi-microphone noise reduction algorithm equals

$$Z(f) = \underbrace{S_d(f)e^{-j2\pi f\tau_1}}_{X_1(f)} \mathbf{W}^H(f)\mathbf{d}(f,\theta). \tag{2.34}$$

Hence, the inner product $\mathbf{W}^H(f)\mathbf{d}(f,\theta)$ completely characterizes the gain of a multi-microphone noise reduction algorithm to sound incident from a particular direction θ and is referred to as the spatial directivity pattern $H(f,\theta)$ (cf. Section 2.2.1).

Under far and free field assumptions, also simple expressions can be derived for the coherence function $\Gamma_{kl}(f)$ of a localized and diffuse noise field. These expressions, which are typically used in the design of fixed superdirective beamformers [9, 10, 28, 123, 245] (cf. Section 2.2.5), will be used in Chapter 3 to theoretically predict the performance of different adaptive noise reduction algorithms. From (2.28) and (2.32), it follows that the coherence $\Gamma_{kl}(f)$ of a *localized sound source* at an angle θ in free and far field equals

$$\boxed{\Gamma_{kl}(f) = e^{j2\pi f\cos\theta(d_l - d_k)/c}.} \tag{2.35}$$

The coherence Γ_{kl} of a *diffuse sound field* equals

$$\boxed{\Gamma_{kl}(f) = \frac{\sin(2\pi f(d_l - d_k)/c)}{2\pi f(d_l - d_k)/c} = \mathrm{sinc}(2\pi f(d_l - d_k)/c)} \tag{2.36}$$

with $d_l - d_k$ the interspacing between microphone l and k [13, 111]. The diffuse sound field is a good approximation of sound fields in strongly reverberant rooms, if the sound sources in the room have no direct path to the microphone array, i.e. for a low direct-to-reverberant ratio.

2.1.4 Performance measures

To evaluate the performance of the multi-microphone noise reduction algorithms, we consider the noise reduction performance, the distortion of the desired speech and the noise sensitivity. The latter is a measure for the sensitivity of the algorithm to random errors encountered in real-life applications (e.g., deviations from idealized assumptions (e.g., microphone mismatch) and processing errors) [28, 29, 198, 245]. First, we define the performance measures in the frequency-domain. Subsequently, we convert them to broadband performance measures. Since the major goal of a noise reduction algorithm for hearing aids is to improve intelligibility, appropriate broadband measures are intelligibility weighted averages of the performance in different frequency bins.

Power transfer function (PTF)

In the definition of the performance measures, we refer to the power transfer functions (PTF) from the algorithm's input to the algorithm's output $Z(f)$ as $G(f)$. Unless stated otherwise, we consider the first microphone signal as the input of the algorithm[2], hence,

$$\boxed{G(f) = P_Z(f)/P_{X_1}(f).} \tag{2.37}$$

Similarly, $G^s(f)$, $G^n(f)$ and $G^i(f)$ refer to the PTF of the speech component, the PTF of the noise component and the PTF of spatially white noise (e.g., internal noise) from the algorithm's input to the output $Z(f)$. For *a homogeneous sound field*, the PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ can be expressed as a function of the filter coefficients $\mathbf{W}(f)$ and the coherence matrices $\mathbf{\Gamma}^s(f)$ and $\mathbf{\Gamma}^n(f)$ using (2.22) and $\mathbf{\Gamma}^i(f) = \mathbf{I}_M$:

$$G^s(f) = P_Z^s(f)/P_X^s(f) = \mathbf{W}^H(f)\mathbf{\Gamma}^s(f)\mathbf{W}(f), \tag{2.38}$$

$$G^n(f) = P_Z^n(f)/P_X^n(f) = \mathbf{W}^H(f)\mathbf{\Gamma}^n(f)\mathbf{W}(f), \tag{2.39}$$

$$G^i(f) = P_Z^i(f)/P_X^i(f) = \mathbf{W}^H(f)\mathbf{W}(f). \tag{2.40}$$

Noise reduction performance and speech distortion

The *noise reduction performance* can be analyzed by considering the PTF $G^n(f)$. A better measure is to consider the SNR improvement $\Delta\mathrm{SNR}(f)$ of the algorithm:

$$\boxed{\Delta\mathrm{SNR}(f) = \mathrm{SNR}_{\mathrm{out}}(f) - \mathrm{SNR}_{\mathrm{in}}(f)} \tag{2.41}$$

[2]The PTFs are typically computed with respect to the first microphone because the multi-channel Wiener filtering technique described in Section 2.4 and Section 2.5 provides an estimate of the speech component in the first microphone signal.

where $\mathrm{SNR}_{\mathrm{in}}(f)$ and $\mathrm{SNR}_{\mathrm{out}}(f)$ are the input SNR (i.e., typically at the first microphone) and output SNR at frequency f:

$$\mathrm{SNR}_{\mathrm{in}}(f) = 10\log_{10}\left(P^s_{X_1}(f)/P^n_{X_1}(f)\right) \quad (2.42)$$

$$\mathrm{SNR}_{\mathrm{out}}(f) = 10\log_{10}\left(P^s_{Z}(f)/P^n_{Z}(f)\right) \quad (2.43)$$

To analyze the *speech distortion*, we consider the PTF $G^s(f)$ of the speech component from the input (i.e., at the first microphone) to the output $Z(f)$ of the algorithm. Ideally, $G^s(f)$ equals 0 dB.

Noise sensitivity

When designing a noise reduction algorithm, assumptions often have to be made about the desired signal model (e.g., the position of the desired source w.r.t. the microphone array, identical microphones with ideal directivity properties, absence of reverberation). These assumptions are, however, seldom satisfied exactly in practice. Microphones are rarely matched in gain and phase and their characteristics may even drift over time [114]. The spatial characteristics of the microphones are affected by the presence of the head and the mounting on the hearing aid. Movements by the desired speaker as well as head movements by the hearing aid user and array misalignment influence the relative position of the desired source w.r.t. the microphone array.

Hence, in addition to offering good noise suppression and low speech distortion, the algorithm should be robust, i.e., insensitive to small errors in these assumptions. The same holds for processing errors (i.e., round off errors due to finite processing accuracy of the DSP) [245]. The sensitivity to errors in the assumed signal model and processing errors is often quantified by the *noise sensitivity* $\Psi(f)$, which is defined as the ratio of the spatially white noise gain $G^i(f)$ to the gain of the desired signal $G^s(f)$ [28, 29, 198, 245]:

$$\boxed{\Psi(f) = \frac{G^i(f)}{G^s(f)}.} \quad (2.44)$$

The higher $\Psi(f)$, the more sensitive the design will be to errors in the assumed signal model. In hearing aids, the array length A_l is limited because of aesthetical and practical reasons. In [198, 245], it is shown that for low $f \cdot d$, i.e. $f \cdot d \ll \frac{c}{2}$ (with d the microphone interspacing), noise reduction is obtained at the expense of an increased noise sensitivity. The increased noise sensitivity indicates that the robustness of multi-microphone noise reduction algorithms may be crucial for practical use in hearing aids.

Remark 2.2 In addition to the sensitivity of the algorithm to signal model errors, the noise sensitivity indicates how much the internal noise of the system (e.g., sensor noise) gets amplified w.r.t. the desired signal. □

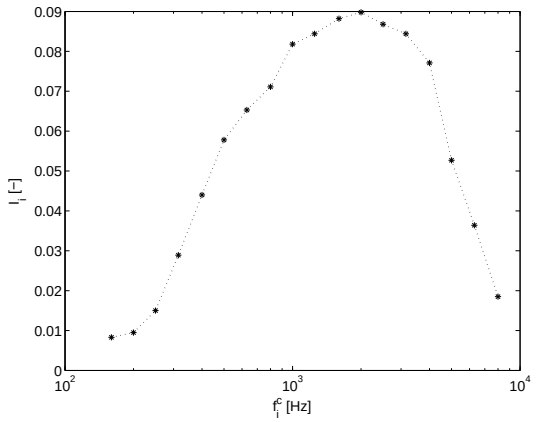


Figure 2.3: Band importance function I_i for average speech as a function of the center frequency f_i^c of the one-third octave band.

Intelligibility weighted performance measures

To compare the performance of different noise reduction algorithms for broadband signals such as speech, broadband performance measures are more suitable. Since the major goal of a noise reduction algorithm in hearing aid applications is to improve intelligibility, appropriate broadband measures are intelligibility weighted averages of the performance (in dB) in different one third octave bands [73, 245]. The weight I_i is referred to as the band importance function and expresses the importance of the i-th one-third octave band with center frequency f_i^c for intelligibility. The weights I_i and the center frequencies f_i^c of the one-third octave bands are defined in [1, Table 3] and depicted in **Figure 2.3** for average speech: frequencies between 1 kHz and 3 kHz are most important while frequencies above 5 kHz are less important for speech intelligibility.

The improvement $\Delta\text{SNR}_{\text{intellig}}$ in *intelligibility weighted SNR* [73] is defined as

$$\boxed{\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i \cdot (\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}),} \tag{2.45}$$

where $\text{SNR}_{i,\text{out}}$ is the output SNR (in dB) and $\text{SNR}_{i,\text{in}}$ is the input SNR (in dB) in the i-th one-third octave band $[2^{-\frac{1}{6}} f_i^c,\, 2^{\frac{1}{6}} f_i^c]$:

$$\text{SNR}_{i,\text{out}} = 10\log_{10}\left(\int_{2^{-\frac{1}{6}} f_i^c}^{2^{\frac{1}{6}} f_i^c} P_Z^s(f)df \Big/ \int_{2^{-\frac{1}{6}} f_i^c}^{2^{\frac{1}{6}} f_i^c} P_Z^n(f)df\right) \tag{2.46}$$

$$\mathrm{SNR}_{i,\mathrm{in}} = 10\log_{10}\left(\int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_{X_1}^s(f)df \Big/ \int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} P_{X_1}^n(f)df\right). \quad (2.47)$$

Similarly, we define an *intelligibility weighted spectral distortion measure* (in dB), $\mathrm{SD}_{\mathrm{intellig}}$, of the desired speech signal as

$$\boxed{\mathrm{SD}_{\mathrm{intellig}} = \sum_i I_i \cdot \mathrm{SD}_i,} \quad (2.48)$$

with SD_i the average spectral distortion (in dB) in the i-the one-third octave band, computed as

$$\mathrm{SD}_i = \int_{2^{-\frac{1}{6}}f_i^c}^{2^{\frac{1}{6}}f_i^c} |10\log_{10} G^s(f)|\, df \Big/ \left[\left(2^{\frac{1}{6}} - 2^{-\frac{1}{6}}\right) f_i^c\right]. \quad (2.49)$$

2.1.5 Acoustical configuration for real-data experiments

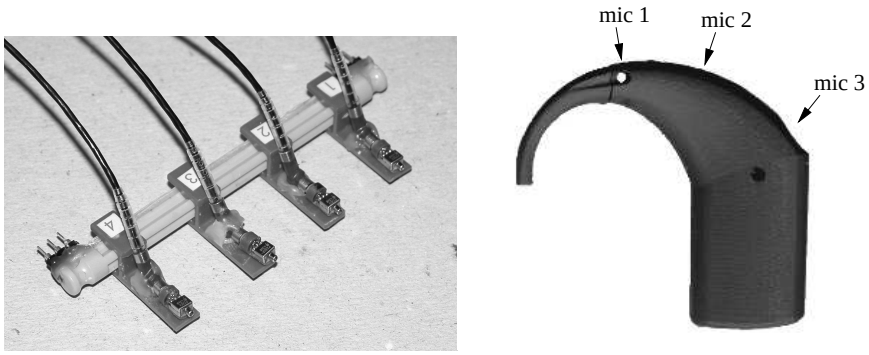
This section describes the set-up of the acoustical configuration that has been used for the real-data experiments in this thesis.

Most experiments have been carried out in an office room with a reverberation time T_{60} of about 700 ms (measured for a stationary speech-weighted noise signal). The reverberation time T_{60} is defined as the time needed for a sound pressure wave to decay to -60 dB of its original level and is a measure for the amount of reverberation in a room: the longer the reverberation time, the more reverberant the room.

Microphone configurations

In the experiments, two microphone configurations have been used. A first experimental configuration, depicted in **Figure 2.4(a)** consists of a *linear endfire array* with 4 omnidirectional microphones (KNOWLES EM-4368) of which the interspacing can be varied between 0.006 m and 0.02 m: typically, a uniform array with microphone interspacing $d = 0.01$ m and $d = 0.02$ m has been used. A second, more realistic configuration, shown in **Figure 2.4(b)** consists of a *BTE hearing aid with* 3 *omnidirectional microphones* (KNOWLES FG-3452). The interspacing $d_2 - d_1$ between the first and the second microphone is about 0.01 m and the interspacing $d_3 - d_2$ between the second and third microphone is about 0.015 m.

The microphone signals are amplified by Larson-Davis 2200-C amplifiers. The signals are recorded by LynxOne sound cards at a sampling frequency $f_s = 16$ kHz. In the experiments, the microphones have been calibrated by means of recordings of a speech-weighted noise signal positioned at 0° in an anechoic



(a) (b)

Figure 2.4: Microphone configurations; (a) Linear endfire array; (b) 3-microphone BTE.

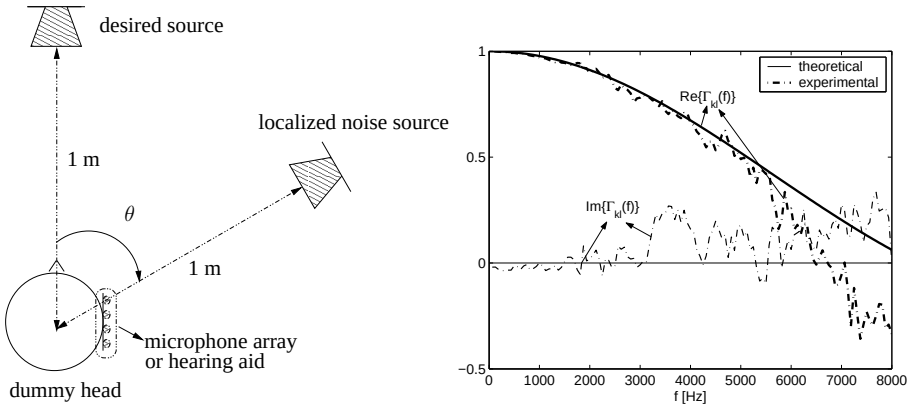
room. In order to take into account the head shadow effect in the calibration, the anechoic recordings were made while the microphone array was mounted on a dummy head (i.e., a manikin head) (cf. Section 2.2.4).

Noise configuration

In the experiments, the microphone array or BTE was mounted on the right ear of a dummy head, as illustrated in **Figure 2.5(a)** (endfire configuration). The desired source is positioned at a distance of 1 m in front of the head ($\theta = 0°$). In the experiments, the noise field consists of one or more localized noise sources or a diffuse noise field:

- *Localized noise sources* are positioned at an angle θ w.r.t. the desired source and at a distance of 1 m w.r.t. the head (cf. Figure 2.5). The noise signals are either stationary noise with the average long-term spectrum of speech, referred to as stationary speech-weighted noise, or spectrally non-stationary, multi-talker babble noise [5]. The desired source signal usually consists of sentences spoken by a male speaker from the HINT-database [185]. In some experiments, stationary speech-weighted noise (i.e., the unmodulated noise signal with male weighted idealized speech spectrum (normal effort) on track 1 of [106] or the speech weighted noise signal from the HINT-database [185]) has been used to predict the average performance based on directivity only.

- The *diffuse sound field* has been generated as proposed in [264], i.e., uncorrelated, stationary speech-weighted noise signals are sent out through



(a) (b)

Figure 2.5: (a) Acoustical configuration of the set-up. (b) Experimental and theoretical coherence function $\Gamma_{kl}(f)$ of two consecutive microphones k and l of the endfire array with interspacing $d = 2$ cm for a diffuse noise field.

8 loudspeakers that are positioned on the edges of a cube centered around the dummy head. As an illustration, **Figure 2.5(b)** depicts the experimentally measured coherence function $\Gamma^n_{kl}(f)$ of two consecutive microphones of the 4-microphone endfire array with $d = 0.02$ m together with the theoretical coherence function (2.36), demonstrating that the generation of a diffuse noise field is quite good.

The desired speech signal and the noise signals are uncorrelated. Both the speech signal and the total noise signal have a level of 70 dB SPL at the center of the head. In case of a multiple noise source scenario, all the noise sources have the same level at the center of the head. For evaluation purposes, the speech and noise signal have been recorded separately in the experiments.

2.2 Fixed beamforming

The first techniques that have been used in hearing aids to spatially suppress the noise are fixed beamformers. Nowadays, fixed beamformers are often combined with some adaptive processing to provide adaptive directivity (cf. Section 2.3). In a fixed beamformer, the filter coefficients $\mathbf{w}_l[k], l = 1, ..., M$ in Figure 2.1 are kept fixed. They are optimized such that a focussed beam is steered towards the target direction (i.e., the forward-look direction $\theta = 0^\circ$) while reducing

background noise coming from other directions. This section discusses several types of fixed beamforming techniques applicable to hearing aids, i.e., filter-and-sum beamforming [39, 120, 142, 188, 191, 280], delay-and-sum beamforming [170, 259, 261], differential microphones [49, 254], sensitivity constrained superdirective beamforming [9, 28, 32, 123, 131, 214, 225, 224, 245] and robust filter-and-sum beamforming [37]. These techniques form a basic building block in some of the adaptive multi-microphone noise reduction techniques that are described in this thesis (cf. Section 2.3 and Chapter 4). Among the fixed beamformers, especially differential microphones have found wide application in commercial hearing aids. However, at low frequencies, these beamformers exhibit a high sensitivity to imperfections (such as microphone mismatch), limiting their practical use for commercial hearing aids to first-order (i.e., two-microphone) designs or at most second-order (i.e., three-microphone) designs at the high frequencies only.

In the mathematical description of the fixed beamformers, we always consider a *linear endfire array* for the sake of simplicity.

2.2.1 Spatial directivity pattern and directivity index

Fixed beamformers are often characterized by their *spatial directivity pattern* $H(f,\theta)$. The spatial directivity pattern $H(f,\theta)$ refers to the sensitivity of a fixed beamformer to sound coming from a particular direction θ:

$$H(f,\theta) = \mathbf{W}^H(f)\mathbf{d}(f,\theta), \tag{2.50}$$

where the steering vector $\mathbf{d}(f,\theta)$ has been defined in (2.33).

The noise reduction performance of a fixed beamformer is typically assessed through the *directivity index* $DI(f)$, defined as the ratio of the array gain for sound incident from the target direction, i.e., $\theta = 0^\circ$, to the array gain for diffuse noise:

$$DI(f) = \frac{\mathbf{W}^H(f)\mathbf{d}(f,0^\circ)\mathbf{d}^H(f,0^\circ)\mathbf{W}(f)}{\mathbf{W}^H(f)\mathbf{\Gamma}^n(f)\mathbf{W}(f)}, \tag{2.51}$$

with $\mathbf{\Gamma^n}(f)$ equal to (2.36).

Similar as in Section 2.1.4, a broadband spatial directivity pattern $H(\theta)$ and directivity index DI can be obtained by taking an intelligibility weighted average of the performance at different frequencies.

2.2.2 Filter-and-sum beamforming

The most general fixed broadband beamformer is the filter-and-sum beamformer [39, 120, 142, 188, 191, 280], which is depicted in **Figure 2.1**. The filter coefficients $\mathbf{w}_l[k]$, $l = 1, \ldots, M$ of the filter-and-sum beamformer are

computed so that the spatial directivity pattern $H(f,\theta)$ optimally fits a (predefined) desired directivity pattern $D(f,\theta)$ in the frequency range Ω and the angular range Θ for a given arbitrary microphone configuration. Several cost functions can be used that realize this objective. We refer to [33, 39] for an overview. An example is to minimize the weighted least squares (WLS) cost function

$$J_{WLS} = \int_{\Omega}\int_{\Theta} L(f,\theta)\left[H(f,\theta) - D(f,\theta)\right]^2 df\, d\theta, \tag{2.52}$$

where $L(f,\theta)$ is a positive real weighting function, assigning more or less importance to certain frequencies and/or angles.

2.2.3 Delay-and-sum beamforming

The simplest fixed beamforming technique, which can be interpreted as a special case of filter-and-sum beamforming, is delay-and-sum (DS) beamforming [259, 261]. A DS beamformer compensates for the arrival time differences of the desired sound signal $s_d[k]$ to each microphone by inserting delays τ_l, $l = 1, \ldots, M$ after each microphone. The time aligned signals at the outputs of the delays are then summed together:

$$z[k] \quad = \tfrac{1}{M} \quad \sum_{l=1}^{M} x_l[k-\tau_l], \tag{2.53}$$

where the delays τ_l equal

$$\tau_l = -\frac{(d_l - d_1)\cos\theta}{c} f_s \tag{2.54}$$

and where $\theta = 0^\circ$ is the direction of the desired sound. Since the delays τ_l are generally non-integer values, they are implemented by all-pass interpolation filters [117, 144].

For a uniform linear endfire array with microphone interspacing $d = \frac{A_l}{M-1}$, the DS beamformer $\mathbf{W}(f)$ equals

$$\mathbf{W}(f) = \frac{1}{M}\mathbf{d}(f, 0^\circ) \tag{2.55}$$

with $\mathbf{d}(f,0^\circ)$ defined in (2.33), and hence, the spatial directivity pattern $H(f,\theta)$ equals

$$H(f,\theta) = \frac{1}{M}\sum_{l=0}^{M-1} e^{j2\pi f l\left(\frac{d}{c} - \frac{d}{c}\cos\theta\right)} = \frac{1}{M}\frac{e^{j\frac{M}{2}\xi}}{e^{j\frac{\xi}{2}}}\frac{\sin(\frac{M\xi}{2})}{\sin(\frac{\xi}{2})} \tag{2.56}$$

with $\xi = 2\pi f \frac{d}{c}(1-\cos\theta)$. For small $\frac{M\xi}{2} \ll \pi$ and hence, for small $fMd \ll c$, the spatial directivity pattern becomes omnidirectional, i.e., $H(f,\theta) \approx 1$. For

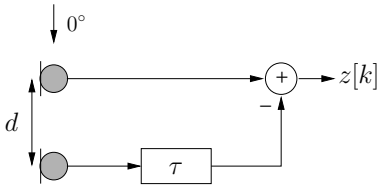


Figure 2.6: Differential microphone.

commercial hearing aids with typical array lengths A_l of $2-3$ cm, the directivity of a delay-and-sum beamformer is negligible, especially at low frequencies. DS beamformers are therefore especially encountered in 2-microphone binaural hearing systems [74, 140, 170, 271], where the distance d between the microphones is about $d \approx 0.15$ m.

However, an advantage of the DS beamformer is that it minimizes the noise sensitivity $\Psi(f)$ and hence, provides a high robustness against errors in the assumed signal model. Using (2.55) and $\mathbf{\Gamma}^s(f) = \mathbf{d}(f, 0^\circ)\mathbf{d}^H(f, 0^\circ)$ in (2.44), (2.38)-(2.40), we obtain

$$\Psi(f) = \frac{1}{M}. \tag{2.57}$$

2.2.4 Differential microphones

In contrast to DS beamformers, differential microphones [49, 254] offer a high frequency independent directivity with small-sized endfire arrays, making them popular in the design of commercial hearing aids.

Remark 2.3 Differential microphones are also referred to as superdirective arrays since their directivity index is higher than that of a DS beamformer [49]. □

First-order differential microphone

Figure 2.6 depicts the concept of a *first-order* differential microphone. A first-order differential microphone consists of two sound pressure sensors (e.g., two omnidirectional microphones) separated by a small distance d. Directivity is achieved either acoustically (e.g., a two-port hardware directional microphone) or electronically by delaying the rear microphone by τ and subtracting the delayed output from the output of the front microphone.

The *spatial directivity pattern* $H(f, \theta)$ equals

$$H(f, \theta) = 1 - e^{-j2\pi f\left(\tau + \frac{d}{c}\cos\theta\right)}. \tag{2.58}$$

Table 2.1: Commonly used spatial directivity patterns.

pattern	α_τ	τ	*property*
cardioid	$\frac{1}{2}$	$\frac{d}{c}$	zero at 180°
dipole	0	0	zero at 90°
supercardioid	$\frac{1}{1+\sqrt{3}}$	$\frac{d}{c\sqrt{3}}$	zero at 125°, maximum front-to-back ratio
hypercardioid	$\frac{1}{4}$	$\frac{d}{3c}$	zero at 109°, maximum directivity index ($DI(f) = 6$ dB)

Assuming a small microphone interspacing d and delay τ (i.e., $2\pi f \frac{d}{c} \ll \pi$ and $2\pi f\tau \ll \pi$),

$$H(f,\theta) \approx 2\pi f \left(\tau + \frac{d}{c}\cos\theta\right). \tag{2.59}$$

From (2.59), it follows that a first-order differential microphone has a first-order high pass frequency characteristic. To provide a distortionless response for the desired signal direction $\theta = 0°$, the high pass characteristic has to be compensated for by an equalization filter $W_{eq}(f)$

$$W_{eq}(f) = 1 - e^{-j2\pi f\left(\tau+\frac{d}{c}\right)} \approx \frac{1}{2\pi f\left(\tau + \frac{d}{c}\right)}. \tag{2.60}$$

Introducing $\alpha_\tau = \frac{\tau}{\tau+\frac{d}{c}}$, the normalized (i.e., after equalization with $W_{eq}(f)$), spatial directivity pattern $\tilde{H}(\theta)$ can be written as

$$\tilde{H}(\theta) \approx \alpha_\tau + (1-\alpha_\tau)\cos\theta, \tag{2.61}$$

resulting in a frequency-independent directivity pattern that has a maximum at the desired direction $\theta = 0°$ and a minimum at an angle $\theta \in [\frac{\pi}{2}, \pi]$ [49]. The value of α_τ (or the delay τ) determines the shape of the directivity pattern. For $\alpha_\tau \leq \frac{1}{2}$, the directivity pattern has a zero between $\frac{\pi}{2}$ and π and for $\frac{1}{2} \leq \alpha_\tau < 1$, the directivity pattern has its minimum at 180°. **Table 2.1** summarizes the most commonly used directivity patterns.

Using (2.59), we find that for small $fd \ll \frac{c}{2}$ and $f\tau \ll \frac{1}{2}$, the noise sensitivity $\Psi(f)$ equals

$$\Psi(f) \approx \frac{1}{2\pi^2 f^2\left(\tau + \frac{d}{c}\right)^2}. \tag{2.62}$$

The noise sensitivity increases dramatically for low frequencies $f \to 0$, resulting in a high sensitivity at low f to errors in the assumed signal model such as microphone mismatch [19, 254]. Consequently, the directivity index $DI(f)$ that is reached in real-life use is considerably less than 6 dB at low frequencies.

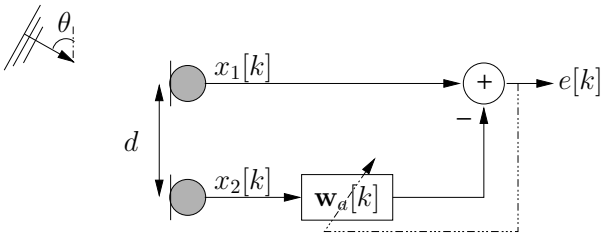


Figure 2.7: Microphone calibration.

For *higher-order* differential microphones, the noise sensitivity further increases proportional to $(fd)^{-2n}$ with $n = M - 1$ the order [49], limiting their practical use for hearing aids to at most a second-order differential microphone at the higher frequencies, i.e., $f > 1$ kHz [203, 254] (cf. Triano 3 of Siemens).

Microphone calibration

To reduce the detrimental effect of microphone mismatch and imperfections in microphone mounting, the microphones are often calibrated by an adaptive noise canceller (ANC) (cf. Section 2.3.1), as depicted in **Figure 2.7**. The ANC $\mathbf{w}_a[k]$ tries to equalize the mismatch by minimizing the power of the difference signal $e[k]$

$$e[k] = x_1[k] - \mathbf{w}_a^T[k]\mathbf{x}_2[k] \tag{2.63}$$

for an anechoic signal coming from a specified angle θ (e.g., $\theta = 90°$). The ANC $\mathbf{w}_a[k]$ models the inner product of the mismatch and the relative propagation delay $\tau_1 - \tau_2 = -\frac{d}{c}\cos\theta$ between the first and the second microphone.

Calibration is, however, never perfect: the mismatch is only compensated for one *single* direction of arrival θ. Furthermore, microphone characteristics drift over time [114]. In addition, the calibration has to be performed for each individual hearing aid, making it cumbersome and expensive. As a result, there is a high demand for more robust noise reduction algorithms.

2.2.5 Robust fixed beamformers

This section describes two procedures for designing fixed beamformers that are robust against microphone mismatch.

Sensitivity constrained superdirective beamforming

A commonly used design criterion consists of finding the filter coefficients $\mathbf{W}(f)$ that maximize the directivity index $DI(f)$ (see (2.51)) with a constraint on the

noise sensitivity $\Psi(f)$ [28, 123, 245], where

$$\Psi(f) = \frac{\mathbf{W}^H(f)\mathbf{W}(f)}{\mathbf{W}^H(f)\mathbf{d}(f,0^\circ)\mathbf{d}^H(f,0^\circ)\mathbf{W}(f)}. \tag{2.64}$$

To avoid distortion of signals coming from the target direction $\theta = 0^\circ$, the constraint $\mathbf{W}^H(f)\mathbf{d}(f,0^\circ) = 1$ is typically imposed on the filter $\mathbf{W}(f)$, leading to

$$\Psi(f) = \mathbf{W}^H(f)\mathbf{W}(f). \tag{2.65}$$

The optimization criterion is then most easily formulated as [28]:

$$\begin{aligned} \mathbf{W}(f) = \min_{\mathbf{W}(f)} & \mathbf{W}^H(f)\mathbf{\Gamma}^n(f)\mathbf{W}(f) + \eta(f)\mathbf{W}^H(f)\mathbf{W}(f) \\ & \text{subject to } \mathbf{W}^H\mathbf{d}(f,0^\circ) = 1 \end{aligned} \tag{2.66}$$

with $\mathbf{\Gamma}^n(f)$ the coherence matrix of diffuse noise (see (2.36)). The solution to (2.66) equals [28]:

$$\boxed{\mathbf{W}(f) = \frac{(\mathbf{\Gamma}^n(f) + \eta(f)\mathbf{I}_M)\,\mathbf{d}(f,0^\circ)}{\mathbf{d}^H(f,0^\circ)\,(\mathbf{\Gamma}^n(f) + \eta(f)\mathbf{I}_M)\,\mathbf{d}(f,0^\circ)}.} \tag{2.67}$$

The value of $\eta(f)$ trades of between maximal directivity and minimal noise sensitivity. For $\eta(f) = 0$, the directivity index $DI(f)$ is maximized. For an endfire array, the maximum $DI(f)$ equals $10\log_{10} M^2$ [49, 198]. For $\eta(f) \to \infty$, the noise sensitivity $\Psi(f)$ is minimized, resulting in the DS beamformer (2.55) and $\Psi(f) = -10\log_{10} M$ [28]. To find the optimum $\mathbf{W}$ for a given noise sensitivity $\Psi(f) = \zeta^2(f)$, one has to search iteratively for the value of $\eta(f)$ that produces the desired $\Psi(f) = \zeta^2(f)$.

Remark 2.4 For $\eta(f) = 0$ and a homogeneous noise field with coherence matrix $\mathbf{\Gamma^n}(f)$, (2.67) corresponds to the minimum variance distortionless response (MVDR) beamformer that minimizes the output noise power subject to a distortionless response in the target direction $\theta = 0^\circ$[260] (cf. Section 2.3.2). □

Filter-and-sum beamforming robust to gain and phase errors

Constraining the noise sensitivity $\Psi(f)$, improves the robustness against random errors. However, there is no straight relationship between the noise sensitivity and the amount of microphone mismatch that may occur, making it difficult to find an optimal value for the noise sensitivity constraint ζ^2 or $\eta(f)$ (2.66). In [37], two procedures for designing robust broadband beamformers have been proposed that allow to incorporate statistical knowledge about the microphone mismatch in the design criterion.

Suppose that the characteristic of the l-th microphone equals

$$\Upsilon_l(f,\theta) = \upsilon_l(f,\theta)e^{-j\phi_l(f,\theta)}, \tag{2.68}$$

where both the gain $\upsilon_l(f,\theta)$ and the angle $\phi_l(f,\theta)$ can be frequency and angle-dependent, and that the probability density function (pdf) $f_{\mathcal{Y}}(\Upsilon_l)$ of the stochastic variable Υ_l (i.e., the joint pdf of the gain υ_l and the angle ϕ_l) is known.

Suppose that $J(\Upsilon_1, \cdots, \Upsilon_M)$ is the cost function that one would have minimized, if the microphone characteristics $\{\Upsilon_1, \cdots, \Upsilon_M\}$ would have been known. A possible general objective is to design a broadband filter-and-sum beamformer $\mathbf{w}[k]$ with a (predefined) desired spatial directivity pattern $D(f,\theta)$ in the frequency range Ω and the angular range Θ, for a given arbitrary microphone array configuration [33, 39] (cf. Section 2.2.2).

A first procedure consists of minimizing a weighted sum of the cost functions $J(\Upsilon_1, \cdots, \Upsilon_M)$ for all possible microphone characteristics $\{\Upsilon_1, \cdots, \Upsilon_M\} \in D$, using the probability density functions of the microphone characteristics as weights [37]:

$$J_{\text{mean}} = \int_{\Upsilon_1} \cdots \int_{\Upsilon_M} J(\Upsilon_1, \cdots, \Upsilon_M) f_{\mathcal{Y}}(\Upsilon_1) \cdots f_{\mathcal{Y}}(\Upsilon_2) d\Upsilon_1 \cdots d\Upsilon_M. \tag{2.69}$$

A second procedure is to minimize the worst-case performance, i.e., the maximum cost $J(\Upsilon_1, \cdots, \Upsilon_M)$ for all feasible microphone characteristics $\{\Upsilon_1, \cdots, \Upsilon_M\} \in D$, resulting in a minimax criterion [37]

$$\min_{\mathbf{W}(f)} \max_{\forall\{\Upsilon_1, \cdots, \Upsilon_M\} \in \mathcal{D}} J(\Upsilon_1, \cdots, \Upsilon_M). \tag{2.70}$$

For further details about the optimization, we refer to [37].

2.3 Adaptive beamforming

In contrast to fixed beamformers, adaptive multi-microphone noise reduction techniques make use of data-dependent filters $\mathbf{w}_l[k]$ (cf. Figure 2.1) that are adapted to respond to time-varying environments. As a result, adaptive techniques generally have a better noise reduction performance than fixed beamforming techniques, particularly if the number of interferers is small (i.e., smaller than the number of microphones M in the array) and in acoustic environments with little reverberation. Adaptive multi-microphone noise reduction techniques can be classified into two categories, i.e., adaptive beamforming techniques and multi-channel Wiener filtering techniques. This section concentrates on adaptive beamforming techniques, while multi-channel Wiener filtering techniques are the scope of Section 2.4 and Section 2.5.

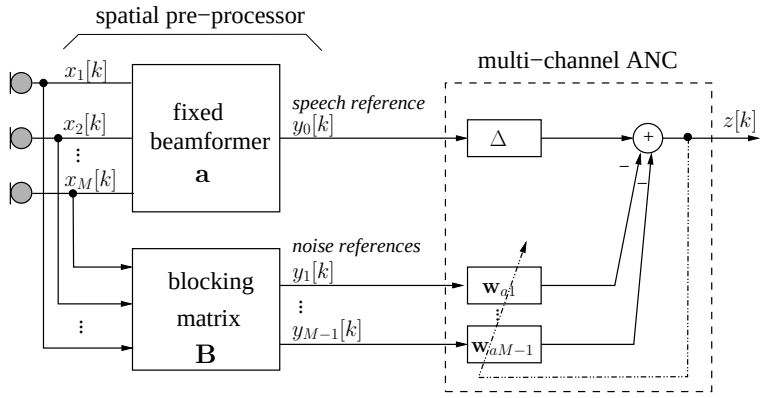


Figure 2.8: Generalized Sidelobe Canceller.

Adaptive beamforming techniques (e.g., the Frost beamformer [62, 261], the Generalized Sidelobe Canceller [20, 77]) typically solve a linearly constrained minimum variance (LCMV) optimization criterion [62, 117, 260, 261], minimizing the output power or output noise power subject to the constraint that signals coming from a certain region or direction of interest (i.e., ideally the direction of the desired speech source) are preserved. An efficient realization of the LCMV that is widely studied for application in hearing aids [74, 131, 140, 198, 263, 271] is the Generalized Sidelobe Canceller (GSC) [20, 77]: the GSC transforms the constrained optimization criterion into an unconstrained optimization criterion through a combination of a fixed spatial pre-processor and an adaptive noise canceller (ANC).

Section 2.3.1 describes the standard GSC. Section 2.3.2 discusses the minimum variance distortionless (MVDR) beamformer, which can be interpreted as a special case of the GSC. The GSC relies on a priori assumptions about the desired signal model and hence, tends to be sensitive to errors in these assumptions such as microphone mismatch. In Section 2.3.3 variants of the GSC are described that exhibit better robustness to signal model errors. Section 2.3.4 describes the (first-order) adaptive differential microphone array (ADMA) that is encountered in the latest two -microphone commercial hearing aids [23, 48, 157, 252]. This ADMA can be interpreted as a special, simplified case of the GSC.

2.3.1 Generalized Sidelobe Canceller

Figure 2.8 depicts the Generalized Sidelobe Canceller (GSC) [20, 77]. The GSC consists of a fixed, spatial pre-processor and a multi-channel adaptive noise canceller (ANC) [273] $\mathbf{w}_a[k]$.

Spatial pre-processor

The spatial pre-processor consists of a fixed beamformer $\mathbf{a} \in \mathbb{R}^{ML_A \times 1}$ and a blocking matrix $\mathbf{B} \in \mathbb{R}^{ML_B \times M-1}$. The *fixed beamformer* $\mathbf{a}$ creates a so-called speech reference

$$y_0[k] = \mathbf{a}^T \mathbf{x}[k] = y_0^s[k] + y_0^n[k], \tag{2.71}$$

with $\mathbf{x}[k] \in \mathbb{R}^{ML_A \times 1}$ by steering a beam towards the direction of the desired signal (i.e., 0°in our case of an endfire array), with a speech contribution $y_0^s[k]$ and a noise contribution $y_0^n[k]$. Typically, a DS beamformer $\mathbf{a}$ is used, but also more advanced fixed beamformers such as superdirective or filter-and-sum beamformers (cf. Section 2.2.5) could be used.

The *blocking matrix* $\mathbf{B} \in \mathbb{R}^{ML_B \times M-1}$ with

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_{M-1} \end{bmatrix} \tag{2.72}$$

creates $M-1$ so-called noise references

$$y_l[k] = \mathbf{b}_l^T \mathbf{x}[k] = y_l^s[k] + y_l^n[k],\ l = 1, ..., M-1, \tag{2.73}$$

with $\mathbf{x}[k] \in \mathbb{R}^{ML_B \times 1}$, by steering zeroes towards 0° so that the noise contributions $y_i^n[k]$ are dominant compared to the speech leakage contributions $y_i^s[k]$. In general, the blocking matrix $\mathbf{B}$ is designed to be orthogonal to the fixed beamformer $\mathbf{a}$. Ideally, the noise references do not contain any speech contribution, i.e., $y_i^s[k] = 0$, for $i = 1, ..., M-1$.

Example 2.1 A simple technique to create the noise references consists of pairwise subtracting the microphone signals $x_l[k]$ that have been time-aligned for 0°, i.e.,

$$y_l[k] \quad = \quad x_1[k] - x_l[k - \tau_l],\ l = 1, ..., M-1 \tag{2.74}$$

with $\tau_l = -\frac{(d_l - d_1)}{c} f_s$. Note that for small microphone interspacings $d_l - d_1$, this corresponds to creating a backward facing (i.e., steered to 180°) cardioid differential microphone (cf. Section 2.2.4). △

During periods of speech + noise, the references $y_i[k]$, $i = 0, ..., M-1$ contain speech + noise contributions. During periods of noise-only, the references only consist of a noise component, i.e., $y_i[k] = y_i^n[k]$. The second order statistics of the noise signal are assumed to be sufficiently stationary so that they can be estimated during periods of noise-only.

Multi-channel ANC

The adaptive noise canceller (ANC) $\mathbf{w}_a[k] \in \mathbb{R}^{(M-1)L \times 1}$ with

$$\mathbf{w}_a[k] \quad = \quad \begin{bmatrix} \mathbf{w}_{a1}^T[k] & \mathbf{w}_{a2}^T[k] & \cdots & \mathbf{w}_{aM-1}^T[k] \end{bmatrix}^T, \tag{2.75}$$

$$\mathbf{w}_{al}[k] \quad = \quad \begin{bmatrix} w_{al,0}[k] & w_{al,1}[k] & \cdots & w_{al,L-1}[k] \end{bmatrix}^T, \tag{2.76}$$

removes the correlation between the (delayed) speech reference $y_0[k-\Delta]$ and the noise references $y_l[k]$, $l = 1, ..., M-1$. The delay Δ is applied to the speech reference to effectively allow for non-causal taps in the filter $\mathbf{w}_a$. Usually, it is set to $\lceil \frac{L}{2} \rceil - 1$, where $\lceil x \rceil$ returns the smallest integer equal or larger than x.

Ideally, the noise references do not contain any speech, i.e., $y_l[k] = y_l^n[k]$ and the ANC $\mathbf{w}_a$ minimizes the output noise power, i.e.,

$$\boxed{\mathbf{w}_a = \arg\min_{\mathbf{w}_a} J(\mathbf{w}_a) = \arg\min_{\mathbf{w}_a} \mathcal{E}\{\left|y_0^n[k-\Delta] - \mathbf{w}_a^T[k]\mathbf{y}^n[k]\right|^2\}} \tag{2.77}$$

where

$$\mathbf{y}^{n,T}[k] = \left[\begin{array}{cccc} \mathbf{y}_1^{n,T}[k] & \mathbf{y}_2^{n,T}[k] & ... & \mathbf{y}_{M-1}^{n,T}[k] \end{array} \right], \tag{2.78}$$

$$\mathbf{y}_i^n[k] = \left[\begin{array}{cccc} y_i^n[k] & y_i^n[k-1] & ... & y_i^n[k-L+1] \end{array} \right]^T, \tag{2.79}$$

leading to

$$\boxed{\mathbf{w}_a = \mathcal{E}\{\mathbf{y}^n[k]\mathbf{y}^{n,T}[k]\}^{-1}\mathcal{E}\{\mathbf{y}^n[k]y_0^n[k-\Delta]\}.} \tag{2.80}$$

However, in the design of the fixed, spatial pre-processor, assumptions are made about the microphone characteristics, the position of the desired speech source and the microphone positions and furthermore reverberation is assumed to be absent. In practice, these assumptions are often violated (e.g. due to microphone mismatch and reverberation) so that speech leaks into the noise references. To limit the effect of such speech leakage on the ANC $\mathbf{w}_a$, the ANC is adapted during periods of noise-only [74, 93, 103, 258], so that (2.77) applies provided that the speech detection algorithm is perfect.

The above described standard GSC (i.e., with DS beamformer $\mathbf{a}$ and blocking matrix $\mathbf{B}$ created as in Example 2.1) with ANC adapted during noise-only, is used as a reference for the other noise reduction algorithms developed in this thesis.

NLMS based adaptation

A simple algorithm for adapting the ANC, is the *least mean squares (LMS)* algorithm [85, 273]. The LMS algorithm approximates the steepest descent algorithm

$$\mathbf{w}_a[j] = \mathbf{w}_a[j-1] + \frac{\rho}{2}\left(-\frac{\partial J(\mathbf{w}_a)}{\partial \mathbf{w}_a}\right)_{\mathbf{w}_a=\mathbf{w}_a[j-1]}, \tag{2.81}$$

where j denotes the iteration index and ρ the step size, using an instantaneous gradient estimate. Replacing the iteration index j by a time index k and leaving out the expectation operators $\mathcal{E}\{\cdot\}$, this results in the following update equation for the ANC $\mathbf{w}_a[k]$ (2.77) during periods of noise-only

$$\mathbf{w}_a[k] = \mathbf{w}_a[k-1] + \rho\mathbf{y}^{n,T}[k]\left(y_0^n[k-\Delta] - \mathbf{w}_a^T[k-1]\mathbf{y}^n[k]\right). \tag{2.82}$$

The step size ρ determines the convergence speed, the steady-state excess error and the stability. The larger ρ, the faster the convergence, but the larger the steady-state excess error. To guarantee convergence in the mean square, which means that the excess error $J(\mathbf{w}_a[k])$ converges to a finite steady-state value, the step size ρ is chosen such that it satisfies the following (conservative) criterion

$$0 < \rho < \frac{2}{L\sum_{i=1}^{M-1}\mathcal{E}\{y_i^n[k]\}}. \tag{2.83}$$

Since the upper bound is inversely proportional to the energy of the input noise vector $\mathbf{y}^n[k]$, the step size ρ is often normalized according to

$$\rho[k] = \frac{\bar{\rho}}{\mathbf{y}^{n,T}[k]\mathbf{y}^n[k] + c},$$

where $0 < \bar{\rho} < 2$ and where $c > 0$ is a small positive constant to prevent division by zero. In practice, the energy $\mathbf{y}^{n,T}[k]\mathbf{y}^n[k] = Lp_y^n[k]$ is efficiently computed by means of a first-order IIR update, i.e.,

$$p_y^n[k] = \gamma p_y^n[k-1] + (1-\gamma)\sum_{l=1}^{N-1}\left(y_l^n[k]\right)^2, \tag{2.84}$$

with $\gamma = 1 - \frac{1}{L}$.

The convergence of the LMS algorithm not only depends on ρ but also on the eigenvalue spread of the input correlation matrix $\mathcal{E}\{\mathbf{y}^n[k]\mathbf{y}^{n,T}[k]\}$: the larger the ratio $\lambda_{\max}/\lambda_{\min}$, with $\lambda_{\max}$ and $\lambda_{\min}$ the maximum and minimum eigenvalue of $\mathcal{E}\{\mathbf{y}^n[k]\mathbf{y}^{n,T}[k]\}$, respectively, the slower the convergence. To achieve fast convergence for spectrally colored input signals $y_l^n[k]$, other algorithms should be used such as RLS [85] (cf. Section 6.2.2 of Chapter 6) or frequency-domain LMS with a frequency dependent step size [219] (cf. Chapter 5).

2.3.2 Minimum variance distortionless response beamformer

A special case of the LCMV beamformer or the GSC is the minimum variance distortionless response beamformer (MVDR) [260, 261]. The MVDR beamformer $\mathbf{W}(f)$ minimizes the *output noise power* subject to a distortionless response in the target direction $\theta = 0^\circ$, i.e.,

$$\begin{aligned}\mathbf{W}(f) = \min_{\mathbf{W}(f)} &\mathbf{W}^H(f)\mathcal{E}\left\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\right\}\mathbf{W}(f)\\ &\text{subject to } \mathbf{W}^H(f)\mathbf{d}(f,0^\circ) = 1.\end{aligned} \tag{2.85}$$

The solution $\mathbf{W}(f)$ of (2.85) equals [260]

$$\mathbf{W}(f) = \frac{\mathcal{E}\left\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\right\}^{-1}\mathbf{d}(f,0^\circ)}{\mathbf{d}^H(f,0^\circ)\mathcal{E}\left\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\right\}^{-1}\mathbf{d}(f,0^\circ)} \tag{2.86}$$

or, for a homogeneous noise field,

$$\mathbf{W}(f) = \frac{\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{d}(f,0^\circ)}{\mathbf{d}^H(f,0^\circ)\mathbf{\Gamma}^{n^{-1}}\mathbf{d}(f,0^\circ)}. \tag{2.87}$$

It can be shown that the MVDR beamformer $\mathbf{W}(f)$ (2.86) corresponds to the GSC with ANC adapted during noise-only periods, if the fixed beamformer $\mathbf{A}(f)$ and the blocking matrix $\mathbf{B}(f)$ satisfy

$$\mathbf{A}(f) = \frac{\mathbf{d}(f,0^\circ)}{\mathbf{d}^H(f,0^\circ)\mathbf{d}(f,0^\circ)}; \quad \mathbf{B}^H(f)\mathbf{A}(f) = \mathbf{0}, \tag{2.88}$$

and if $\begin{bmatrix} \mathbf{B}(f) & \mathbf{A}(f) \end{bmatrix}$ is a non-singular matrix of dimensions $M \times M$ (cf. Section 3.2.1) [33, 116, 261].

Remark 2.5 For spatially uncorrelated noise, i.e., $\mathbf{\Gamma}^n(f) = \mathbf{I}_M$, the MVDR beamformer corresponds to the DS beamformer (2.55). For a diffuse noise field, the MVDR beamformer equals the superdirective beamformer (2.67) with $\eta(f) = \beta(f)$, where $\beta(f)$ denotes the internal-to-external noise ratio at the different microphones. □

In the literature, the MVDR beamformer is also often used to refer to the beamformer that minimizes the *output power* subject to the linear constraint $\mathbf{W}^H\mathbf{d}(f,0^\circ) = 1$, i.e.,

$$\mathbf{W}(f) = \min_{\mathbf{W}(f)} \mathbf{W}^H(f)\mathcal{E}\left\{\mathbf{X}(f)\mathbf{X}^H(f)\right\}\mathbf{W}(f) \text{ subject to } \mathbf{W}^H(f)\mathbf{d}(f,0^\circ) = 1, \tag{2.89}$$

resulting in

$$\mathbf{W}(f) = \frac{\mathcal{E}\left\{\mathbf{X}(f)\mathbf{X}^H(f)\right\}^{-1}\mathbf{d}(f,0^\circ)}{\mathbf{d}^H(f,0^\circ)\mathcal{E}\left\{\mathbf{X}(f)\mathbf{X}^H(f)\right\}^{-1}\mathbf{d}(f,0^\circ)}. \tag{2.90}$$

In [33, 260], it has been shown that in the presence of a single desired source, the filter (2.90) equals the MVDR beamformer (2.86) if the steering vector $\mathbf{d}(f,0^\circ)$ equals the desired signal direction, i.e., $\mathbf{d}(f,0^\circ) = e^{-j2\pi f\tau}\mathbf{H}(f)$ with $\mathbf{H}(f)$ the vector with the acoustic transfer functions of the desired signal (cf. Section 2.1.3) and τ a scalar.

2.3.3 Variants of the standard GSC with improved robustness

The standard GSC implementation assumes perfect knowledge of the microphone positions and characteristics, the location of the desired speaker and furthermore, the absence of reverberation. In practice, these assumptions are violated so that speech leakage into the noise references occurs. Distortion

of the desired signal can then not be avoided, even when adapting the ANC during periods of noise-only. Note that in case of speech leakage, the speech component $z^s[k]$ at the GSC output equals

$$z^s[k] = y_0^s[k-\Delta] - \mathbf{w}_a^T[k]\mathbf{y}^s[k]. \tag{2.91}$$

Therefore, different variants of the standard GSC have been proposed in the literature. They are based on reducing the speech leakage $\mathbf{y}^s[k]$ in the presence of signal model errors by an improved design of the spatial pre-processor and/or limiting the distortion caused by the ANC $\mathbf{w}_a[k]$ by sensitivity constrained adaptation.

Improved design of the spatial pre-processor

While in case of small-sized arrays, a fixed DS beamformer $\mathbf{a}$ offers a high robustness[3] to signal model errors, the directivity pattern and, in particular, the location of the zero of the noise references may be quite sensitive to small errors, resulting in a large amount of speech leakage $\mathbf{y}^s[k]$ when signal model errors occur.

Example 2.2 At small $f \cdot d$, the noise references in Example 2.1 correspond to a backward facing cardioid differential microphone and hence, exhibit a high sensitivity at low frequencies to signal model errors such as microphone mismatch [19, 254]. △

In the literature, speech leakage due to misteer (e.g., due to movements of the desired speaker or head movements of the hearing aid user) and microphone placement uncertainty is reduced by incorporating multiple linear constraints in the design of the fixed spatial pre-processor [96] or by using a spatial filter designed blocking matrix [25, 187], which minimizes the speech leakage for a specified angular region around 0° instead of for 0° only [187]. Leakage due to reverberation, on the other hand, is avoided by incorporating a transfer function model in the design of the spatial pre-processor [65] instead of delays only. Finally, instead of using a fixed blocking matrix, also an adaptive blocking matrix can be used to reduce speech leakage [93, 103, 258].

In Chapter 3, we will show that, among the different signal model errors, especially microphone mismatch is detrimental to the GSC. If statistical knowledge about the microphone characteristics is available, speech leakage in the noise references can be minimized for all possible microphone characteristics by using the robust filter-and-sum design procedures in [37] (cf. Section 2.2.5). Alternatively, sensitivity constrained superdirective beamformers that are steered to

[3]Note that this is required to avoid distortion of the speech component $y_0^s[k-\Delta]$ in the speech reference.

the back, could be used as noise references. The design procedures described in Section 2.2.5 can also be applied to create a more optimized fixed beamformer that has a better noise suppression than the DS beamformer but is still sufficiently robust to signal model errors. In this thesis, we concentrate on the design of a robust adaptive stage. Improvements of the spatial pre-processor like the ones suggested above will be studied in further research (cf. Section 8.2).

Sensitivity constrained adaptation of the ANC

Although the speech leakage is reduced by the above techniques, it can never be completely avoided. Hence, additional measures are required to limit speech distortion caused by the ANC $\mathbf{w}_a[k]$.

A common approach to increase the robustness of the ANC is to apply a *quadratic inequality constraint* (QIC) [29, 93, 96, 103, 255] to the ANC $\mathbf{w}_a[k]$:

$$\boxed{\begin{aligned}\mathbf{w}_a[k] \quad = \quad & \arg\min_{\mathbf{w}_a[k]} \mathcal{E}\{\left|y_0^n[k-\Delta] - \mathbf{w}_a^T[k]\mathbf{y}^n[k]\right|^2\} \\ & \text{subject to } \mathbf{w}_a^T[k]\mathbf{w}_a[k] \leq \zeta^2.\end{aligned}} \tag{2.92}$$

The QIC avoids excessive growth of the filter coefficients $\mathbf{w}_a[k]$. Hence, it reduces the undesired speech distortion when speech leaks into the noise references. In the remainder, we refer to the GSC with QIC (2.92) as QIC-GSC.

The QIC-GSC (2.92) is an efficient approximation for imposing a *constraint on the noise sensitivity* $\Psi_{\text{GSC}}(f)$ of the GSC, which was originally proposed to increase the robustness of LCMV or GSC beamforming against random errors [29, 108]. In [29, 255], it is shown that a constraint on the noise sensitivity $\Psi_{\text{GSC}}(f)$ corresponds to a QIC constraint on the filters $\mathbf{B}(f)\mathbf{W}_a(f)$, i.e.

$$\mathbf{W}_a^H(f)\mathbf{B}^H(f)\mathbf{B}(f)\mathbf{W}_a(f) \leq \zeta^2, \tag{2.93}$$

if $\mathbf{B}^H(f)\mathbf{A}(f) = 0$. Hence, for a blocking matrix $\mathbf{B}(f)$ that satisfies

$$\mathbf{B}^H(f)\mathbf{B}(f) = \mathbf{I}_{M-1}; \quad \mathbf{B}^H(f)\mathbf{A}(f) = \mathbf{0} \tag{2.94}$$

at each frequency f, the QIC (2.92) equals the noise sensitivity constraint (NSC) (2.93) [29, 255]. In Chapter 3, the robustness of the GSC with noise sensitivity constraint (NSC), i.e., (2.93) with $\mathbf{B}^H(f)\mathbf{A}(f) = \mathbf{0}$, will be compared with the GSC. In the sequel, the GSC with noise sensitivity constraint (NSC) is denoted by NSC-GSC.

In [29, 103], the QIC-GSC (2.92) is efficiently implemented by the *scaled projection algorithm (SPA)*: at each update step, the quadratic inequality constraint is imposed to the newly obtained ANC filter $\mathbf{w}_a[k]$ by scaling the filter coeffi-

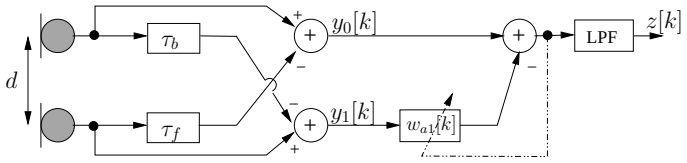


Figure 2.9: Adaptive differential microphone array (ADMA).

cients $\mathbf{w}_a[k]$ as

$$\mathbf{w}_a[k] \leftarrow \frac{\zeta}{\|\mathbf{w}_a[k]\|_2}\mathbf{w}_a[k] \quad \text{if } \|\mathbf{w}_a[k]\|_2^2 \geq \zeta^2. \tag{2.95}$$

Although this technique works well for LMS updating, it does not appear to be as effective for RLS as for LMS [255]. In [255], the quadratic inequality constraint has been implemented by using *variable loading*. For RLS, this technique provides a better approximation to the optimal solution (2.92) than the SPA. For LMS, variable loading does not appear to offer any performance advantage over the cheaper, scaled projection LMS.

Remark 2.6 In [25, 187], the QIC (2.92) is approximated by the optimization criterion (cf. Section 2.2.5)

$$\mathbf{w}_a = \arg\min_{\mathbf{w}_a} \mathcal{E}\{\left|y_0^n[k-\Delta] - \mathbf{w}_a^T[k]\mathbf{y}^n[k]\right|^2 + \eta\mathbf{w}_a^T[k]\mathbf{w}_a[k]\}, \tag{2.96}$$

which is also referred to as fixed loading [255]. For LMS updating, this optimization criterion leads to the very cheap, leaky LMS algorithm [85, 273]:

$$\mathbf{w}_a[k] = (1-\rho\eta)\mathbf{w}_a[k-1] + \rho\mathbf{y}^{n,T}[k]\left(y_0^n[k-\Delta] - \mathbf{w}_a^T[k-1]\mathbf{y}^n[k]\right) \tag{2.97}$$

A drawback of this approach is, however, that the required η for a given QIC value ζ^2 depends on the noise scenario. □

In this thesis, we concentrate on the widely studied *standard GSC with ANC adapted during noise-only periods* with and without NSC or QIC. In Chapter 4, we will propose an alternative technique for increasing the robustness of the ANC that achieves a better noise reduction than the QIC-GSC for a given maximum speech distortion level.

2.3.4 Adaptive differential microphone array

In [23, 48, 157, 252], a first-order adaptive differential microphone array (ADMA) has been proposed with the aim of adaptively steering a zero in the direction

of the strongest interferer. The ADMA is efficiently realized by combining two closely spaced, identical omnidirectional microphones in an endfire combination as depicted in **Figure 2.9**.

The ADMA can be interpreted as a special, simplified case of the GSC, consisting of a forward facing cardioid $\mathbf{a}$ ($\tau_f = \frac{d}{c}$), a backward facing cardioid $\mathbf{B}$ ($\tau_b = \frac{d}{c}$) and a one-tap ANC $w_{a1}[k]$. To compensate for the high pass frequency characteristic of the forward facing cardioid $\mathbf{a}$, the output of the ADMA is filtered by a low pass filter (LPF) (cf. Section 2.2.4). Because of its simplicity, the ADMA has found wide application in the latest two-microphone digital hearing aids (e.g., Phonak Claro dAZ [138] , GN ReSound Canta7 [157], Widex Senso Diva [113]) (cf. Section 1.4.1).

Under the assumptions of *identical perfectly omnidirectional microphones, far and free field propagation and small* $fd \ll \frac{c}{2}$, it can be shown that for this back-to-back cardioid configuration (i.e., $\tau_f = \tau_b = \frac{d}{c}$), a one-tap adaptive filter w_{a1} suffices to steer a zero to θ_n. In Appendix 2.A, we show that the directivity pattern $H(f, \theta)$ of the ADMA has a zero at θ_n iff (if and only if) the scalar w_{a1} equals [157, 252]

$$w_{a1} = \frac{1 + \cos\theta_n}{1 - \cos\theta_n}. \tag{2.98}$$

Note that $0 \leq w_{a1} \leq 1$ when $\theta_n \in [90°, 270°]$.

The one-tap ANC $w_{a1}[k]$ is adapted using NLMS. To avoid cancellation of the target signal in the front region $[-\theta_f, \theta_f]$, the scalar $w_{a1}[k]$ is constrained after each update to $[0, \zeta]$ with $\zeta = \frac{1+\cos\theta_f}{1-\cos\theta_f}$:

$$\begin{aligned} \tilde{w}_{a1}[k] &= w_{a1}[k-1] + \frac{\bar{\rho}}{y_1^2[k] + c} y_1[k] \left(y_0[k] - w_{a1}[k-1] y_1[k]\right) \\ w_{a1}[k] &= \min\left\{\max(\tilde{w}_{a1}[k], \zeta), 0\right\} \end{aligned} \tag{2.99}$$

Typically, $\zeta = 1$, so that only noise sources in the back hemisphere are cancelled [157, 252]. Note that (2.99) corresponds to imposing the QIC $\|w_{a1}\|_2^2 \leq \zeta^2$ on the ANC $w_{a1}[k]$ with the SPA (2.95).

Practical issues

In practice, the assumptions of identical omnidirectional microphones and free-field propagation are violated. The microphones do not have identical characteristics. Furthermore, the presence of the head and the mounting of the microphones onto the hearing aid affect the directivity patterns of the microphones.

Although an QIC is imposed on the ANC $w_{a1}[k]$, the ADMA is sensitive to signal model errors such as microphone mismatch, especially at low $f \cdot d$. Note that for $0 \leq w_{a1} \leq \infty$, the ADMA equals a first order differential microphone,

which exhibits a large sensitivity at low $f \cdot d$ (cf. Section 2.2.4). At low $f \cdot d$, a small model error will affect the directivity pattern of the forward and backward facing cardioid, resulting in significant speech distortion of the speech reference as well as speech leakage in the noise references. In [252], a second-order ADMA has been proposed, resulting in an even higher sensitivity. Like fixed differential microphones, the practical use of the ADMA for commercial hearing aids is therefore limited to first-order or at most second-order designs at the high frequencies only.

While theoretically a one-tap filter $w_{a1}[k]$ seems sufficient to efficiently suppress a single noise source in the free and far field at small $f \cdot d \ll \frac{c}{2}$, the performance of a one-tap filter is significantly reduced in practice. The one-tap assumption relies on a perfect back-to-back cardioid configuration, which only holds true for perfect identical microphones and small $f \cdot d \ll \frac{c}{2}$. In addition, the presence of internal noise[4] affects the adaptation of the one-tap ANC $w_{a1}[k]$, sacrificing the efficiency of a one-tap filter. Furthermore, the one-tap ANC $w_{a1}[k]$ may not suffice to suppress reverberation and multiple noise sources. In [252], a frequency-domain implementation of the ADMA is proposed so that multiple, spatially distinct noise sources with non-overlapping spectra can be suppressed.

2.4 Multi-channel Wiener filtering

A second category of adaptive multi-microphone noise reduction techniques is referred to as *Multi-channel Wiener Filtering* (MWF) [2, 35, 36, 33, 162, 189, 190, 220, 235, 212]. These techniques provide a minimum mean square error (MMSE) estimate of either the clean *speech source* signal $s_d[k]$ [220, 162], the (reverberant) speech component $x_l^s[k]$ in one of the microphone signals [35, 36, 33, 189, 190, 235, 212] or a reference signal [2, 189, 190]. Hence, they inevitably introduce some speech distortion.

Among the different MWF techniques, we can distinguish between three classes:

GSC with single-channel post-filter

The MWF that provides an estimate of the clean speech source signal $s_d[k]$, can be decomposed into an optimum MVDR beamformer (or, if some conditions are fulfilled[5], an optimum GSC) cascaded with a single-channel Wiener post-filter [220]. This motivates the combination of an *MVDR or a GSC with a single-channel post-filter* [11, 14, 56, 146, 162, 169, 220, 282, 283] for noise reduction.

[4] Note that, at low $f \cdot d$, the internal noise level at the output of the backward cardioid is large w.r.t. the external noise. The ANC w_1 will then focus on reducing the internal noise instead of the external noise, especially at low f.

[5] Note that the MVDR beamformer corresponds to a GSC with distortionless constraint $\mathbf{W}^H(f)\mathbf{d}(f,\theta) = 1$ where θ refers to the target signal direction.

The cascade of an MVDR or a GSC with a single-channel post-filter, however, relies on the same assumptions about the desired signal model as the MVDR or the GSC and is thus equivalently robust against errors in the assumed signal model.

The use of post-filtering was initially proposed to increase the noise reduction performance of a beamformer (e.g., DS beamformer or GSC) in reverberant rooms [282, 283]. The single-channel post-filter further reduces the noise at the beamformer output by exploiting short-time spectral characteristics of the noise and the speech. Different methods for estimating the post-filter have been suggested in the literature. Many of these techniques are based on the assumption of spatially uncorrelated noise [56, 162, 282, 283]. As a result, they provide suboptimal performance in localized or diffuse noise fields, especially when dealing with small microphone arrays. Improved post-filtering methods for spatially correlated noise have been suggested in [11, 14, 146, 169, 220].

MWF using calibration sequences

In [189, 190], an *MWF is proposed that uses calibration sequences*. The calibration sequences are recorded through the actual system when no noise is present. The MWF then provides an MMSE estimate of a calibration observation measured at one of the microphones or at a separate microphone located at the source position. The calibration scenario is assumed to be representative of the practical scenario (i.e., same microphone characteristics and microphone configuration, same nominal position and long-term power spectral density (PSD) of the desired source, same room acoustics etc.). Like for the GSC, deviations from these assumptions will result in performance loss.

Remark 2.7 Instead of using calibration sequences, a mathematical model of the source signal $s_d[k]$ is used in [70] to compute the MWF. This technique requires a priori information about the desired signal model (e.g., microphone configuration) and the PSD of the source signal $s_d[k]$. To reduce the sensitivity to errors in the desired signal model, the mathematical model is computed for a specified region around the source signal $s_d[k]$. □

MWF with unknown reference

In [35, 36, 212], *an MWF with unknown reference* has been developed. In contrast to the GSC and the previous MWF techniques, neither a priori information about the desired signal model nor calibration is required, making it an appealing approach from a robustness point of view. This MWF provides an MMSE estimate of the (reverberant) speech component in one of the microphone signals. It is uniquely based on estimates of the second order statistics of the speech and the noise signal.

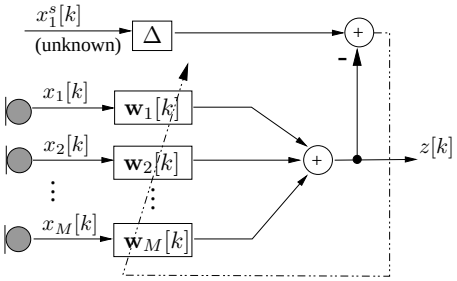


Figure 2.10: Multi-channel Wiener filter.

Remark 2.8 In [2], an alternative MWF has been proposed where the reference signal $d[k]$ of the MWF is generated by using a single-channel spectral subtraction algorithm based on minimum statistics [163]. □

In this thesis, we only focus on the third class of MWF, i.e., *MWF with unknown reference.* In the remainder, we will use the term MWF to refer to this technique.

In this section, we discuss the basic concepts of the MWF. Section 2.4.1 describes the optimization criterion of the MWF, while Section 2.4.2 briefly discusses possible (recursive) implementations. In the next section, we propose a subband implementation of the MWF to improve intelligibility and reduce computational complexity. Below, the formulas of the MWF are generalized to complex input signals so that they can be applied directly in the subband implementation of section 2.5.

2.4.1 Multi-channel Wiener filter (with unknown reference)

Figure 2.10 illustrates the concept of the MWF. The MWF $\mathbf{w}[k] \in \mathbb{C}^{ML\times 1}$

$$\mathbf{w}[k] = \left[\begin{array}{cccc} \mathbf{w}_1^H[k] & \mathbf{w}_2^H[k] & \cdots & \mathbf{w}_M^H[k] \end{array}\right]^H, \tag{2.100}$$

$$\mathbf{w}_l[k] = \left[\begin{array}{cccc} w_{l,0}[k] & w_{l,2}[k] & \cdots & w_{l,L-1}[k] \end{array}\right]^T, \tag{2.101}$$

minimizes the mean square error (MSE) between a delayed[6] version of the (unknown) speech component $x_l^s[k-\Delta]$ in the l-th (e.g. first) microphone signal and the sum $z[k] = \mathbf{w}^H\mathbf{x}[k]$ of the M filtered, received microphone

[6]The delay Δ is applied to effectively allow for non-causal taps of the filter $\mathbf{w}[k]$. Usually, it is set to $\left\lceil \frac{L}{2} \right\rceil - 1$.

signals [36], i.e.,

$$\boxed{\mathbf{w}[k] = \arg\min_{\mathbf{w}[k]} \mathcal{E}\left\{\left|x_l^s[k-\Delta] - \mathbf{w}^H[k]\mathbf{x}[k]\right|^2\right\},} \tag{2.102}$$

leading to

$$\boxed{\mathbf{w}[k] = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\mathcal{E}\{\mathbf{x}[k]x_l^{s,*}[k-\Delta]\}.} \tag{2.103}$$

Because of double-talk of speech and noise, the reference signal $x_l^s[k-\Delta]$ is unknown. If a robust speech detection is used, noise-only observations can be made during speech pauses (denoted with time index k'), where $x_l[k'] = x_l^n[k']$. This allows to estimate the second order statistics of the noise. These noise statistics are assumed to be sufficiently stationary, i.e.,

$$\mathcal{E}\{\mathbf{x}^n[k]x_l^{n,*}[k-\Delta]\} = \mathcal{E}\{\mathbf{x}^n[k']x_l^{n,*}[k'-\Delta]\}. \tag{2.104}$$

In addition, we assume that the speech and noise signals are statistically independent, i.e.,

$$\mathcal{E}\{\mathbf{x}^n[k]x_l^{s,*}[k-\Delta]\} = \mathbf{0}. \tag{2.105}$$

so that the correlation vector $\mathcal{E}\{\mathbf{x}^s[k]x_l^{s,*}[k-\Delta]\}$ can be estimated as

$$\mathcal{E}\{\mathbf{x}^s[k]x_l^{s,*}[k-\Delta]\} = \mathcal{E}\{\mathbf{x}[k]x_l^*[k-\Delta]\} - \mathcal{E}\{\mathbf{x}^n[k]x_l^{n,*}[k-\Delta]\}. \tag{2.106}$$

Using (2.104)-(2.106), the MWF (2.103) becomes:

$$\mathbf{w}[k] = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\left(\mathcal{E}\{\mathbf{x}[k]x_l^*[k-\Delta]\} - \mathcal{E}\{\mathbf{x}^n[k]x_l^{n,*}[k-\Delta]\}\right), \tag{2.107}$$

where the second order statistics $\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}$ and $\mathcal{E}\{\mathbf{x}^n[k]x_l^{n,*}[k-\Delta]\}$ are estimated during periods of speech + noise and periods of noise-only, respectively. In contrast to the GSC, the MWF does not make any a priori assumptions about the desired signal model. It is uniquely based on estimates of the second order statistics of the speech and noise signal.

In the sequel, we will assume, without loss of generality, that the speech contribution in the first microphone signal is estimated, i.e. $x_1[k-\Delta]$, and that the delay $\Delta = \lceil \frac{L}{2} \rceil - 1$, where $\lceil x \rceil$ represents the smallest integer larger than or equal to x.

Remark 2.9 The MMSE criterion (2.102) of the MWF can be generalized to allow for a trade-off between noise reduction and speech distortion [36], as will be elaborated in Section 4.1. □

2.4.2 Implementation of the MWF

In [33, 35, 36, 212, 213], recursive implementations based on a generalized singular value decomposition (GSVD) or a QR decomposition (QRD) of a speech + noise data matrix and a noise data matrix have been proposed for the MWF (2.107).

GSVD based MWF

In [33, 35, 36, 40], a GSVD based implementation for the MWF (2.107) has been proposed. Define the filter matrix $\mathbf{W}[k] \in \mathbb{C}^{ML \times ML}$ as

$$\mathbf{W}[k] = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\} - \mathcal{E}\{\mathbf{x}^n[k]\mathbf{x}^{n,H}[k]\}, \tag{2.108}$$

so that the i-th column of $\mathbf{W}[k]$ corresponds to the MWF (2.107) that provides an MMSE estimate of $x_l^s[k-\Delta]$ with $l = \left\lfloor \frac{i-1}{L} \right\rfloor + 1$ and $\Delta = \mathrm{mod}(i-1, L)$, where $\lfloor x \rfloor$ represents the largest integer smaller than or equal to x and where $\mathrm{mod}(i-1, L)$ denotes the remainder of the division $\frac{i-1}{L}$.

The filter matrix $\mathbf{W}[k]$ is approximated at time instant k by means of a GSVD of an input data matrix $\mathbf{X}[k] \in \mathbb{C}^{p_k \times ML}$ and a noise data matrix $\mathbf{X}^n[k] \in \mathbb{C}^{q_k \times ML}$, i.e.,

$$\mathbf{X}[k] = \begin{bmatrix} \varphi[k]\mathbf{x}^T[k] \\ \varphi[k-1]\mathbf{x}^T[k-1] \\ \vdots \\ \varphi[k-p_k+1]\mathbf{x}^T[k-p_k+1] \end{bmatrix}, \tag{2.109}$$

$$\begin{aligned} \mathbf{X}^n[k] &= \begin{bmatrix} (1-\varphi[k])\mathbf{x}^T[k] \\ (1-\varphi[k-1])\mathbf{x}^T[k-1] \\ \vdots \\ (1-\varphi[k-q_k+1])\mathbf{x}^T[k-q_k+1] \end{bmatrix} \\ &= \begin{bmatrix} (1-\varphi[k])\mathbf{x}^{n,T}[k] \\ (1-\varphi[k-1])\mathbf{x}^{n,T}[k-1] \\ \vdots \\ (1-\varphi[k-q_k+1])\mathbf{x}^{n,T}[k-q_k+1] \end{bmatrix}, \end{aligned} \tag{2.110}$$

collecting respectively p speech+noise and q noise-only signal vectors (corresponding to speech pauses) up to time instant k. The sequence $\varphi[k]$ denotes the output of the speech detection algorithm at time k, where $\varphi[k] = 1$ represents a speech + noise observation and where $\varphi_k = 0$ represents a noise-only observation. The scalars p_k and q_k are chosen such that

$$\sum_{l=k-p_k+1}^{k} \varphi[l] = P; \qquad \sum_{l=k-q_k+1}^{k} (1-\varphi[l]) = Q. \tag{2.111}$$

The GSVD of the data matrices $\mathbf{X}[k]$ and $\mathbf{X}^n[k]$ is defined as

$$\begin{cases} \mathbf{X}[k] &= \mathbf{U}[k]\mathrm{diag}\{\sigma_i[k]\}\mathbf{Q}^H[k] \\ \mathbf{X}^n[k] &= \mathbf{U}^n[k]\mathrm{diag}\{\eta_i[k]\}\mathbf{Q}^H[k] \end{cases} \tag{2.112}$$

with $\mathbf{U}[k]$ and $\mathbf{U}^n[k]$ orthogonal matrices, $\mathbf{Q}$ an invertible but not necessarily orthogonal matrix containing the generalized singular vectors and $\frac{\sigma_i}{\eta_i}$ the generalized singular values.

Using time averaging to estimate the 2nd order statistics[7]

$$\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\} \approx \mathbf{X}^T[k]\mathbf{X}^*[k]/p \quad (2.113)$$

$$\mathcal{E}\{\mathbf{x}^n[k]\mathbf{x}^{n,H}[k]\} \approx \mathbf{X}^{n,T}[k]\mathbf{X}^{n,*}[k]/q \quad (2.114)$$

and substituting (2.112)-(2.114) in (2.108), gives:

$$\mathbf{W}_k^* \simeq \mathbf{Q}^{-H}\mathrm{diag}\left\{\max\left(1-\frac{p}{q}\frac{\eta_i^2[k]}{\sigma_i^2[k]},0\right)\right\}\mathbf{Q}^H. \quad (2.115)$$

Since theoretically $\mathcal{E}\left\{1-\frac{p}{q}\frac{\eta_i^2[k]}{\sigma_i^2[k]}\right\} \geq 0$, negative values of $1-\frac{p}{q}\frac{\eta_i^2[k]}{\sigma_i^2[k]}$ are set to 0. The filter $\mathbf{w}[k]$ in (2.107) corresponds to the ith column of $\mathbf{W}[k]$ with $i = \Delta + lL + 1$.

Finally, suffice it to say that, in practice, the GSVD is not calculated from scratch at each time k but updated recursively using a square-root free implementation, reducing the complexity from $\mathcal{O}(16M^3L^3)$ to $\mathcal{O}(21.5M^2L^2)$ (cf. Table 2.2) [35, 40]. A further decrease in complexity can be obtained by reducing the number of filter and GSVD updates by a factor s_f and s_g, respectively, at the expense of tracking performance [40]. In the sequel, $s_f = s_g = 1$.

Remark 2.10 To allow for tracking of changing scenarios, exponential weighting with forgetting factors λ_{s+n} and λ_n is incorporated in the update formulas for the speech + noise and noise-only correlation matrix, respectively [33, 212]. An exponential weighting factor λ_{s+n} or λ_n corresponds to $p \approx \frac{1}{1-\lambda_{s+n}}$ and $q \approx \frac{1}{1-\lambda_n}$, respectively, in (2.113) and (2.114). For a sampling frequency $f_s = 16$ kHz, typical values for λ_{s+n} and λ_n are around 0.99998, which corresponds to a time-window $1/[f_s(1-\lambda_{s+n})] = 1/[f_s(1-\lambda_n)] \approx 3$ seconds, i.e., much longer than the average short-time stationarity of speech. Hence, the performance of the MWF is largely dependent on the *long-term average* spectral and spatial characteristics of the speech and noise sources so that no short-time effects, such as residual musical noise, will occur in this MWF. This implies that the MWF can also be used to suppress spectrally non-stationary noise, provided that the long-term, spectral and spatial characteristics of the speech and the noise vary slowly in time [33] (cf. Chapter 3). □

Other implementations

In [210, 212, 213], a QRD based and a fast QRD based implementation of the MWF have been proposed, resulting in a lower cost than recursive GSVD for a comparable performance.

Table 2.2 summarizes the computational complexity, i.e., the number of floating point operations (additions and multiplications) per second (flops) of the

[7] This assumes stationarity and ergodicity of the signals $x_l[k]$ and $x_l^n[k]$ with $l = 1, \ldots, M$.

Table 2.2: Complexity (expressed in flops) of the recursive square-root free GSVD, the QRD and fast QRD implementation of MWF ($f_s = 16$ kHz).

Algorithm	*Flops*	*E.g., $f_s = 16$ kHz, $L = 24$*	
		$M = 3$	$M = 2$
GSVD	$f_s \left(\frac{17.5}{s_g} + \frac{4}{s_f} \right) L^2 M^2$	1783,3 Mflops	792,6 Mflops
QRD	$f_s \left(3.5\, L^2 M^2 + 15.5\, ML + M + 2 \right)$	209,1 Mflops	129,3 Mflops
fast QRD	$f_s \left((21L - 11.5)\, M^2 + 19ML - 3.5M \right)$	92,6 Mflops	46,0 Mflops

square-root free recursive GSVD, the QRD and the fast QRD implementation. As an example, the computational complexity has been computed for $f_s = 16$ kHz, filter length $L = 24$ and $M = 2$ and $M = 3$ microphones, respectively. Although the computational complexity of a fast QRD implementation is considerably lower than the cost of GSVD, it is still far too high for implementation in commercial hearing aids, which typically can execute only 2 to 5 MIPS (cf. Section 1.4).

In the next section, we propose a subband MWF implementation [235] that results in improved intelligibility at a significantly lower cost compared to a fullband approach. In order to further reduce computational complexity, we will derive low-cost, time-domain and frequency-domain stochastic gradient algorithms for the MWF in Chapter 5.

2.5 Subband MWF implementation

In this section, we develop a subband implementation of the MWF. We show that in the case of colored signals, the subband implementation improves intelligibility more than the fullband approach. In addition, a significant complexity reduction is achieved.

2.5.1 Improved intelligibility at lower computational cost

In hearing aid applications, the main purpose of a noise reduction algorithm is to improve speech intelligibility, rather than to maximize SNR or to minimize MSE. The quality of a noise reduction algorithm is therefore assessed in terms of speech intelligibility weighted measures [73] such as the intelligibility weighted $\text{SNR}_{\text{intellig}}$ (cf. Section 2.1.4). These measures, based on the articulation index (AI), reflect the fact that distinct frequency bands contribute independently and differently to intelligibility.

Most acoustic interferers are broadband with their power concentrated at the low frequencies, e.g., car noise, speech interference. Hence, their spectrum resembles the long-term average speech spectrum. In this case, minimizing the fullband MSE will weigh the MSE (i.e. sum of distortion and noise) more at the high-energy low frequencies than at the low-energy high frequencies. If the number of degrees of freedom (and thus the filter length) is small, the filter taps are mainly used to improve the MSE at the low frequencies, resulting in a small improvement and large distortion of the desired signal at higher frequencies (see also Section 2.5.4).

However, all frequency bands contribute independently to intelligibility. Hence, an SNR improvement in any of the frequency bands, especially in the frequency range from 500 to 4000 Hz, can increase the intelligibility of speech. Maximization of intelligibility can thus better be achieved by minimizing the MSE in separate subbands. By doing so, more attention is paid to the low-energy frequency bands, i.e., in the case of an interferer with a low pass spectrum, the high frequencies. Given that a small-sized array can better reduce correlated noise at the high frequencies than at the low frequencies, the additional noise reduction at the higher frequencies may lead to a significant increase in $\text{SNR}_{\text{intellig}}$.In addition to a better noise suppression, a subband MWF implementation results in less speech distortion at high frequencies. Hence, it may also improve the speech quality, provided no degradation occurs at the lower frequencies.

Finally, in a subband implementation, the computation of the MWF can be performed at a lower rate, leading to a significant reduction in computational complexity (cf. Section 2.5.3)

Remark 2.11 Although the subband implementation is developed here for the MWF with unknown reference, its benefits over a fullband implementation also apply to other noise reduction algorithms such as the GSC and other MWF techniques (cf. Section 2.4). □

Remark 2.12 In [150], a pre-whitening operation has been proposed to improve the intelligibility weighted performance of the LCMV beamformer or the GSC. Such pre-whitening is intrinsically present in a subband implementation. When using pre-whitening, the energy of the input data and hence, the modelling effort of the fullband adaptive filter, is more evenly distributed throughout the frequency band of interest. However, to design a perfect pre-whitening operation, the power spectral density of the interferer has to be known. Furthermore, compared to a subband implementation, pre-whitening results in a significantly larger computational complexity, while not offering better performance [235]. In the forthcoming chapters, the pre-whitening operation is applied when evaluating fullband algorithms. In this case, the microphone signals are pre-whitened prior to processing. As a pre-whitening filter,

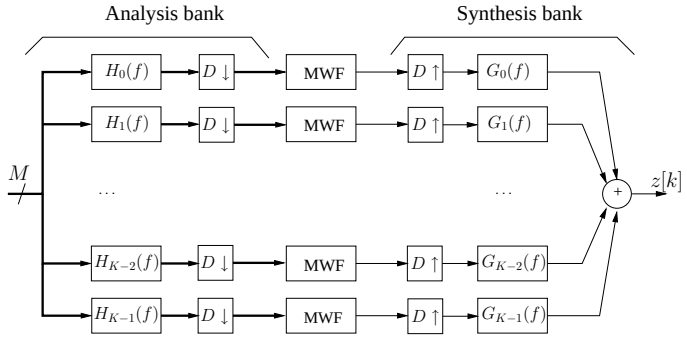


Figure 2.11: Subband implementation of the MWF

we use a 20th order FIR filter that models the inverse of the long-term average speech spectrum. The output of the noise reduction algorithm is de-whitened by filtering it with the inverse of the pre-whitening filter. □

2.5.2 Uniform subband implementation

The subband implementation of the MWF is illustrated in **Figure 2.11**. The M microphone signals are each split into K frequency bands by a (nearly) perfect reconstruction (PR), oversampled uniform discrete Fourier transform (DFT) modulated filterbank (see below) [50]. These filterbanks can be implemented efficiently using a polyphase decomposition of the prototype filter $H_0(f)$ and a fast Fourier transform (FFT). Since the fullband signals are real and (roughly) half of the complex DFT modulated analysis filters are the complex conjugate of the other half, only half [8] of the complex subband signals need to be processed.

The MWF filter operates on the subband signals at the decimated sampling frequency $\frac{f_s}{D}$. To avoid audible interband aliasing, the downsampling factor D is kept smaller than the number of subbands K. In [270], it is shown that for a system identification set-up, the aliasing level in the subbands corresponds to a non-linear distortion and thus presents a lower bound for the minimum achievable MSE in each subband. This indicates that the amount of aliasing determines the performance and hence should be kept small if the MSE is minimized in each subband independently (and thus no fullband adaptation criterion is used). A large stopband attenuation and small amount of interband aliasing are thus required.

[8] Actually, two real and $\frac{K}{2}-1$ complex signals need to be processed if K is even and one real and $\frac{K-1}{2}$ complex signals need to be processed if K is odd. We will approximate this in the complexity calculations later on by $\frac{K}{2}$ complex signals.

If no filtering (or only a delay Δ) is performed in the subbands, which e.g. occurs in a noise free scenario (since then $\mathbf{w}[k] = \delta[k - \Delta]$ in (2.102)-(2.103)), the effect of the analysis and synthesis bank should approximate a delay Δ_f. This is called a perfect reconstruction (PR) filterbank. The delay Δ_f caused by a DFT modulated filterbank equals [50]

$$\Delta_f = D \left\lceil \frac{L_p}{D} \right\rceil + D - 1 \tag{2.116}$$

samples at f_s (with $f_s = 16$ kHz), with L_p the length of the prototype filter $H_0(f)$. The filter bank delay Δ_f as well as the computational cost of the filterbank implementation increases with increasing filter length L_p. In hearing aid applications, the total processing delay should be smaller than 20 ms in order not to degrade intelligibility due to asynchronism with lip reading [247]. The total processing delay of the subband MWF algorithm consists of the delay Δ_f of the filterbank and the delay of the MWF (i.e., Δ_{sub} samples at $\frac{f_s}{D}$). Typically, $\Delta_{sub} = \left\lceil \frac{L_{sub}}{2} \right\rceil - 1$ with L_{sub} the filter length per channel used in the subbands. Hence, the total processing delay equals

$$\underbrace{D \left\lceil \frac{L_p}{D} \right\rceil + D - 1}_{\Delta_f} + \underbrace{D \left(\left\lceil \frac{L_{sub}}{2} \right\rceil - 1 \right)}_{\Delta_{sub}} \tag{2.117}$$

samples at f_s. For reasonable L_{sub}, the delay of the MWF is negligible compared to the delay Δ_f of the filterbank. In order to achieve a large stopband attenuation[9] with reasonable L_p, so-called "nearly PR" filter banks [50] are used. Note that the perfect reconstruction property would be lost anyway when a filtering operation is performed on the subband signals. The deviation from perfect reconstruction is expressed by the reconstruction error [270]

$$RE = \sum_{n=-\infty}^{\infty} (t[n] - \delta[n - \Delta_f])^2, \tag{2.118}$$

where $t[n]$ is the distortion function of the filterbank and Δ_f the delay caused by the filterbank. It should be noted that, in contrast to aliasing, the distortion caused by the "nearly PR" filterbank is a *linear* distortion. Together with the stopband attenuation, the reconstruction error is a measure for the minimum achievable MSE for the fullband signal [270].

In the sequel, two nearly PR oversampled DFT modulated filter banks are used: a 16-channel filter bank with $D = 12$, prototype filter length $L_p = 120$, delay $\Delta_f = 8.2$ ms and reconstruction error $RE = -34$ dB, and a 32-channel filter bank with $D = 20$, $L_p = 240$, $\Delta_f = 16.2$ ms and $RE = -40$ dB.

[9] Remember that a large stopband attenuation avoids aliasing and hence, non-linear distortion.

2.5.3 Computational complexity

Since in the subband approach, the MWF operates at a D times lower rate, a significant complexity reduction is achieved. The overall computational cost of the subband MWF algorithm consists of the complexity of the filter bank and the complexity of the subband MWF. Any of the algorithms discussed in Section 2.4.2 can be used to implement the subband MWF (i.e., a square-root free recursive GSVD, a QRD or a fast QRD implementation). As an example, we consider the recursive GSVD implementation.

When using a DFT modulated filter bank, only half the number of complex subband signals have to be processed at the rate $\frac{f_s}{D}$. However, the computational cost of a complex MWF is about 4 times the computational cost of a real MWF.

Example 2.3 When using the square-root free GSVD implementation, the subband MWF filters cost about

$$2K\frac{f_s}{D}(21.5)L_{sub}^2M^2 \tag{2.119}$$

flops, where L_{sub} equals the filter length per channel used in the subbands. $\triangle$

The implementation of a DFT modulated analysis or synthesis filter bank [50] requires

$$(2L_p + 2K\log_2 K)\frac{f_s}{D} \tag{2.120}$$

flops. The total complexity of the filter bank implementation (i.e., M analysis banks and 1 synthesis bank) thus equals

$$(M+1)(2L_p + 2K\log_2 K)\frac{f_s}{D}. \tag{2.121}$$

In the case of a GSVD implementation, the cost of the filter bank will be negligible compared with the cost of the subband MWF. From (2.119) and Table 2.2, it follows that the subband GSVD will roughly be a factor $\frac{D^3}{2K}$ less complex if the same temporal window or temporal dimension T is used. The temporal window T, expressed as the number of samples at sampling rate f_s, equals the filter length L for a fullband implementation and $L_{sub}D$ for a subband approach. This amounts to roughly a factor 54 and 125, respectively, for the above given filter banks.

Table 2.3 compares the computational complexity (expressed in Mflops) of the recursive fullband and subband GSVD approach (16 and 32 subbands) for different number of microphones M and a temporal window $T = 24$. A subband implementation significantly reduces the complexity of the GSVD based

Table 2.3: Comparison of the computational complexity (expressed in Mflops) of fullband and subband MWF implementations for a temporal window $T = LD = 24$. The sampling frequency $f_s = 16$ kHz.

Algorithm		*Fullband*	*Subband* ($T = L_{sub}D = 24$)	
		$T = L = 24$	$K = 16$, $D = 12$	$K = 32$, $D = 20$
GSVD	$M = 2$	792.6 Mflops	16.2 Mflops	8.3 Mflops
	$M = 3$	1783.3 Mflops	35.0 Mflops	16.8 Mflops
	$M = 4$	3170.0 Mflops	61.2 Mflops	28.6 Mflops
QRD	$M = 2$	129.3 Mflops	6.7 Mflops	5.1 Mflops
	$M = 3$	209.5 Mflops	11.5 Mflops	8.0 Mflops
fast QRD	$M = 2$	46.0 Mflops	9.6 Mflops	6.7 Mflops
	$M = 3$	92.6 Mflops	18.1 Mflops	11.8 Mflops

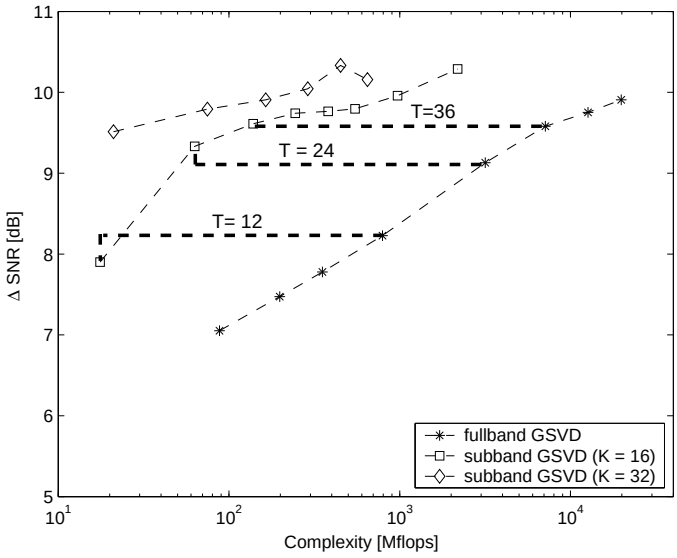
MWF. A further decrease in complexity can be obtained by using a subband implementation of the QRD or fast QRD algorithm (cf. Section 2.4.2), making an implementation in current 2-microphone commercial hearing aids almost feasible. For $M > 2$, a further reduction in complexity can be obtained by using only two microphones at the low frequencies.

Remark 2.13 In addition to a reducing the computational cost, a subband implementation of the matrix-based GSVD or QRD algorithm considerable reduces the memory requirements. In the fullband GSVD or QRD algorithm, $LM \times LM$-dimensional matrices have to be stored, resulting in a memory requirement of $\mathcal{O}(L^2M^2)$. In a subband implementation, $\frac{K}{2}$ complex valued matrices of dimensions $\frac{L}{D}M \times \frac{L}{D}M$ are stored, reducing the memory requirements by a factor $\frac{K}{D^2}$, i.e., $\mathcal{O}(\frac{K}{D^2}L^2M^2)$. □

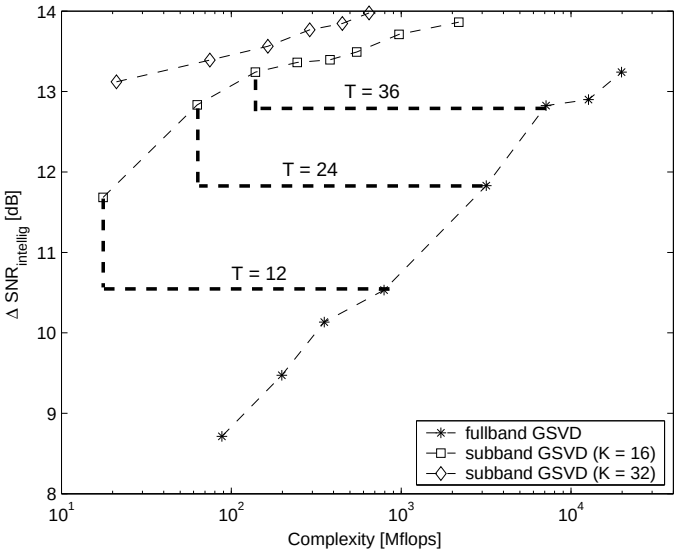
2.5.4 Experimental results

In this section, we compare the performance of the fullband and the subband GSVD based MWF for a desired source at $\theta = 0°$ and a noise source at $\theta = 90°$ based on recordings with the 4−microphone endfire array with uniform microphone interspacing $d = 0.02$ m (cf. Section 2.1.5). To get an idea of the average performance based on directivity only, the desired and noise signal are both stationary speech-weighted noise signals (i.e., track 1 of [106]).

Figure 2.12 (a) and (b) plot the improvement in ΔSNR (fullband SNR) and the improvement $\Delta\text{SNR}_{\text{intellig}}$, obtained by varying the filter length L of the fullband GSVD and the filter length per subband L_{sub} in the subband GSVD, as a function of complexity (expressed in Mflops). The dashed lines compare



(a)



(b)

Figure 2.12: (a) ΔSNR of fullband and subband GSVD as a function of complexity; (b) $\Delta\text{SNR}_{\text{intellig}}$ of fullband and subband GSVD. The dashed lines connect points of the fullband GSVD and subband GSVD ($K = 16$) with the same temporal window T.

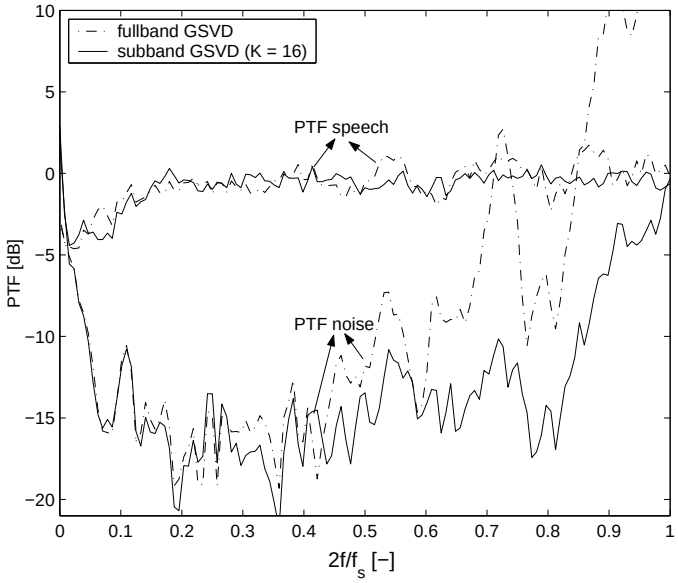


Figure 2.13: PTF of the desired signal and the noise signal for the fullband and subband GSVD. The sampling frequency $f_s = 16$ kHz.

points of the fullband GSVD and the subband GSVD ($K = 16$) with the same temporal window T. For the same complexity, a significant performance gain is achieved with subband GSVD. The gain in $\text{SNR}_{\text{intellig}}$ (2 to 4 dB) is larger than the gain in SNR (1 to 2.5 dB). For the same temporal window T, the subband GSVD achieves a better $\Delta\text{SNR}_{\text{intellig}}$, especially for small filter lengths L, while the improvement ΔSNR is comparable. This indicates that for colored signals such as speech, the subband approach indeed tends to provide improved intelligibility compared to the fullband method.

Figure 2.13 plots the PTF $G^s(f)$ and $G^n(f)$ from the first microphone to the algorithm's output of the desired and interfering signal obtained with the fullband and the 16-subband approach for a temporal window $T = 24$. The sampling frequency f_s equals 16 kHz. As expected, the subband implementation results in less distortion and larger noise reduction at the higher frequencies (i.e., $f > 2.3$ kHz) than the fullband implementation, while maintaining the performance at the low frequencies.

In short, a subband MWF implementation achieves a larger improvement in $\Delta\text{SNR}_{\text{intellig}}$ at a significantly lower cost.

2.6 Conclusion

In this chapter, we have discussed several multi-microphone noise reduction techniques and have defined the appropriate performance measures for evaluating them for hearing aid applications. Special attention has also been paid to the algorithms that are encountered in the latest commercial hearing aids, i.e., fixed and adaptive differential microphone arrays (ADMA). At low frequencies, these algorithms exhibit a high sensitivity to signal model errors such as microphone mismatch, which exponentially increases with an increasing number of microphones. As a result, their practical use for hearing aids is limited to first-order designs or at most second-order designs at the high frequencies only.

Multi-microphone noise reduction techniques can be broadly classified into three classes, namely, fixed beamforming, adaptive beamforming and multi-channel Wiener filtering. In contrast to fixed beamformers, adaptive beamformers and MWF techniques make use of data-dependent filters that are adapted to respond to time-varying environments. As a result, they generally have a better noise reduction performance than fixed beamforming techniques.

A widely studied adaptive, beamforming algorithm is the Generalized Sidelobe Canceller, of which the (first order) ADMA encountered in the latest two-microphone commercial hearing aids is a special case. The GSC consists of a fixed beamformer, a blocking matrix and a multi-channel ANC. Like fixed beamformers, the GSC relies on a priori knowledge of the desired signal model and hence, tends to be sensitive to errors in the assumed signal model such as microphone mismatch. Several variants of the GSC have been discussed to reduce its sensitivity, such as limiting the adaptation of the ANC to periods of noise-only and/or constraining the noise sensitivity of the GSC (e.g., by imposing a QIC on the ANC). The GSC with ANC adapted during noise-only periods and the more robust GSC with noise sensitivity constraint (NSC-GSC) or quadratic inequality constraint (QIC-GSC) will be used as a reference for the noise reduction algorithms developed in this thesis.

A second category of adaptive multi-microphone noise reduction techniques is referred to as multi-channel Wiener filtering. In this thesis, we concentrate on the MWF with unknown reference: in contrast to fixed beamformers, adaptive beamformers (e.g., the GSC) and other MWF techniques, this MWF technique does not require a priori knowledge about the desired signal model nor calibration. Hence, it seems promising from a robustness point of view. In this chapter, we have developed a subband implementation of this MWF that leads to a significant reduction in computational cost, making it feasible for implementation in future hearing aids. In addition, we have shown that a subband implementation offers an improved intelligibility compared to a fullband realization. In the next chapter, we will compare the MWF and the GSC in terms of performance and robustness for small-sized arrays as used in commercial hearing aids.

2.A Appendix: Adaptive differential microphone array

In this appendix, we show that under the assumption of identical, perfectly omnidirectional microphones, far field propagation and small $fd \ll \frac{c}{2}$, a one-tap adaptive filter w_{a1} in the ADMA (described in Section 2.3.4) suffices to steer a zero to θ_n.

Assuming identical, perfect omnidirectional microphones and far field propagation, the directivity pattern $H(f,\theta)$ of the ADMA in Figure 2.9 can be written as

$$H(f,\theta) = H_f(f,\theta) - W_{a1}(f)H_b(f,\theta) \tag{2.A.1}$$

where $H_f(f,\theta)$ is the spatial directivity pattern of the forward-facing cardioid, i.e.,

$$\begin{aligned} H_f(f,\theta) &= 1 - e^{-j(2\pi f\tau_f + \frac{d}{c}\cos\theta)} \\ &= 2je^{-j\pi f(\tau_f + \frac{d}{c}\cos\theta)}\sin(\pi f(\tau_f + \frac{d}{c}\cos\theta)) \end{aligned} \tag{2.A.2}$$

with $\tau_f = \frac{d}{c}$, and where $H_b(f,\theta)$ is the spatial directivity pattern of the backward facing cardioid

$$\begin{aligned} H_b(f,\theta) &= e^{-j2\pi fd\frac{\cos\theta}{c}} - e^{-j2\pi f\tau_b} \\ &= 2je^{-j\pi f(\tau_b + \frac{d}{c}\cos\theta)}\sin(\pi f(\tau_b - \frac{d}{c}\cos\theta). \end{aligned} \tag{2.A.3}$$

Using (2.A.2) and (2.A.3), we find

$$\begin{aligned} H(f,\theta) &= 2je^{-j\pi f(\tau_b + \frac{d}{c}\cos\theta)}\Bigg[e^{-j\pi f(\tau_f - \tau_b)}\sin(\pi f(\tau_f + \frac{d}{c}\cos\theta)) \\ &\quad -W_{a1}(f)\sin\left(\pi f(\tau_b - \frac{d}{c}\cos\theta)\right)\Bigg] \end{aligned} \tag{2.A.4}$$

For $\tau_b = \frac{d}{c}$, the frequency characteristic $H(f,0^\circ)$ for the direction of interest $\theta = 0^\circ$ is independent of w_{a1} and equals the frequency characteristic $H_f(f,0^\circ)$ of the forward-facing cardioid. From (2.A.4), it follows that the spatial directivity pattern $H(f,\theta)$ of the ADMA has a zero at $\theta = \theta_n$ iff

$$W_{a1}(f) = e^{-j\pi f(\tau_f - \tau_b)}\frac{\sin\left(\pi f\left(\tau_f + \frac{d}{c}\cos\theta_n\right)\right)}{\sin\left(\pi f\left(\tau_b - \frac{d}{c}\cos\theta_n\right)\right)}. \tag{2.A.5}$$

For $\tau_f = \tau_b = \frac{d}{c}$ and small $f \cdot d \ll \frac{c}{2}$, (2.A.5) can be approximated by a scalar w_{a1}

$$w_{a1} = \frac{1+\cos\theta_n}{1-\cos\theta_n}. \tag{2.A.6}$$

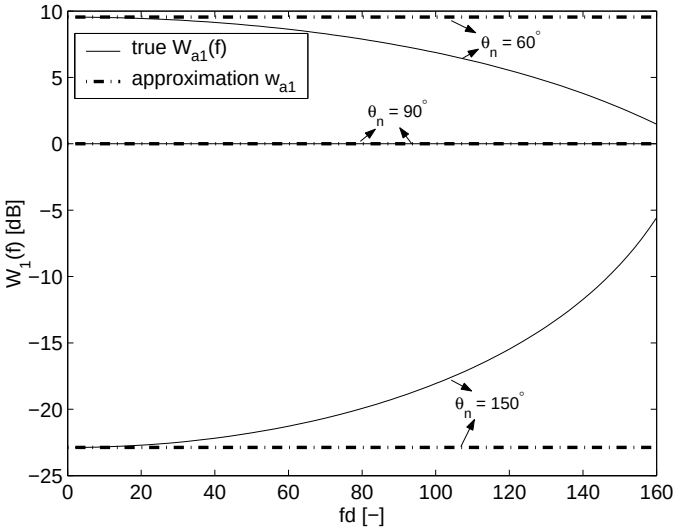


Figure 2.A.14: Comparison of the true $W_{a1}(f)$ (2.A.5) and its approximation by the scalar w_{a1} (2.A.6).

As an illustration, **Figure 2.A.14** compares (2.A.5) and the approximation (2.A.6) as a function of $f \cdot d$ for different angles θ_n. For $f \cdot d < 60$, (2.A.6) seems to be a good approximation of (2.A.5).

Chapter 3

Robustness analysis of GSC and MWF for hearing aids

FOR small-sized arrays such as hearing aids, noise reduction is obtained at the expense of an increased sensitivity to errors in the assumed signal model, i.e. microphone mismatch, variations in speaker and microphone positions, reverberation. However, the noise reduction algorithm should still be robust, i.e., insensitive to small model errors. In this chapter, we evaluate the robustness of the GSC with ANC adapted during noise-only periods (cf. Section 2.3.1), the NSC-GSC (cf. Section 2.3.3) and the MWF (cf. Section 2.4 and Section 2.5) to different model errors both analytically and experimentally. In the analysis, we focus on small-sized arrays as used in hearing aid applications.

In Section 3.2, mathematical expressions are derived to assess the optimal performance, i.e., noise reduction and speech distortion, of the GSC, the NSC-GSC and the MWF.

In Section 3.3, theoretical analysis of the noise sensitivity determines under which conditions (e.g., noise scenario, microphone configuration, etc.) the noise sensitivity increases and hence robustness issues become crucial for the GSC. Especially complicated noise scenarios and low $f \cdot d$ result in a high noise sensitivity. In addition, we theoretically and experimentally show that because of the high noise sensitivity of small-sized arrays, the number of useful microphones in a BTE is limited to three, provided that the noise reduction algorithm is robust enough.

Section 3.4 evaluates the robustness of the GSC, the NSC-GSC and the MWF to different types of errors in the assumed signal model, i.e. microphone mismatch, robustness to microphone placement uncertainty and robustness to variations

in speaker location. Among the different signal model errors microphone gain mismatch in particular is found to be harmful to the GSC, even if the ANC is adapted during noise-only periods. As a consequence, a noise sensitivity constraint is essential, which goes at the expense of less noise reduction. The MWF on the other hand, is not degraded by microphone mismatch, so that compared to a robust NSC-GSC, more noise is reduced and the benefit of a third microphone in a BTE can be fully exploited.

Section 3.5 and Section 3.6 study the sensitivity of the GSC, the NSC-GSC and the MWF to inaccurate estimation of the second order statistics, e.g., caused by non-stationarities in the noise or errors by the speech detection algorithm. Simulations indicate that the MWF preserves its benefit over the NSC-GSC, provided that the spatial characteristics and long-term average PSD of the noise vary relatively slowly in time (on the order of seconds) and that the speech detection error rate is smaller than 20 %.

The theoretical analysis assumes stationary signals, free field propagation of speech and noise sources and infinite filter lengths. In practice, finite filter lengths are used, reverberation is present and signals are non-stationary. In addition, the second order statistics of speech + noise and noise-only are estimated from a finite amount of incoming data and at different moments in time. Therefore, the major theoretical results are validated for real-data recordings with the three-microphone BTE mounted on a dummy head.

3.1 Preliminaries

In hearing aids, the array length is limited because of aesthetical and practical reasons. In [198, 245], it is shown that noise can be reduced even at low $f \cdot d$, i.e. $f \cdot d \ll \frac{c}{2}$, be it at the expense of an increased noise sensitivity (cf. Section 2.1.4) and hence an increased sensitivity to errors in the assumed signal model. The increased noise sensitivity shows that the *robustness* of multi-microphone noise reduction algorithms may be crucial for practical use in hearing aids.

A widely studied, adaptive noise reduction algorithm for hearing aids is the LCMV beamformer [198, 74, 96, 131, 140, 141, 157, 263, 261, 271], which is often implemented using the GSC [77] (cf. Section 2.3.1). The LCMV beamformer and GSC rely on a priori assumptions about the desired signal model (e.g, the location of the desired speaker, perfectly calibrated microphones, no reverberation). As a result, they tend to be sensitive to small errors in these assumptions when used for small-sized arrays such as hearing aids. To reduce the sensitivity of the GSC to signal model errors, different variants of the GSC have been proposed in the literature (cf. Section 2.3.3). Common techniques are to adapt the ANC during periods of noise-only [74, 92, 258] and/or to constrain the noise sensitivity by applying a quadratic inequality constraint to the

ANC [29, 92, 96, 103, 157] or by injection of spatially uncorrelated noise [108]. In the remainder, we therefore concentrate on the widely studied *standard GSC with ANC adapted during noise-only periods* with and without noise sensitivity constraint.

In contrast to the GSC and other MWF techniques, the MWF with unknown reference [33, 35, 36, 212, 235] described in Section 2.4 does not require a priori information about the desired signal model nor calibration. This MWF provides an MMSE estimate of the (reverberant) speech component in one of the microphone signals. It is uniquely based on estimates of the second order statistics of the speech and the noise signal. To obtain these estimates, the MWF assumes that the second order statistics of the noise are sufficiently stationary so that they can be estimated during periods of noise-only. Like for the GSC, a robust speech detection is thus needed. Simulations with a 5-microphone array of 0.2 m showed that the GSVD based algorithm is indeed more robust than the GSC [35] in case of significantly violated assumptions about the desired signal model. Compared to the GSC, no or a less severe constraint on the noise sensitivity is expected to be needed so that more noise may be reduced. However, since the MWF does not rely on a priori information, it is expected to be sensitive to inaccurate estimation of the second order statistics caused by e.g., non-stationarities in the noise or speech detection errors.

Analytical calculations have drawn the attention of many researchers to predict the optimal (i.e., stationary signals, infinite filter length) *noise reduction performance* achievable by single and multi-microphone noise reduction systems [12, 13, 14, 27, 66, 162, 192, 198, 260, 266]. In [13, 14, 192, 198], the theoretically optimal noise reduction performance of the GSC has been derived as a function of the complex coherence functions between the microphone noise signals. In [14, 162], a similar theoretical analysis evaluates a microphone array with post-filtering. In [27, 108, 260, 266] the sensitivity of the MVDR beamformer[1] (cf. Section 2.3.2) and GSC against signal model errors has been analytically investigated for antenna array applications. Most of these analyses focus on an MVDR beamformer that minimizes the total output power and not the output noise power: then, performance degradation due to signal model errors especially occurs for high input SNR [27, 96, 108, 199, 260, 266]. The sensitivity to errors is significantly reduced when only the output noise power is minimized [260], which is obtained by the GSC with ANC adapted during noise-only periods.

In this chapter, we present a unified theoretical framework based on the complex coherence functions to assess both the *broadband performance* of the GSC with ANC adapted during noise-only periods, the NSC-GSC and the MWF and their *robustness* to different types of model errors (i.e., errors in the assumed

[1] Note that the MVDR beamformer $\mathbf{W}(f)$ can be considered as a special case of LCMV beamforming or GSC with the distortionless constraint $\mathbf{W}^H(f)\mathbf{d}(f,\theta) = 1$, where θ refers to the target signal direction (cf. Section 2.3.2)[261].

signal model as well as inaccurate estimation of the statistics) for different noise scenarios. The robustness is analyzed for *small-sized arrays as used in hearing aids and in terms of speech intelligibility weighted performance measures*. The theoretical results are verified through experiments with recorded data.

3.1.1 Assumptions of the theoretical analysis

In the theoretical analysis, we assume *stationary signals* and *infinite filters* so that the signals and filters can be represented in the frequency domain. Unless stated otherwise, we assume perfect estimation of the second order statistics $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}\}$ of the speech and the noise.

For the time being, we assume a *homogeneous speech and noise field*, i.e., that the PSD of the received speech, $P_X^s(f)$, the internal and external noise $P_X^i(f)$ and $P_X^e(f)$, the received noise, $P_X^n(f) = P_X^i(f) + P_X^e(f)$, and the received microphone signals $P_X(f) = P_X^s(f) + P_X^n(f)$ are the same at each microphone so that the expressions (2.24)-(2.27) of Section 2.1.3 apply, i.e.,

$$\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} = P_X(f)\mathbf{\Gamma}(f); \tag{3.1}$$

$$\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} = P_X^s(f)\mathbf{\Gamma}^s(f); \tag{3.2}$$

$$\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} = P_X^n(f)\mathbf{\Gamma}^n(f). \tag{3.3}$$

In Section 3.4.1, the coherence matrices $\mathbf{\Gamma}(f)$, $\mathbf{\Gamma}^n(f)$ and $\mathbf{\Gamma}^s(f)$ in (2.24)-(2.27) are modified to include spectral differences between the microphone signals.

To simplify the theoretical analysis, we assume a *linear endfire microphone array with the desired speaker at* 0° (cf. Section 2.1.3). We assume *free and far field propagation and perfect omnidirectional microphones* so that the mathematical expressions (2.35)-(2.36) of Section 2.1.3 can be used in combination with (2.28)-(2.31) to compute the coherence of the speech and the noise.

3.1.2 Experimental validation: configuration and implementation issues

The theoretical analysis assumes stationary signals, free field propagation of speech and noise sources, infinite filter lengths and, apart from Section 3.5 and Section 3.6, perfect estimation of the second order statistics of speech and noise. In practice, finite filter lengths are used, reverberation is present and signals are non-stationary. In addition, the second order statistics of speech + noise and noise-only are estimated from a finite amount of incoming data and at a different moment in time. Therefore, we will verify the major theoretical results based on real-data recordings in a reverberant, office room and when finite filters are used. In most experiments, we assume a perfect speech detection. The sensitivity to speech detection errors is analyzed in Section 3.6.

Configuration

Unless stated otherwise, the recordings for the experimental validation have been made with the 3-microphone BTE mounted on the dummy head (cf. Section 2.1.5).

Implementation issues

To compensate for possible microphone mismatch, the fixed stage of the GSC, i.e., the fixed DS beamformer and the blocking matrix (created as in Example 2.1 of Section 2.3.1), has been optimized for the BTE (or array) mounted on the dummy head. To model the delays in the DS beamformer and the blocking matrix, FIR filters are used. The FIR filters are designed so that they minimize the energy of the noise references for a signal from 0°, recorded in an anechoic room (cf. Section 2.2.4). To improve intelligibility, the microphone signals are pre-whitened in the GSC prior to processing by means of a so-called pre-emphasis filter [150] (cf. Remark 2.12 of Section 2.5.1). As a pre-emphasis filter, we use a high pass filter that models the inverse of the long-term average speech spectrum. At the GSC output, the effect of the pre-emphasis filter on the desired signal spectrum is cancelled by filtering the output with the inverse of the pre-emphasis filter. The ANC filters are implemented using exponentially weighted RLS with forgetting factor λ [85, 173] (cf. Section 6.2.2). The performance of the NSC-GSC is obtained by injecting spatially uncorrelated noise in the pre-emphasized microphone signals that are used to compute the filters.

For the MWF, the recursive subband GSVD based implementation described in Section 2.5.2 is used with 16 subbands, a downsampling ratio $D = 12$ and forgetting factors $\lambda_{s+n}^D = \lambda_n^D$ where λ_{s+n} and λ_n are the equivalent forgetting factors in a fullband implementation. The subband implementation intrinsically performs a pre-whitening operation (cf. Section 2.5). The filters of the GSC and the MWF are sufficiently long so that a further increase in filter length does not result in a significant increase in performance.

3.2 Theoretical performance analysis

In this section, we derive mathematical expressions for the performance of the GSC (cf. Section 2.3.1), the GSC with noise sensitivity constraint (NSC-GSC) (cf. Section 2.3.3) and the MWF (cf. Section 2.4) as a function of the complex coherence matrices $\mathbf{\Gamma}^n(f)$ and $\mathbf{\Gamma}^s(f)$ of the noise and the speech (cf. Section 2.1.3). In the forthcoming sections, these expressions will be used to theoretically predict the robustness of the GSC and the MWF against different model errors for a small-sized linear endfire array. In addition, we describe the microphone configuration and implementation of the GSC and MWF that have been used in the experimental validations.

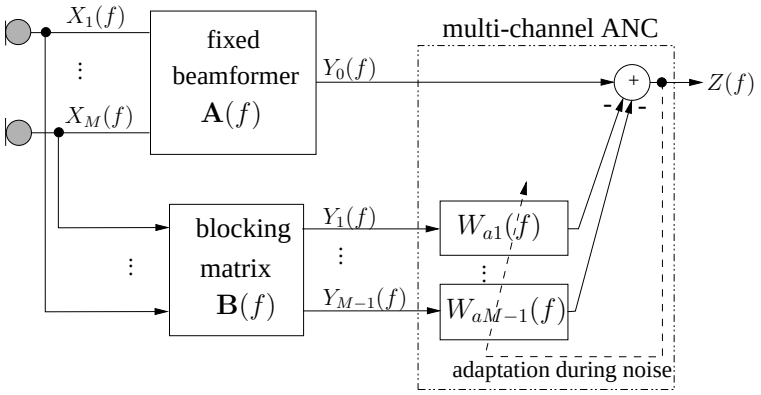


Figure 3.1: Generalized Sidelobe Canceller in the frequency-domain.

Below, we derive theoretical formulas for PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ of the speech, the noise and the internal noise from the first microphone to the output $Z(f)$ of the noise reduction algorithms using (2.38)-(2.40) of Section 2.1.4. From the PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ and the input PSDs $P_X^s(f)$ and $P_X^n(f)$, the noise sensitivity and the broadband intelligibility weighted performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ (cf. Section 2.1.4) can then be computed using (2.44) and (2.45)-(2.49). Since in practice, the desired and interfering signals are often speech-like, the improvement $\Delta\text{SNR}_{\text{intellig}}$ will be computed using the long-term average speech PSD of the signals found on [106] for $P_X^s(f)$ and $P_X^n(f)$.

3.2.1 GSC with and without noise sensitivity constraint

In [13, 192], the theoretical noise reduction $NR(f) = 1/G^n(f)$ of the GSC has been expressed in terms of the complex coherence $\Gamma_{kl}(f)$ between the k-th and l-th microphone signals. In addition to the PTF of the noise signal $G^n(f)$, also the PTF of the desired speech signal $G^s(f)$ and the internal noise $G^i(f)$ can be expressed as a function of $\Gamma_{kl}(f)$. These results are reviewed in this section. In addition, the formulas for the more robust NSC-GSC are derived.

Figure 3.1 reviews the structure of the GSC in the frequency-domain (cf. Section 2.3.1). For simplicity, but without loss of generality, we assume that there is no delay Δ in the speech reference path (cf. Figure 2.8 of Section 2.3.1). The $M \times 1$ fixed beamformer $\mathbf{A}(f)$ creates a speech reference

$$Y_0(f) = \mathbf{A}^H(f)\mathbf{X}(f), \tag{3.4}$$

and the blocking matrix $\mathbf{B}(f)$ creates $M-1$ noise references

$$Y_l(f) = \mathbf{B}^H(l,:)\mathbf{X}(f),\ l = 1, \cdots, M-1. \tag{3.5}$$

To avoid possible speech cancellation, the ANC $\mathbf{W}_a(f)$ is adapted during periods of noise-only [74, 92, 103, 258]. In the sequel, we use a DS beamformer, i.e.,

$$\mathbf{A}(f) = \frac{1}{M}\left[\begin{array}{cccc} 1 & e^{j2\pi f\frac{d_2-d_1}{c}} & \cdots & e^{j2\pi f\frac{d_M-d_1}{c}} \end{array}\right]^H \tag{3.6}$$

with $d_l - d_1$ the interspacing between the l-th and the first microphone. For e.g., $M = 3$, the $M \times (M-1)$ blocking matrix $\mathbf{B}$ is set to (cf. Example 2.1 in Section 2.3.1)

$$\mathbf{B} = \left[\begin{array}{ccc} 1 & -e^{j2\pi f\frac{d_2-d_1}{c}} & 0 \\ 1 & 0 & -e^{j2\pi f\frac{d_3-d_1}{c}} \end{array}\right]^H. \tag{3.7}$$

PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ of the GSC

The PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ of the GSC can be expressed as a function of $\mathbf{\Gamma}^s(f)$ and $\mathbf{\Gamma}^n(f)$, the fixed beamformer $\mathbf{A}(f)$ and the blocking matrix $\mathbf{B}(f)$. Define the stacked vector $\mathbf{Y}(f)$ as

$$\mathbf{Y}(f) = \left[\begin{array}{cccc} Y_1(f) & Y_2(f) & \cdots & Y_{M-1}(f) \end{array}\right]^T. \tag{3.8}$$

The ANC $\mathbf{W}_a(f) \in \mathbb{C}^{M-1\times 1}$

$$\mathbf{W}_a(f) = \left[\begin{array}{cccc} W_{a1}(f) & W_{a2}(f) & \cdots & W_{aM-1}(f) \end{array}\right]^H \tag{3.9}$$

minimizes the output noise power $P_Z^n(f)$, i.e.,

$$\begin{aligned} \mathbf{W}_a(f) &= \arg\min_{\mathbf{W}_a(f)} \mathcal{E}\{\left|Y_0(f) - \mathbf{W}_a^H(f)\mathbf{Y}(f)\right|^2\} \\ &= \arg\min_{\mathbf{W}_a(f)} \mathcal{E}\{\left|\mathbf{A}^H(f)\mathbf{X}(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\mathbf{X}(f)\right|^2\} \end{aligned} \tag{3.10}$$

(cf. Section 2.3.1), leading to:

$$\mathbf{W}_a(f) = \mathcal{E}\{\mathbf{B}^H(f)\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\mathbf{B}(f)\}^{-1}\mathcal{E}\{\mathbf{B}^H(f)\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\mathbf{A}(f)\} \tag{3.11}$$

$$\boxed{\mathbf{W}_a(f) = \left(\mathbf{B}^H(f)\mathbf{\Gamma}^n(f)\mathbf{B}(f)\right)^{-1}\mathbf{B}^H(f)\mathbf{\Gamma}^n(f)\mathbf{A}(f).} \tag{3.12}$$

For a homogeneous noise field and infinite filter lengths, the ANC $\mathbf{W}_a(f)$ thus only depends on the spatial characteristics of the noise $\mathbf{\Gamma}^n(f)$ and not on $P_X^n(f)$.

The ANC output $Z(f)$ equals

$$Z(f) = \left(\mathbf{A}^H(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\right)\mathbf{X}(f). \tag{3.13}$$

Hence, from (2.22) and (2.38)-(2.40), it follows that the PTFs $G^s(f)$, $G^n(f)$ and $G^i(f)$ can be expressed as a function of $\mathbf{\Gamma}^s(f)$ and $\mathbf{\Gamma}^n(f)$ as:

$$G^s(f) = \left(\mathbf{A}^H(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\right)\mathbf{\Gamma}^s(f)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}_a(f)\right), \tag{3.14}$$

$$G^n(f) = \left(\mathbf{A}^H(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\right)\mathbf{\Gamma}^n(f)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}_a(f)\right), \tag{3.15}$$

$$G^i(f) = \left(\mathbf{A}^H(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\right)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}_a(f)\right). \tag{3.16}$$

In [33, 116], it has been shown that if $\mathbf{B}^H(f)\mathbf{A}(f) = \mathbf{0}$ and $\begin{bmatrix}\mathbf{B}(f) & \mathbf{A}(f)\end{bmatrix}$ is a non-singular matrix of dimensions $M \times M$,

$$\mathbf{A}(f) - \mathbf{W}_a(f)\mathbf{B}(f) = \frac{\mathbf{A}^H(f)\mathbf{A}(f)}{\mathbf{A}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f)}\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f). \tag{3.17}$$

For $\mathbf{A}(f) = \frac{\mathbf{H}(f)}{\mathbf{H}^H(f)\mathbf{H}(f)}$ with $\mathbf{H}(f)$ defined in (2.19) of Section 2.1.3, the filter $\mathbf{A}(f) - \mathbf{W}_a(f)\mathbf{B}(f)$ corresponds to an *optimum* MVDR beamformer that minimizes the output noise power subject to a distortionless response for the desired source signal $S_d(f)$ (cf. Section 2.3.2).

PTFs $\tilde{G}^s(f)$, $\tilde{G}^n(f)$ and $\tilde{G}^i(f)$ of the NSC-GSC

To increase the robustness of the GSC, a constraint on the noise sensitivity is often applied [29, 96, 103, 108] (cf. NSC-GSC in Section 2.3.3). In [29], it is shown that for $\mathbf{B}^H(f)\mathbf{A}(f) = 0$, a constraint on the noise sensitivity $\Psi(f)$ corresponds to a quadratic inequality constraint on the filters $\mathbf{B}(f)\mathbf{W}_a(f)$, i.e.

$$\mathbf{W}_a^H(f)\mathbf{B}^H(f)\mathbf{B}(f)\mathbf{W}_a(f) \leq \zeta^2. \tag{3.18}$$

In [108], a similar constraint is imposed by inserting spatially uncorrelated noise in the microphone signals[2] used to design the filters. In [29], it is demonstrated that the method of noise injection is equivalent with a quadratic inequality constraint for a certain level of noise injection and a certain value ζ^2. Hence, the PTFs $\tilde{G}^s(f)$, $\tilde{G}^n(f)$ and $\tilde{G}^i(f)$ of the NSC-GSC are obtained by replacing $\mathbf{\Gamma}^n(f)$ in the computation of $\mathbf{W}_a(f)$ with

$$\mathbf{\Gamma}^n(f) + \eta(f)\mathbf{I}_M, \tag{3.19}$$

where $\eta(f)$ is the ratio of the injected noise power to the input noise power $P_X^n(f)$:

$$\tilde{G}^s(f) = \left(\mathbf{A}^H(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\right)\mathbf{\Gamma}^s(f)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}_a(f)\right), \tag{3.20}$$

$$\tilde{G}^n(f) = \left(\mathbf{A}^H(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\right)\mathbf{\Gamma}^n(f)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}_a(f)\right), \tag{3.21}$$

$$\tilde{G}^i(f) = \left(\mathbf{A}^H(f) - \mathbf{W}_a^H(f)\mathbf{B}^H(f)\right)\left(\mathbf{A}(f) - \mathbf{B}(f)\mathbf{W}_a(f)\right). \tag{3.22}$$

[2] Since $\mathbf{B}^H(f)\mathbf{A}(f) = 0$, this corresponds to adding a diagonal loading term to the coherence matrix $\mathbf{\Gamma}^n(f)$ of the noise in the denominator of (3.12).

with

$$\boxed{\mathbf{W}_a(f) = \left[\mathbf{B}^H(f)\left(\mathbf{\Gamma}^n(f) + \eta(f)\mathbf{I}_M\right)\mathbf{B}(f)\right]^{-1}\mathbf{B}^H(f)\left(\mathbf{\Gamma}^n(f) + \eta(f)\mathbf{I}_M\right)\mathbf{A}(f).} \tag{3.23}$$

Given the GSC parameters $\mathbf{A}(f)$ and $\mathbf{B}(f)$, the coherences $\mathbf{\Gamma}^s(f)$, $\mathbf{\Gamma}^n(f)$ and the PSD $P_X^s(f)$, $P_X^n(f)$, the PTFs (3.14)-(3.16) and (3.20)-(3.22) and hence the intelligibility weighted performance measures (2.45) and (2.48) of the GSC and the NSC-GSC can be computed.

3.2.2 Multi-channel Wiener Filter (MWF)

Similarly, the theoretically optimal performance of the MWF, i.e., the PTFs $G^s(f)$,$G^n(f)$ and $G^i(f)$ as well as the intelligibility weighted performance measures, can be expressed as a function of $\mathbf{\Gamma}^n(f)$ and $\mathbf{\Gamma}^s(f)$.

The MWF $\mathbf{W}(f)$

$$\mathbf{W}(f) = \begin{bmatrix} W_1(f) & W_2(f) & \cdots & W_M(f) \end{bmatrix}^H \tag{3.24}$$

that provides an MMSE estimate $Z(f) = \mathbf{W}^H(f)\mathbf{X}(f)$ of the speech component[3] $X_1^s(f)$ in the first microphone, can be written as (cf. (2.103)):

$$\mathbf{W}(f) = \mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}^{-1}\left(\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}\right)\mathbf{e}_1 \tag{3.25}$$

where the $M \times 1$ vector $\mathbf{e}_1$ equals the first canonical vector defined as $\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$.

For the time being, we assume perfect estimation of the second order statistics of speech and noise, i.e.,

$$\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} = \mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}. \tag{3.26}$$

The sensitivity of the MWF to inaccurate estimation of the speech + noise and noise statistics is analyzed later in Section 3.5 and Section 3.6. Using (3.1)-(3.3) and (3.26), formula (3.25) can be rewritten as:

$$\boxed{\mathbf{W}(f) = \frac{P_X^s(f)}{P_X^s(f) + P_X^n(f)}\mathbf{\Gamma}^{-1}(f)\mathbf{\Gamma}^s(f)\mathbf{e}_1,} \tag{3.27}$$

where, using (2.30),

$$\mathbf{\Gamma}(f) = \frac{\mathbf{\Gamma}^s(f) + \alpha(f)\mathbf{\Gamma}^n(f)}{1 + \alpha(f)}, \tag{3.28}$$

$$\alpha(f) = \frac{P_X^n(f)}{P_X^s(f)}. \tag{3.29}$$

[3] In the derivation of $\mathbf{W}(f)$, we assume, without loss of generality, that the delay $\Delta = 0$.

In contrast to the GSC, the filter $\mathbf{W}(f)$ not only depends on the spatial characteristics of the noise $\mathbf{\Gamma}^n(f)$, but also on the input noise-to-signal ratio $\alpha(f)$ and the spatial characteristics of the speech signal $\mathbf{\Gamma}^s(f)$. Note that the ratio $\frac{P_X^s(f)}{P_X^s(f)+P_X^n(f)}$ in (3.27) corresponds to a *spectral* Wiener filter, which results in additional performance if the spectra of the desired and undesired signal differ. Hence, (3.27) truly combines spatial and spectral filtering.

Recall from Section 2.1.4 that the PTFs $G^n(f)$, $G^s(f)$ and $G^i(f)$ can be expressed as a function of $\mathbf{W}(f)$ and $\mathbf{\Gamma}^s(f)$ and $\mathbf{\Gamma}^n(f)$ as:

$$\boxed{\begin{aligned} G^s(f) &= \mathbf{W}^H(f)\mathbf{\Gamma}^s(f)\mathbf{W}(f), && (3.30)\\ G^n(f) &= \mathbf{W}^H(f)\mathbf{\Gamma}^n(f)\mathbf{W}(f), && (3.31)\\ G^i(f) &= \mathbf{W}^H(f)\mathbf{W}(f). && (3.32) \end{aligned}}$$

Single speech source

In the presence of a *single speech source* with PSD $P_{S_d}(f)$, $\mathbf{\Gamma}^s(f)$ is rank-one and equals

$$\mathbf{\Gamma}^s(f) = \frac{P_{S_d}(f)}{P_X^s(f)}\mathbf{H}(f)\mathbf{H}^H(f) \tag{3.33}$$

(cf. (2.28) of Section 2.1.3). Using (3.33) and the matrix inversion lemma [69], we obtain

$$(\mathbf{\Gamma}^s(f)+\alpha(f)\mathbf{\Gamma}^n(f))^{-1} = \frac{1}{\alpha(f)}\left(\mathbf{\Gamma}^{n^{-1}} - \frac{\mathbf{\Gamma}^{n^{-1}}\mathbf{H}(f)\mathbf{H}^H(f)\mathbf{\Gamma}^{n^{-1}}}{\mathbf{H}(f)^H\mathbf{\Gamma}^{n^{-1}}\mathbf{H}(f)+\frac{P_X^n(f)}{P_{S_d}(f)}}\right), \tag{3.34}$$

so that $\mathbf{W}(f)$ in (3.27) can be re-written as

$$\boxed{\mathbf{W}(f) = \underbrace{\Lambda(f)\left(P_X^n(f)\mathbf{\Gamma}^n(f)\right)^{-1}\mathbf{H}(f)}_{\mathbf{W}_m(f)}\underbrace{\left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)}_{W_c(f)}H_1^*(f),} \tag{3.35}$$

with

$$\Lambda(f) = \left[\mathbf{H}(f)^H(P_X^n(f)\mathbf{\Gamma}^n(f))^{-1}\mathbf{H}(f)\right]^{-1}. \tag{3.36}$$

The MWF $\mathbf{W}(f)$ in (3.35) can be decomposed into three parts:

- The *multi-channel filter* $\mathbf{W}_m(f)$ corresponds to an optimum MVDR beamformer with steering vector $\mathbf{H}(f)$ (cf. Section 2.3.2) and hence provides a distortionless response for the target signal $S_d(f)$, i.e.,

$$\mathbf{W}_m^H(f)P_X^s(f)\mathbf{\Gamma}^s(f)\mathbf{W}_m(f) = P_{S_d}(f). \tag{3.37}$$

The residual noise at the output of the filter $\mathbf{W}_m(f)$ equals

$$\mathbf{W}_m^H(f)P_X^n(f)\mathbf{\Gamma}^n(f)\mathbf{W}_m(f) = \Lambda(f). \tag{3.38}$$

- The *single-channel filter* $W_c(f)$ corresponds to a single-channel Wiener filter at the output of the MVDR $\mathbf{W}_m(f)$. This filter increases noise suppression at the expense of distortion of the target signal. Note that the cascade $\mathbf{W}_m(f)W_c(f)$ corresponds to an MWF that provides an MMSE estimate of the source speech signal $S_d(f)$ [220].
- The *single-channel filter* $H_1^*(f)$ translates the MMSE estimate of the source signal $S_d(f)$ to an MMSE estimate of the speech component $X_1^s(f)$ in the first microphone signal.

Using (3.35), the PTF $G^n(f)$, $G^s(f)$ and $G^i(f)$ in (3.30) and (3.32) can be simplified to:

$$G^s(f) = \left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)^2, \tag{3.39}$$

$$G^n(f) = |H_1(f)|^2\left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)^2\frac{\Lambda(f)}{P_X^n(f)}, \tag{3.40}$$

$$G^i(f) = |H_1(f)|^2\left(\frac{P_{S_d}(f)}{P_{S_d}(f)+\Lambda(f)}\right)^2\frac{\mathbf{H}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}(f)}{\left(\mathbf{H}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}(f)\right)^2} \tag{3.41}$$

Speech distortion occurs at frequencies f where the SNR $P_{S_d}(f)/P_X^n(f)$ is low and where the amount of spatial noise reduction $P_X^n(f)/\Lambda(f)$ is limited, i.e., for small $f \cdot d$ and difficult noise scenarios (e.g., when the spatial separation between the speech and the noise sources is poor [189]). Theoretically, the MWF achieves the same SNR improvement $\frac{G^s(f)}{G^n(f)}$ w.r.t. the first microphone signal as an optimum MVDR beamformer steered to $\mathbf{H}(f)$. Hence, in the absence of errors in the signal model assumed by GSC (such as microphone mismatch and reverberation), the GSC and the MWF achieve the same SNR improvement $\frac{G^s(f)}{G^n(f)}$.

3.3 Analysis of the noise sensitivity

In contrast to the GSC, the MWF does not require a priori information about the desired signal model: theoretically, the MWF consists of an optimum MVDR cascaded with a single-channel spectral filter (see (3.35)) and is thus expected to be more robust than the GSC to signal model errors. To quantify the sensitivity to errors in the assumed signal model, the *noise sensitivity*

is often used [29, 198, 245] (cf. Section 2.1.4). In this section, we study the conditions (noise scenario, microphone configuration) under which the noise sensitivity of the GSC and the MWF is high and hence robustness of the GSC against signal model errors is expected to be crucial. In addition, we will show that because of the high noise sensitivity of small-sized arrays, the number of useful microphones in a commercial BTE is limited to three, provided that the noise reduction algorithm is sufficiently robust. The results are illustrated for a linear microphone endfire array with the dimensions of a BTE.

3.3.1 Noise sensitivity of GSC and MWF

Using (3.14), (3.16), (3.30) and (3.32), the noise sensitivity $\Psi(f)$ of the GSC and the MWF is derived as:

$$\boxed{\Psi(f) = \frac{G^i(f)}{G^s(f)}.} \tag{3.42}$$

Equations (3.16) and (3.32) for the internal noise PTF $G^i(f)$ show that, for both algorithms, large filter amplitudes $|W_l(f)|$ and $|W_{al}(f)|$, $l = 1, \cdots, M$ with respect to $G^s(f)$ correspond to a high noise sensitivity $\Psi(f)$.

Using (3.17) in (3.14) and (3.16), $\Psi_{\text{GSC}}(f)$ of the GSC can be written as:

$$\Psi_{\text{GSC}}(f) = \frac{\mathbf{A}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f)}{\mathbf{A}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{s}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{A}(f)}. \tag{3.43}$$

The noise sensitivity $\Psi_{MWF}(f)$ of the MWF can be derived from (3.39) and (3.41):

$$\Psi_{MWF} = |H_1(f)|^2 \frac{\mathbf{H}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}(f)}{\left(\mathbf{H}^H(f)\mathbf{\Gamma}^{n^{-1}}(f)\mathbf{H}(f)\right)^2}. \tag{3.44}$$

Using (3.43) and (3.44), it can be shown that *in the case of a single speech source*, i.e., $\mathbf{\Gamma}^{\boldsymbol{s}}(f) = P_{S_d}(f)\mathbf{H}^H(f)\mathbf{H}(f)/P_X^s(f)$, the noise sensitivity of the GSC and MWF coincide [33, Chapter 3], i.e.,

$$\boxed{\Psi_{\text{GSC}}(f) = \Psi_{\text{MWF}}(f),} \tag{3.45}$$

if the fixed beamformer $\mathbf{A}(f)$ *of the GSC equals* $\mathbf{A}(f) = c(f)\mathbf{H}(f)$, where $c(f)$ *is a frequency-dependent scalar.* Hence, in the absence of errors in the signal model assumed by the GSC, i.e., $\mathbf{A}(f) = |c|e^{-j2\pi f\tau}\mathbf{H}(f)$, with c and τ (frequency-independent) scalars, the noise sensitivity of the GSC and MWF coincide.

For both algorithms, the noise sensitivity$\Psi(f)$ depends on the coherence matrix $\mathbf{\Gamma}^n(f)$ of the noise field and hence on the noise scenario and the microphone configuration.

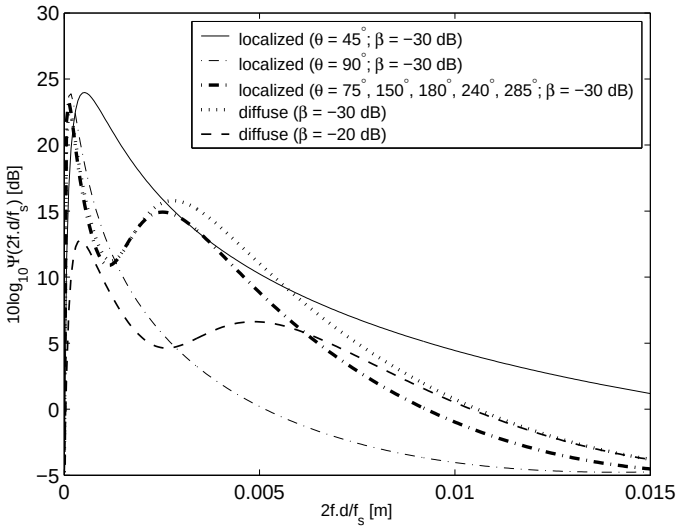


Figure 3.2: Influence of the noise scenario on the noise sensitivity $\Psi_{\mathrm{GSC}}(f) = \Psi_{\mathrm{MWF}}(f)$ (in dB). The sampling frequency $f_s = 16$ kHz.

3.3.2 Influence of the noise scenario

Figure 3.2 investigates the influence of the noise scenario on the noise sensitivity $\Psi(f)$ of the GSC and the MWF. Remember that in the case of a single speech source and in the absence of signal model errors, both algorithms result in the same noise sensitivity (cf. Section 3.3.1). The sensitivity $\Psi(f)$ (in dB) is plotted as a function of $\frac{2f}{f_s}d$ (in meters) with d the microphone interspacing and $f_s = 16$ kHz the sampling frequency, for a linear, uniform 3-microphone endfire array. Five different noise scenarios are shown: a localized single noise source at 45° and 90° with respect to the speech source, both with internal-to-external noise ratio $\beta = -30$ dB, a multiple noise source scenario consisting of five localized noise sources at 75°, 150°, 180°, 240°, 285° with the same PSD and a diffuse noise field with $\beta = -20$ dB and $\beta = -30$ dB. Typical values for $\beta(f)$ lie between -20 dB and -40 dB [14]. Figure 3.2 shows that the noise sensitivity $\Psi(f)$ grows with decreasing $f \cdot d$, decreasing internal-to-external noise ratio $\beta(f)$ and increasing complexity of the noise scenario. Diffuse noise, multiple noise sources and noise sources close to the speech source result in a high $\Psi(f)$ compared to, e.g., a single, localized noise source at 90°.

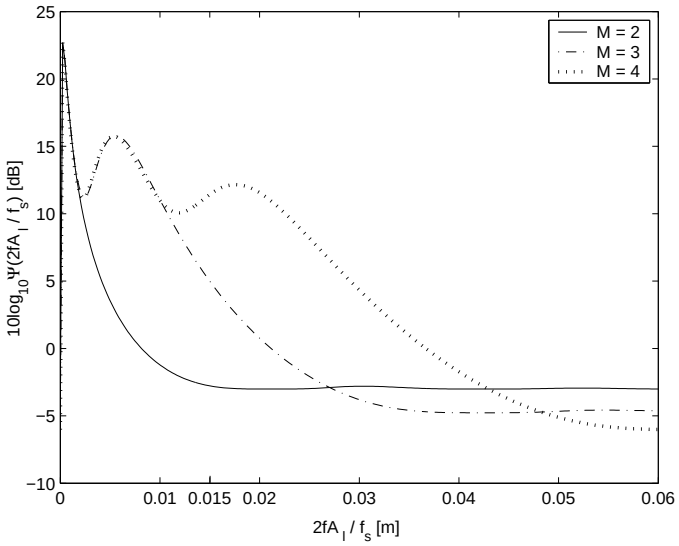


Figure 3.3: Influence of the number of microphones M on the noise sensitivity $\Psi_{\mathrm{GSC}}(f) = \Psi_{\mathrm{MWF}}(f)$ (in dB) for the GSC and the MWF in a diffuse noise field with $\beta(f) = -30$ dB. The sampling frequency $f_s = 16$ kHz.

3.3.3 Influence of the number of microphones

Noise sensitivity as a function of frequency

Figure 3.3 depicts the influence of the number of microphones M per array length $A_l = (M-1)d$ on the noise sensitivity $\Psi(f)$ of the GSC and the MWF for a diffuse noise field with $\beta(f) = -30$ dB. To illustrate the effect of the array length A_l, the noise sensitivity is plotted as a function of $\frac{2f}{f_s}A_l$ (in meters) with $f_s = 16$ kHz. For small-sized arrays (i.e. small $\frac{2f}{f_s}A_l$), an increase of the number of microphones per array length corresponds to an equal or higher noise sensitivity $\Psi(f)$. Hence, the larger the number of microphones per array length, the more robust the used noise reduction algorithm will have to be.

Performance as a function of frequency

Because of the high noise sensitivity of small-sized arrays, the number of useful microphones in a BTE is limited. To illustrate this, **Figure 3.4** shows the influence of the number of microphones M per array length $A_l = (M-1)d$ on the SNR improvement $\frac{G^s(f)}{G^n(f)}$ of the GSC and the MWF for a diffuse noise field with $\beta(f) = -30$ dB. Recall from Section 3.3.1 that in the case of a single speech source and in the absence of signal model errors, the performance $\frac{G^s(f)}{G^n(f)}$ of the GSC and the MWF are the same.

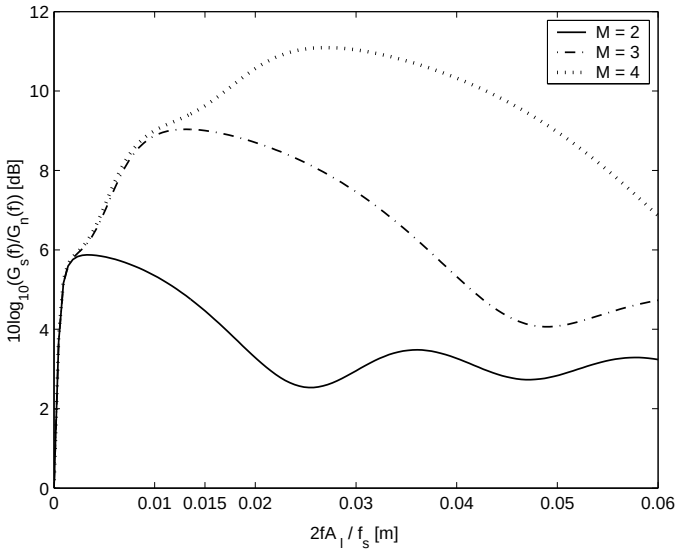


Figure 3.4: Influence of the number of microphones M on the improvement in SNR $\frac{G^s(f)}{G^n(f)}$ (in dB) for the GSC and the MWF in a diffuse noise field with $\beta(f) = -30$ dB ($f_s = 16$ kHz).

For small-sized arrays (i.e., at low $\frac{2f}{f_s}A_l$), external noise is reduced at the expense of internal noise amplification: there is indeed an overall noise reduction $NR(f)$ that optimally trades off internal noise amplification $G^i(f)$ and external noise reduction. The maximum internal noise amplification and hence the external noise reduction will be lower for high internal-to-external ratio $\beta(f)$ (cf. Figure 3.2). At very low $\frac{2f}{f_s}A_l$, the noise sensitivity and hence also the improvement in SNR reach a maximum and do not further increase by adding more microphones. In Figure 3.3 and Figure 3.4, this maximum is reached at low $\frac{2f}{f_s}A_l$, below 0.01 m. As a consequence, the number of useful microphones per array length is limited by the internal-to-external noise ratio $\beta(f)$ [198] and by the noise sensitivity level $\Psi(f)$ tolerated by the noise reduction algorithm (see Section 3.4). For an array with a length $A_l = 3$ cm, which is similar to the dimensions of a commercial BTE hearing aid, no significant SNR improvement is obtained with an additional fourth microphone at frequencies below 4 kHz, i.e., for $\frac{2fA_l}{f_s} \leq 0.015\,\text{m}$.

Intelligibility weighted performance

To illustrate the effect of the number of microphones on the broadband performance, **Table 3.1** gives the improvement in $\text{SNR}_{\text{intellig}}$ as a function of the number of microphones M per array length A_l for the *diffuse noise scenario* and a *multiple noise source scenario* (i.e., 5 noises at 60°, 150°, 180°, 240° and

Table 3.1: $\Delta\text{SNR}_{\text{intellig}}$[dB] as a function of the number of microphones M per array length A_l for diffuse noise and a multiple noise scenario (i.e., 5 noises at 60°, 150°, 180°, 240° and 300°), $\beta(f) = -30$ dB and $\text{SNR}_{\text{in}}(f) = 0$ dB.

scenario	A_l	$M = 2$	$M = 3$	$M = 4$
diffuse	$A_l = 0.03$ m	5.2 dB	7.2 dB	7.5 dB
	$A_l = 0.06$ m	4.6 dB	7.5 dB	8.7 dB
multiple noises	$A_l = 0.03$ m	5.8 dB	9.3 dB	10.5 dB

300°) and $\beta(f) = -30$ dB. For $A_l = 0.03$ m, a third microphone offers a significant improvement, but the benefit of a fourth microphone on the $\text{SNR}_{\text{intellig}}$ is indeed small.

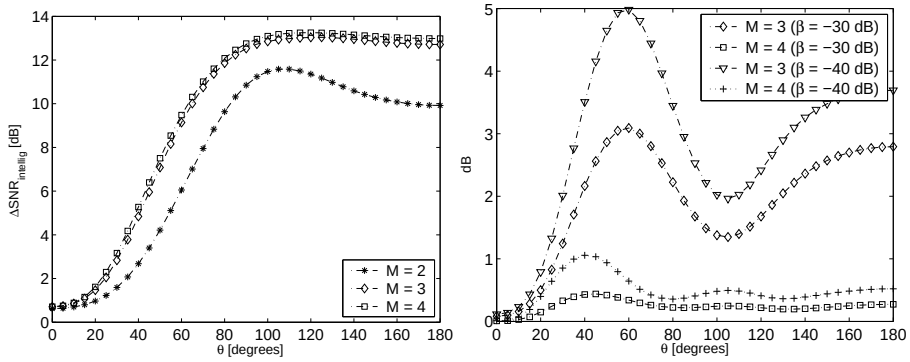
Remark 3.1 In the case of a *single noise source scenario*, a two-microphone array theoretically (i.e., in the absence of reverberation) suffices to cancel one noise source effectively, provided that no spatial aliasing occurs (i.e., $d < \frac{2c}{f}$ for $f \leq \frac{f_s}{2}$). Hence, in the absence of reverberation, the benefit of additional microphones in a BTE will be quite small (cf. Appendix 3.A). In the presence of reverberation, the benefit of additional microphones will be larger, since then not only the localized noise source but also the reverberation has to be suppressed (cf. Section 3.3.4). To illustrate this theoretically, **Figure 3.5(a)** plots the improvement in $\text{SNR}_{\text{intellig}}$ obtained with $M = 2$, $M = 3$ and $M = 4$ for a noise source scenario ($\beta = -30$ dB) consisting of a single directional noise component at angle[4] $\theta \in [0°, 180°]$ and a diffuse noise component, i.e.,

$$\Gamma^n_{kl}(f) = \frac{1}{1+\beta(f)} \left(0.8 e^{j2\pi f \frac{d_l - d_k}{c} \cos\theta} + 0.2\text{sinc}(2\pi f \frac{d_l - d_k}{c}) \right). \qquad (3.46)$$

The diffuse noise component is used here as a model for reverberation. **Figure 3.5(b)** depicts the additional improvement in $\text{SNR}_{\text{intellig}}$ thanks to a third and fourth microphone for $\beta = -30$ dB and $\beta = -40$ dB, demonstrating that the benefit of a third microphone is especially large for $\theta \approx \pm 60°$ and $\theta \approx \pm 180°$ and for small $\beta(f)$. For $\theta = \pm 109°$, the 2-microphone array has the largest improvement and the increase in performance thanks to the third microphone is small. Recall from Section 2.2.4 that for small $f \cdot d$, the largest diffuse noise suppression with $M = 2$ is obtained with a hypercardioid directivity pattern, which has a zero at 109°. For all θ, the benefit of an additional 4th microphone is rather small. □

In short, the number of useful microphones in a commercial BTE is limited to three, provided that the noise reduction algorithm is sufficiently robust so that no constraint on $\Psi(f)$ is needed.

[4]Note that the theoretical performance for a noise source at $\theta \in [180°, 360°]$ equals the performance for a noise source at $360° - \theta$.



(a) (b)

Figure 3.5: Influence of the number of microphones M on $\Delta\text{SNR}_{\text{intellig}}$ for a noise scenario with a directional noise component (i.e., noise source at θ) and a diffuse noise component; $\beta(f) = -30$ dB. (a) $\Delta\text{SNR}_{\text{intellig}}$ for $M = 2$, $M = 3$ and $M = 4$; (b) Additional improvement in $\Delta\text{SNR}_{\text{intellig}}$ thanks to a 3rd ($M = 3$) and 4th microphone ($M = 4$) for $\beta = -30$ dB and $\beta = -40$ dB.

3.3.4 Experimental validation: number of microphones

The theoretical analysis in Section 3.3.3 indicated that the number of useful microphones in a BTE is limited to three. In this section, we verify this result for the subband MWF algorithm based on real-data recordings with the 4-microphone linear endfire array described in Section 2.1.5. The array length A_l is set to 0.03 m. We compare the output $\text{SNR}_{\text{intellig}}$ obtained with two, three and four microphones of the array for different noise scenarios. For the two-microphone array the outer two microphone signals are used and for the three-microphone array, the outer two and the second microphone signals are used. To predict the average performance based on directivity only, the desired signal and the noise signals are both stationary noise signals having the long-term average spectrum of speech (i.e., track 1 of [106]). The level of the desired signal and the total noise signal equals 70 dB SPL at the center of the dummy head.

Figure 3.6 plots the $\text{SNR}_{\text{intellig}}$ at the input (i.e., first microphone) and output of the subband MWF with $M = 2$, $M = 3$ and $M = 4$ for *a single noise source* as a function of the angle θ_n between the noise source and the desired signal. The input $\text{SNR}_{\text{intellig}}$ illustrates the head shadow effect: noise signals coming from $\theta > 180°$ (cf. Figure 2.5(a)) are shielded by the head, which explains the higher input SNR for $\theta > 180°$ and hence the differences in $\Delta\text{SNR}_{\text{intellig}}$ for θ and $360° - \theta$. The use of an additional third microphone offers an additional improvement of about 2 dB for $90° < \theta < 270°$ and about 4 dB for $\theta \approx 60°$. The improvement thanks to a fourth extra microphone is less pronounced.

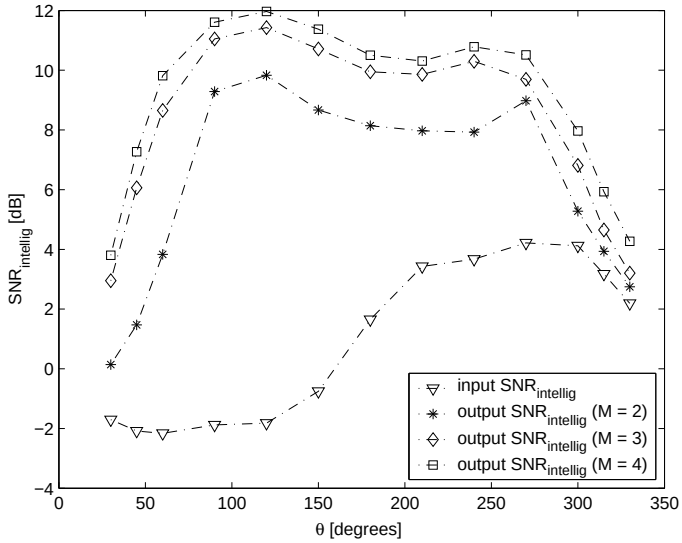


Figure 3.6: $\mathrm{SNR}_{\mathrm{intellig}}$ of subband MWF for different number of microphones M as a function of the angle θ of the noise source. Array length $A_l = 0.03$ m

Table 3.2: $\Delta\mathrm{SNR}_{\mathrm{intellig}}$[dB] as a function of the number of microphones M per array length A_l for diffuse noise and a multiple noise scenario (i.e., 5 noises at 60°, 150°, 180°, 240° and 300°).

scenario	A_l	*input* $\mathrm{SNR}_{\mathrm{intellig}}$	$M = 2$	$M = 3$	$M = 4$
diffuse	$A_l = 0.03$ m	3.1 dB	5.2 dB	6.6 dB	6.8 dB
	$A_l = 0.06$ m	3.2 dB	4.8 dB	6.6 dB	7.5 dB
multiple noises	$A_l = 0.03$ m	0.4 dB	3.1 dB	7.3 dB	8.3 dB

For a single noise source in the back hemisphere (i.e., $90° < \theta < 270°$), a 2-microphone adaptive noise reduction algorithm performs quite well so that the benefit of additional microphones might seem negligible. In real life situations though, multiple noise sources may be present. An example is the cocktail party effect where speech has to be understood among multiple competitive speakers. **Table 3.2** gives the input and output $\text{SNR}_{\text{intellig}}$ for a *diffuse noise* and *a multiple noise source scenario*. Here, an additional third microphone offers an improvement of the performance of the subband MWF of 1.4 to 4 dB, which is considerable compared to the improvement of only 2 dB a 3 dB offered by the 2-microphone array. For $A_l = 0.03$ m, the improvement of an additional fourth microphone is again small. Note that the increase in performance thanks to an additional third and fourth microphone corresponds well to the theoretically predicted values.

3.4 Sensitivity to errors in the assumed signal model

Section 3.3 showed that for small-sized arrays, complex noise scenarios and an increasing number of microphones result in a high noise sensitivity. In this section, we evaluate the robustness of the GSC, the NSC-GSC and the MWF to different types of errors in the assumed signal model, i.e. microphone mismatch, microphone placement and speaker location. The theoretical robustness analysis allows to identify which type of model errors will be most detrimental to the GSC when used for hearing aids. The theoretical results are illustrated for a three-microphone linear endfire array with microphone interspacings $d_2 - d_1 = 0.01$ m and $d_3 - d_2 = 0.015$ m and a multiple noise source scenario with five localized noise sources at 75°, 150°, 180°, 240°, 285° and $\beta(f) = -30$ dB. The major results are verified based on real-data recordings with the three-microphone BTE.

3.4.1 Robustness to microphone mismatch

In practice, microphones are rarely matched in gain and phase. In [114], gain and phase differences between microphone characteristics of up to 6 dB and 10°, respectively, have been reported. The fixed stage of the GSC can be designed to compensate for the differences in microphone characteristics. However, calibrating the fixed stage for each hearing aid individually is cumbersome and expensive. In addition, microphone characteristics may drift over time [114]. Hence, the robustness of the algorithm to gain and phase mismatch is crucial.

We mimic a deviation from matched microphone characteristics by modifying the coherence functions $\Gamma^s_{kl}(f)$, $\Gamma^n_{kl}(f)$ and $\Gamma^i_{kl}(f)$. Suppose that the k-th mi-

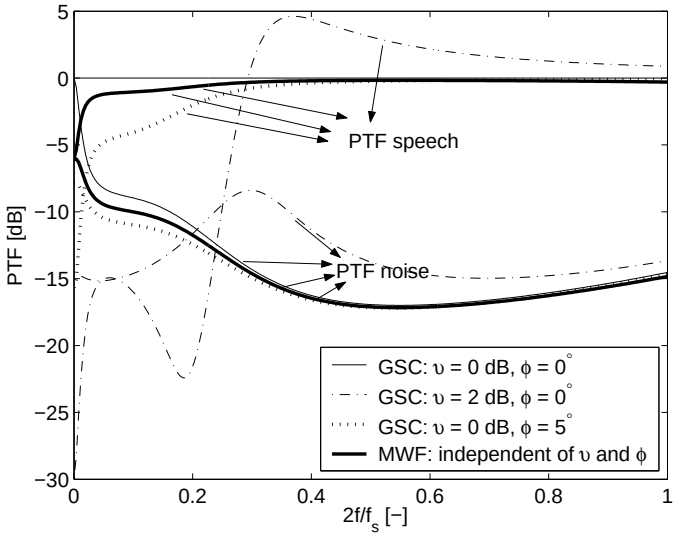


Figure 3.7: Effect of microphone gain mismatch υ and phase mismatch ϕ on the PTF $G^s(f)$ and $G^n(f)$ of GSC and MWF for a multiple noise source scenario and $\beta(f) = -30$ dB. The sampling frequency $f_s = 16$ kHz.

crophone has a gain mismatch $\upsilon(f)$ and a phase mismatch $\phi(f)$ at frequency f with respect to microphone l. The matched microphones coherence function $\Gamma_{kl}(f)$ should then be replaced by $\tilde{\Gamma}_{kl}(f)$

$$\tilde{\Gamma}_{kl}(f) = \begin{cases} \upsilon(f)\exp(j\phi(f))\Gamma_{kl}(f) & \text{for } l \neq k; \\ |\upsilon(f)|^2\Gamma_{kl}(f) & \text{for } l = k. \end{cases} \tag{3.47}$$

Figure 3.8 plots the PTF $G^s(f)$ and $G^n(f)$ of the GSC when the second microphone has a gain mismatch $\upsilon(f) = 2$ dB or a phase mismatch $\phi(f) = 5°$ with respect to the first microphone. The desired signal gets cancelled at lower f where $\Psi(f)$ is high (cf. Figure 3.2). Due to the high $\Psi(f)$ and the cancellation of the desired speech, internal noise may become audible. The PTFs $G^s(f)$ and $G^n(f)$ of the MWF (for noise-to-signal ratio $\alpha(f) = 0$ dB) are also shown. As demonstrated in Section 3.2.2, the MWF attenuates the speech at frequencies where the output SNR is small, i.e., in this example at low f.

Theorem 3.1 *Theoretically, the PTF $G_s(f)$ and $G_n(f)$ of the MWF are independent of a deviation in microphone gain and phase.*

Proof : Suppose that the received microphone signals $\tilde{X}_i(f)$, with $i = 1, ..., M$ are perturbed by $\Upsilon_i(f) = \upsilon_i(f)\exp(j\phi_i(f))$ with $\upsilon_i(f)$ a gain deviation and

$\phi_i(f)$ a phase deviation, with respect to the signals $X_i(f)$ received in the absence of microphone mismatch, i.e.,

$$\tilde{\mathbf{X}}(f) = \mathbf{D}(f)\mathbf{X}(f), \tag{3.48}$$

with

$$\mathbf{D}(f) \quad = \quad \text{diag}\{\Upsilon_1(f), ..., \Upsilon_M(f)\} \tag{3.49}$$

and with $\mathbf{D}(f)$ of full rank, i.e., $\Upsilon_i(f) \neq 0$. The MWF $\tilde{\mathbf{W}}(f)$ in (3.25) then equals

$$\begin{aligned}\tilde{\mathbf{W}}(f) &= \mathcal{E}\{\tilde{\mathbf{X}}(f)\tilde{\mathbf{X}}^H(f)\}^{-1}\left(\mathcal{E}\{\tilde{\mathbf{X}}(f)\tilde{\mathbf{X}}^H(f)\} - \mathcal{E}\{\tilde{\mathbf{X}}^n(f)\tilde{\mathbf{X}}^{n,H}(f)\}\right)\mathbf{e}_1, \\ &= \mathbf{D}^{-H}(f)\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}^{-1} \\ &\quad \cdot \left(\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}\right)\mathbf{D}^H(f)\mathbf{e}_1 \end{aligned} \tag{3.50}$$

This results in the output $\tilde{Z}(f) = \Upsilon_1(f)\mathbf{W}^H(f)\mathbf{X}(f)\mathbf{e}_1$, which is a scaled version of the output $Z(f) = \mathbf{W}^H(f)\mathbf{X}(f)\mathbf{e_1}$ that would have been obtained in the absence of mismatch. Both the input $\tilde{X}_1(f)$ and the output $\tilde{Z}(f)$ of the MWF are scaled by $\Upsilon_1(f)$. Hence, the PTFs $G^s(f)$, $G^n(f)$ of the MWF are not affected by the mismatch. □

In **Figure 3.8(a)-(d)**, the performance measures $\Delta\text{SNR}_{\text{intellig}}$ (Figure 3.8(a) and (c)) and $\text{SD}_{\text{intellig}}$ (Figure 3.8(b) and (d)) are shown as a function of the amount of gain mismatch υ and phase mismatch ϕ, respectively, at the second microphone. The performance of the more robust NSC-GSC, described in Section 3.2.1, is depicted too for two levels of noise injection $\eta(f)$, $\eta(f) = -20$ dB and $\eta(f) = -10$ dB. Without noise sensitivity constraint, a small amount of mismatch leads to an unacceptable degradation of the GSC performance in complicated noise scenarios and at low $f \cdot d$, even if the ANC is adapted during noise-only periods. Phase mismatch is seen to be less detrimental than gain mismatch. Constraining the noise sensitivity $\Psi(f)$ results in a higher robustness at the expense of less noise reduction. The performance of the MWF on the other hand is again seen to be robust to gain and phase mismatch, so that, depending on the processing accuracy, no (or a less severe) constraint on the noise sensitivity is needed. As a consequence, the MWF outperforms the robust NSC-GSC.

3.4.2 Robustness to deviations in microphone positions

The positions of the microphones vary slightly from one hearing aid to the other. Since microphones are quite accurately positioned in a hearing aid, the microphone placement uncertainty is expected to be small. A change Δd (in meters) in the microphone position of the k-th microphone with respect to the

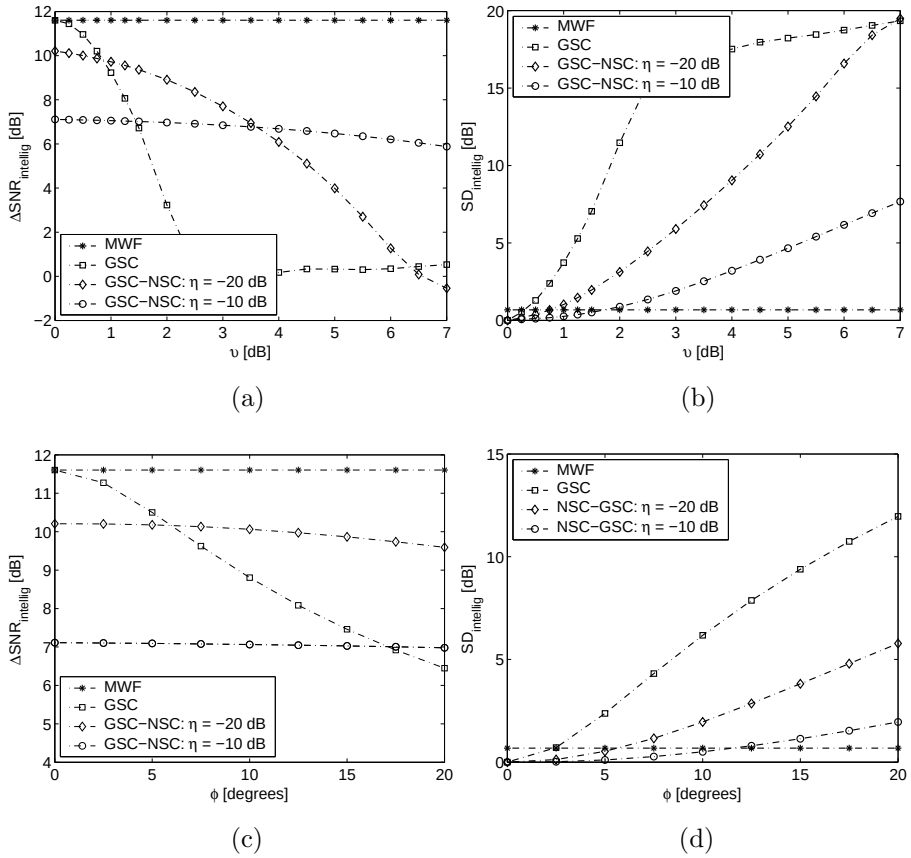


Figure 3.8: Performance ($\Delta\text{SNR}_{\text{intellig}}$ (a),(c) and $\text{SD}_{\text{intellig}}$ (b),(d)) of GSC, NSC-GSC and MWF in a multiple noise source scenario as a function of microphone mismatch; (a),(b) Microphone gain mismatch υ (in dB); (c),(d) Microphone phase mismatch ϕ (in degrees).

assumed position d_k can be incorporated in the coherence functions $\Gamma_{kl}(f)$, $\Gamma^n_{kl}(f)$ and $\Gamma^s_{kl}(f)$ by replacing d_k with $d_k + \Delta d$.

Figure 3.9 depicts the PTF of speech and noise for GSC and MWF when the 2nd microphone is moved by $\Delta d = -0.001$ m and $\Delta d = -0.002$ m, respectively. The performance of the GSC is not strongly affected by such a small change in microphone positions, even when $\Psi(f)$ is rather large. The speech and noise are slightly attenuated, but the amount of noise reduction remains quite constant. In contrast to microphone mismatch, the largest effect of a change in microphone position does not occur at the lowest frequencies. A change by Δd in the coherence function (2.35) of a localized sound source at angle θ corresponds to a frequency dependent phase mismatch in the k-th microphone frequency characteristic of

$$2\pi f \frac{\cos\theta}{c} \Delta d \tag{3.51}$$

(in rad). On the one hand, the impact of phase mismatch is largest for high $\Psi(f)$ and hence at low f, on the other hand the amount of phase mismatch (3.51) grows with increasing frequency. At low f where $\Psi(f)$ is high, the phase change is very small for $\Delta d = -0.001$ m and $\Delta d = -0.002$ m, hence the limited effect of a change in microphone position. Note that the impact still increases with increasing $\Psi(f)$. In applications where the uncertainties Δd are larger, a constraint on the noise sensitivity of GSC may be required. Figure 3.9 shows that the PTF of speech and noise of the MWF are hardly modified when the microphones are slightly moved.

3.4.3 Robustness to variations in speaker location

In addition to being robust to microphone mismatch and variations in microphone positions, the algorithm should also be effective when the signal-of-interest does not arrive from the front but from an angle $\Delta\theta$ with respect to the front. The effect of a change $\Delta\theta$ can easily be studied analytically by incorporating the angle $\Delta\theta$ with respect to 0° in $\Gamma^s_{kl}(f)$ (cf. (2.35) in Section 2.1.3), i.e.,

$$\Gamma^s_{kl}(f) = e^{j2\pi f \cos(\Delta\theta)\frac{d_l - d_k}{c}}. \tag{3.52}$$

Since the ANC filters of the GSC are completely determined by $\Gamma^n_{kl}(f)$ and do not depend on $\Gamma^s_{kl}(f)$ (cf. Section 3.2.1), the noise PTFs $G^n(f)$ of the GSC are not affected by $\Delta\theta$.

Figure 3.10 gives the effect of speaker variations up to 45° for input noise-to-signal ratios $\alpha(f) = 0$ dB and $\alpha(f) = 3$ dB (cf. (3.29)), respectively, on the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the GSC and MWF. In contrast to microphone mismatch, this effect is not detrimental to the GSC: for small to moderate variations up to 20°, the GSC even outperforms the MWF because of its smaller amount of speech distortion. The improvement

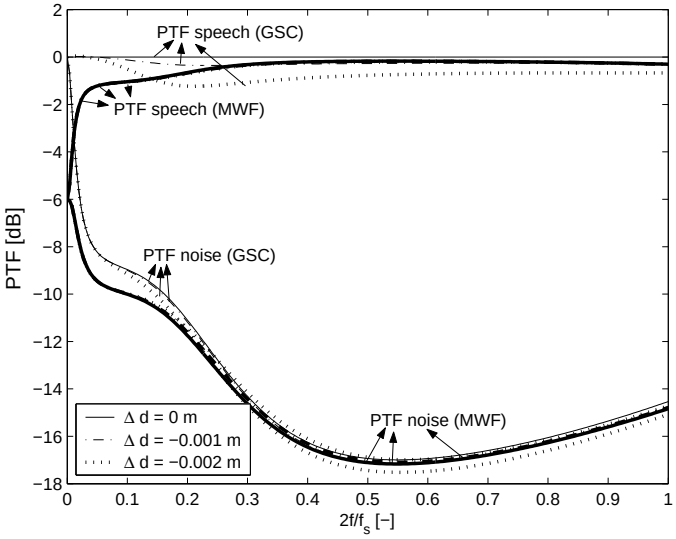


Figure 3.9: Effect of a deviation Δd (in meters) in the position d_k of the 2nd microphone on the PTF $G_s(f)$ and $G_n(f)$ of the GSC and the MWF. The sampling frequency $f_s = 16$ kHz.

$\Delta\text{SNR}_{\text{intellig}}$ decreases with increasing $\Delta\theta$. The amount of speech distortion $\text{SD}_{\text{intellig}}$ caused by the MWF grows with decreasing output SNR, i.e., with increasing $\Delta\theta$ and increasing $\alpha(f)$. In free field, the enhancement $\Delta\text{SNR}_{\text{intellig}}$ is the same for positive and negative $\Delta\theta$. When mounted on the head, the input SNR and hence $\alpha(f)$ will be different for positive and negative $\Delta\theta$ due to the head shadow effect: when speech arrives from a negative angle $\Delta\theta$ being shielded by the head, $\alpha(f)$ will be larger than when speech arrives from a positive angle $\Delta\theta$. Hence, the speech will in practice be more attenuated for negative $\Delta\theta$ (cf. Figure 3.14 of Section 3.4.4).

3.4.4 Experimental validation

In this section, we experimentally analyze the robustness of the GSC and the MWF to microphone mismatch and variations in speaker location based on real-data recordings with the three-microphone BTE. The noise source scenario consists of 5 spectrally non-stationary, multi-talker babble noise sources [5] positioned at 75°, 120°, 180°, 240° and 285°. The desired signal source consists of English sentences spoken by a male speaker from the HINT-database [185]. The input SNR at the first microphone equals 0 dB. In the experiments, the forgetting factors λ, λ_{s+n} and λ_n of the MWF and GSC are set to 0.99998. This implies the assumption of a quasi-stationary environment during at least

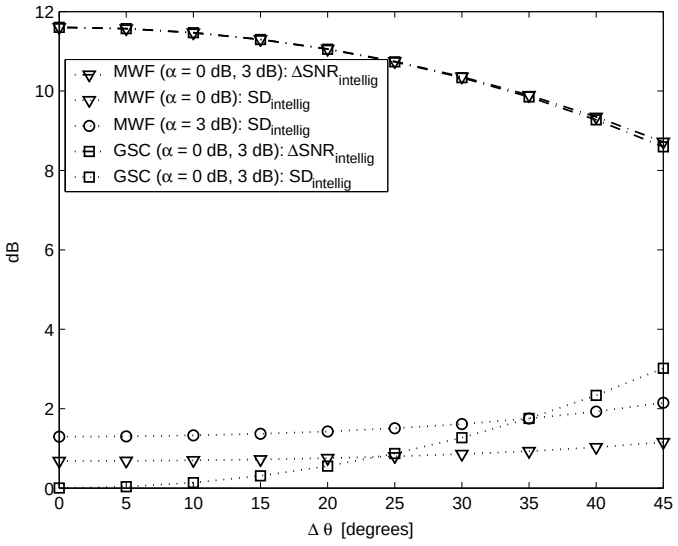


Figure 3.10: Influence of a variation in speaker location on the broadband performance of the GSC and the MWF for input noise-to-signal ratio $\alpha = 0$ dB and $\alpha = 3$ dB. The sampling frequency $f_s = 16$ kHz.

$2/[f_s(1 - 0.99998)] \approx 6$ seconds. A larger λ results in better estimation of the second order statistics in quasi-stationary environments, while for tracking in highly non-stationary environments, a smaller λ may be preferred. For the effect of the forgetting factor λ, λ_{s+n} and λ_n on the performance of the GSC and MWF, we refer to Section 3.5.1.

Robustness to microphone mismatch: PTFs

Figure 3.11 shows the PTF $G^s(f)$ and $G^n(f)$ of the GSC and MWF when the second microphone has a gain mismatch of $v(f) = 2$ dB with respect to the first microphone. A small amount of gain mismatch is indeed detrimental to the performance of the GSC at low $f \cdot d$. An additional constraint on the noise sensitivity is indispensable to guarantee robustness in complicated noise scenarios. The MWF on the other hand is independent of the applied deviation, resulting in improved noise reduction. Compared to Figure 3.7, the MWF attenuates frequencies around $2f/f_s = 0.2$ (with $f_s = 16$ kHz) because the input SNR of the real speech and noise is lower around $2f/f_s = 0.2$ than at other f, resulting in attenuation by the spectral Wiener filter $W_2(f)$ (see (3.35)).

Figure 3.12 plots the noise sensitivity $\Psi(f)$ for both algorithms. For comparison, the theoretical $\Psi(f)$ for internal-to-external noise ratio $\beta = -30$ dB is

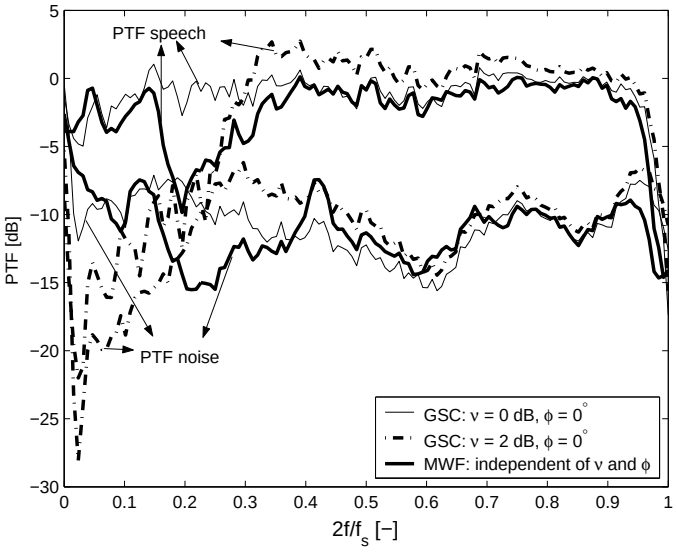


Figure 3.11: Effect of microphone gain mismatch υ on the PTF $G^s(f)$ and $G^n(f)$ of the GSC and MWF for a non-stationary multiple noise source scenario. The sampling frequency $f_s = 16$ kHz.

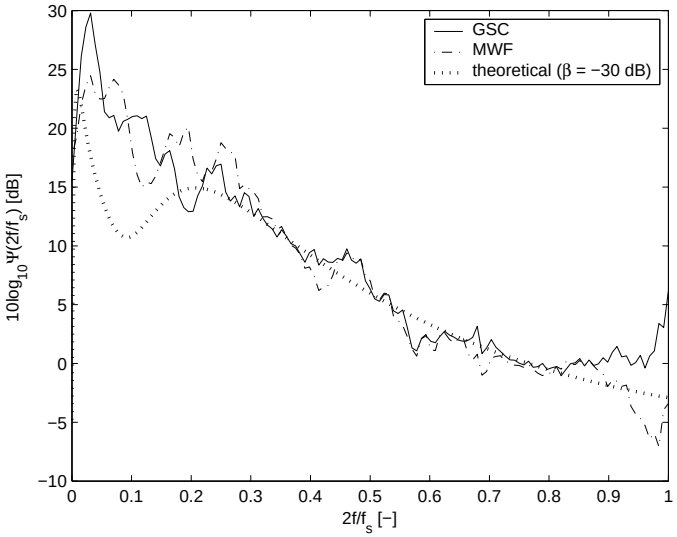


Figure 3.12: Experimentally obtained noise sensitivity of the GSC and MWF in a multiple noise source scenario. The sampling frequency $f_s = 16$ kHz.

depicted too. Theoretically, the noise sensitivity $\Psi(f)$ of the GSC and MWF coincide (see Section 3.3): the small differences are due to reverberation. At low f, the noise sensitivity $\Psi(f)$ is larger than the theoretical $\Psi(f)$ for $\beta = -30$ dB because $\beta(f)$ is lower than -30 dB due to the high energy of the external speech-weighted noise source at low frequencies. As a consequence, the noise signal is attenuated more at low f than in Figure 3.7.

Robustness to microphone mismatch: intelligibility weighted performance

In **Figure 3.13(a)-(b)** and **Figure 3.13(c)-(d)**, the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the GSC and MWF are given as a function of the amount of microphone gain mismatch υ and microphone phase mismatch ϕ, respectively, at the second microphone. For comparison, the experimentally found performance of the GSC with $\beta(f)$ increased to about -30 dB is also shown. Although differences occur between Figure 3.7(a)-(d) and Figure 3.13(a)-(d) due to reverberation, a different $\beta(f)$, the presence of the head and non-stationary signals, the experimental results are quite well predicted by the theory: the performance of the MWF is indeed independent of υ and ϕ, whereas a small amount of gain mismatch suffices to deteriorate the performance of the GSC. Again, phase mismatch is found to be less detrimental than gain mismatch. In the absence of mismatch, the performance of the MWF is a bit worse than the GSC: as explained in Section 2.4, the MWF tries to estimate the reverberant speech component in the first microphone signal, and hence, trades off noise reduction and cancellation of the reverberant part in the speech. The GSC on the other hand only tries to estimate the direct path. In addition, the estimation of the second order statistics of the speech by the MWF is not perfect, resulting in a minor, additional degradation in noise reduction performance (cf. Section 3.5). The speech distortion of the MWF is larger than in Figure 3.7, due to the spectral filtering (see also Figure 3.11). The performance curves of the NSC-GSC are depicted too. The noise injection levels are approximately set to $\eta(f) = -20$ dB and $\eta(f) = -10$ dB with respect to the power of the noise in the pre-emphasized, first microphone signal. The curves for $\eta = -20$ dB and $\eta = -10$ dB confirm the ones predicted in Figure 3.7: a noise sensitivity constraint results in a smaller performance degradation in case of gain and/or phase mismatch at the expense of less noise reduction. The stronger the constraint, the lower the noise reduction that is achieved with the robust GSC. Given the amount of mismatch that may occur, an appropriate constraint on $\Psi(f)$ should be chosen so that the performance is not severely affected in the presence of the mismatch: to allow for a gain mismatch up to 6 dB, $\eta(f) = -10$ dB is a good choice. The MWF outperforms the robust NSC-GSC, even when the noise is spectrally non-stationary.

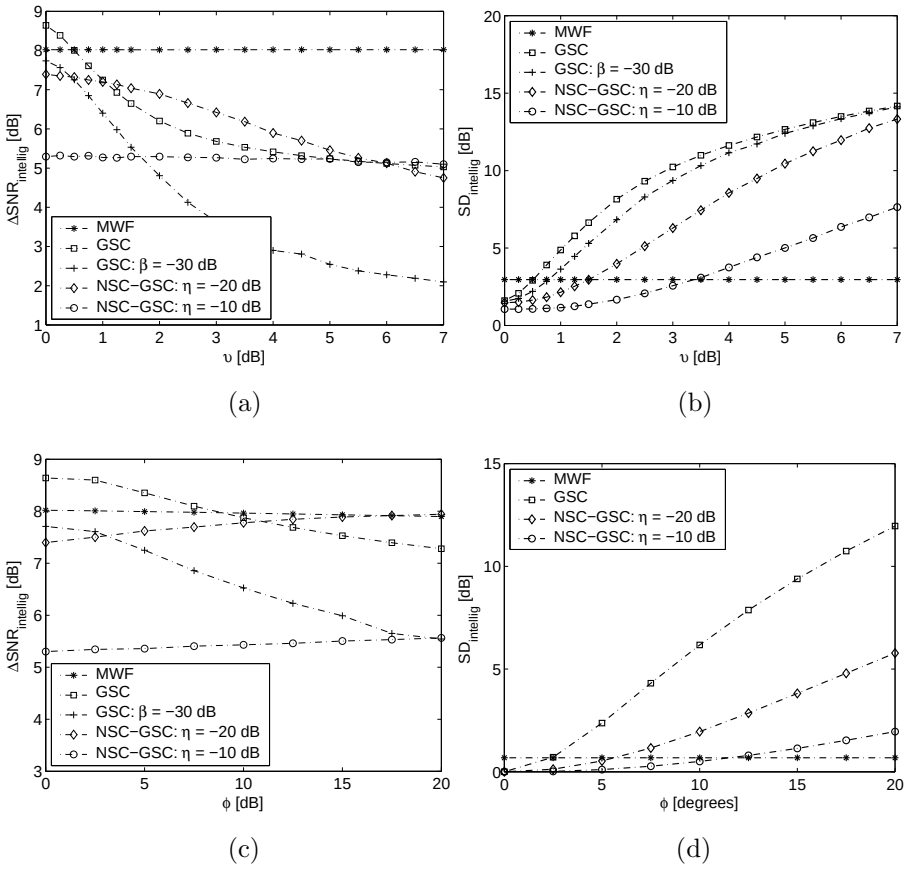


Figure 3.13: Intelligibility weighted performance measures ($\Delta\text{SNR}_{\text{intellig}}$ (a),(c) and $\text{SD}_{\text{intellig}}$(b),(d)) of the GSC, NSC-GSC and MWF in a multiple noise source scenario as a function of the amount of microphone mismatch; (a),(b) Microphone gain mismatch v (in dB); (c),(d) Microphone phase mismatch ϕ (in degrees).

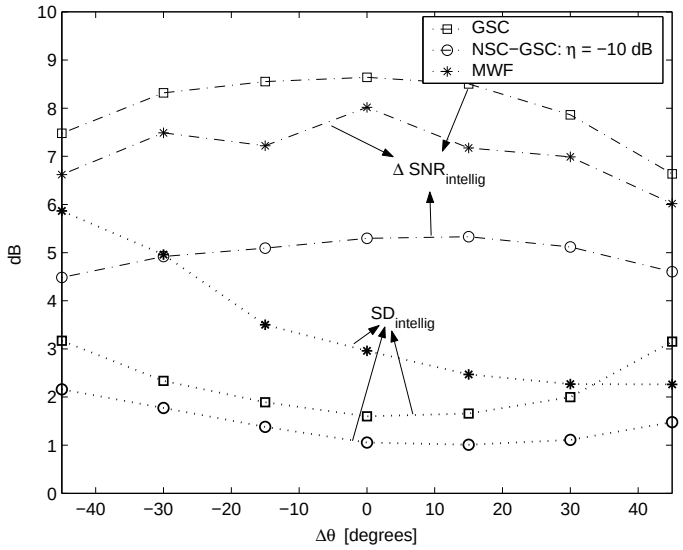


Figure 3.14: Influence of a variation $\Delta\theta$ (in degrees) in speaker location on the improvement $\Delta\text{SNR}_{\text{intellig}}$ (dashed-dotted lines) and the distortion $\text{SD}_{\text{intellig}}$ (dotted lines) of the GSC, the NSC-GSC ($\eta = -10$ dB) and the MWF.

Robustness to variations in speaker location

Figure 3.14 illustrates the effect of a variation $\Delta\theta$ (in degrees) in speaker location with respect to $\theta = 0°$ on the improvement $\Delta\text{SNR}_{\text{intellig}}$ and on the speech distortion $\text{SD}_{\text{intellig}}$ for the MWF, the GSC and the robust NSC-GSC with $\eta(f) = -10$ dB. As predicted theoretically, no significant degradation of the performance of the GSC occurs for speaker variations $\Delta\theta$ up to 45° and $-45°$. In contrast to Figure 3.10, the performance measures are asymmetric for positive and negative $\Delta\theta$. As explained in Section 3.4.3, this is due to the head shadow effect. Speech arriving from negative angles $\Delta\theta$ is shielded by the head, hence negative $\Delta\theta$ result in a higher input noise-to-signal ratio $\alpha(f)$. As shown in Figure 3.10, a higher $\alpha(f)$ induces a larger amount of speech distortion by the MWF.

In conclusion, the GSC seems extremely sensitive to microphone gain mismatch when $\Psi(f)$ is high, i.e., at low $f \cdot d$ and in complicated noise scenarios. The effect of variations in speaker position and microphone positions on the performance also grows for increasing $\Psi(f)$. However, even for high $\Psi(f)$, they are found to be not detrimental for small-sized arrays. To guarantee performance in the presence of microphone gain mismatch, a constraint on the noise sensitivity should be imposed at the expense of less noise reduction. The MWF does not suffer from performance degradation in the presence of microphone mismatch

and variations in speaker and microphone positions, so that, depending on the processing accuracy, no or a less severe constraint on the noise sensitivity is needed, and hence, an improved performance is achieved. In addition, the performance benefit of a third microphone in a BTE, which is determined by $\beta(f)$ and the allowable noise sensitivity $\Psi(f)$ (cf. Section 3.3.3), is larger for the MWF than for the GSC.

3.5 Sensitivity to inaccurate estimation of the second order statistics

In the theoretical analysis of the previous sections, we assumed perfect estimation of the second order statistics $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ of the noise by the GSC and perfect estimation of the second order statistics $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ of the speech and the noise by the MWF. Because of double-talk of speech and noise, the clean speech signal is not available. Recall from Section 2.4 that the MWF estimates the speech correlation matrix $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}$ as

$$\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} = \mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}, \tag{3.53}$$

where $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ are estimated during periods of speech + noise and periods of noise-only, respectively.

In practice, the estimation of the second order statistics may be inaccurate because of several reasons:

- The correlation matrices $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ are estimated from a *finite amount of incoming data* and *at a different moment in time* so that (3.53) does not exactly hold true. This estimation error has been included in all the experiments. To allow for tracking of time-varying environments, the amount of data (time window) used for the estimation of the second order statistics should be kept small. However, the smaller the amount of data, the worse the accuracy of the estimated second order statistics. Section 3.5.1 experimentally studies the influence of the amount of data on the performance of the MWF and GSC for a spatially and long-term spectrally stationary noise scenario.
- Secondly, because of the non-continuous estimation of the noise correlation matrix by the GSC[5] and the noise and speech + noise correlation matrices by the MWF, tracking of *non-stationarities in the noise* is reduced. Non-stationarities of the noise then also lead to a wrong estimate of $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}$ by the MWF.

[5]Note that the ANC of the GSC is adapted during noise-only periods to avoid speech distortion. Hence, changes in the noise scenario that occur during speech + noise periods, are not immediately tracked by the GSC.

- Finally, both the GSC and the MWF rely on a speech detection algorithm. *Speech detection errors* deteriorate the accuracy of the estimated second order statistics by the GSC and the MWF.

Since the MWF does not rely on a priori information, the MWF is expected to be more sensitive to the accuracy of the second order statistics. In this and the next section, we investigate the sensitivity of the GSC, the NSC-GSC and the MWF to inaccurate estimation of the second order statistics. In this section, we concentrate on the sensitivity to inaccurate second order statistics due to estimation from a finite amount of data (Section 3.5.1) and due to non-stationarities in the noise (Section 3.5.2). The average effect of speech detection errors on the performance of the GSC, the NSC-GSC and the MWF is elaborated in Section 3.6. The theoretical results are illustrated for a linear 3-microphone endfire array with microphone interspacings $d_2 - d_1 = 0.01$ m and $d_3 - d_2 = 0.015$ m.

3.5.1 Estimation from finite amount of data

In this section, we experimentally analyze the influence of the amount of data that is used to estimate the second order statistics, on the performance of the GSC and the MWF.

In the experiments of Section 3.4.4, the forgetting factors λ, λ_{s+n} and λ_n of the GSC and MWF were set to 0.99998 to allow for tracking of slowly time-varying environments. A larger λ results in a more accurate estimation of the second order statistics in quasi-stationary environments, while for tracking in highly non-stationary environments, a smaller λ is preferred.

Figure 3.15 plots the improvement $\Delta\text{SNR}_{\text{intellig}}$ of the GSC and MWF as a function of the time window $1/\left[(1-\lambda)f_s\right]$ (in seconds) with $\lambda = \lambda_{s+n} = \lambda_n$ for the multiple noise source scenario (cf. Section 3.4) with 5 spectrally non-stationary noise sources (i.e., multi-talker babble noise [5]) and 5 stationary noise sources (i.e., stationary speech-weighted noise from the HINT-database [185]). The spatial characteristics of the noise sources are fixed. For *spectrally stationary noise sources*, $\Delta\text{SNR}_{\text{intellig}}$ for the GSC and the MWF is not significantly affected by a decrease in λ. However, for small λ_{s+n} and λ_n, the MWF depends on the short-term characteristics of the speech and noise sources so that short-time effects such as musical noise may occur. For *spectrally non-stationary noise sources,* the MWF is affected more than the GSC by a decrease in λ, especially for small λ: because of the short-time non-stationary spectrum, the noise PSD estimated during speech + noise periods differs from the noise PSD estimated during noise-only periods, especially when λ is small. As will be demonstrated in Section 3.5.2, this results in a large error in the speech correlation matrix estimated by the MWF, and consequently, in a decrease in $\Delta\text{SNR}_{\text{intellig}}$. For time windows $1/[(1-\lambda)f_s]$ larger than 0.5 s, the performance

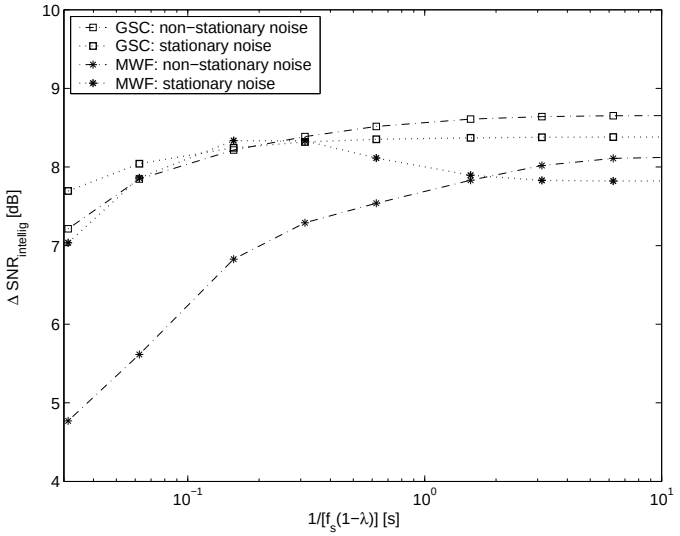


Figure 3.15: Influence of forgetting factor $\lambda = \lambda_{s+n} = \lambda_n$ on the improvement $\Delta\text{SNR}_{\text{intellig}}$ of GSC and MWF for spectrally non-stationary and stationary noise sources, respectively.

loss of MWF is limited: the MWF then depends on the average, long-term characteristics of speech and noise so that spectrally non-stationary noise is successfully suppressed. Hence, the MWF preserves its benefit in robustness and performance over the GSC and the NSC-GSC for spectrally non-stationary noise, provided that the spatial characteristics and long-term average PSD of the noise vary relatively slowly in time (i.e., on the order of seconds).

3.5.2 Non-stationarities in the noise scenario

Section 3.5.1 demonstrated that the MWF is able to suppress spectrally non-stationary noise, provided that the spatial and long-term spectral characteristics of the noise are quasi-stationary. If, however, also the *spatial and long-term spectral characteristics of the noise* are non-stationary, the performance of the MWF is expected to be affected. In this section, we theoretically and experimentally analyze the sensitivity of the MWF, the GSC and the NSC-GSC to non-stationarities in the noise scenario.

Theoretical analysis

As an example to theoretically illustrate the sensitivity of the MWF, the GSC and NSC-GSC to erroneous estimation of the second order statistics due to non-stationarities in the noise scenario, we assume that the noise correlation matrix

$\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ is wrongly estimated because of a change in noise scenario. Suppose that the estimated noise correlation matrix $\mathcal{E}\{\mathbf{X}^n_{est}(f)\mathbf{X}^{n,H}_{est}(f)\}$ equals

$$\begin{aligned}\mathcal{E}\{\mathbf{X}^n_{est}(f)\mathbf{X}^{n,H}_{est}(f)\} &= \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} \\ &\quad +\delta\left(\mathcal{E}\{\tilde{\mathbf{X}}^n(f)\tilde{\mathbf{X}}^{n,H}(f)\} - \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}\right), \quad (3.54)\end{aligned}$$

where $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ corresponds to the new noise correlation matrix after the change in noise scenario, $\mathcal{E}\{\tilde{\mathbf{X}}^n(f)\tilde{\mathbf{X}}^{n,H}(f)\}$ corresponds to the correlation matrix of the previous noise scenario and the scalar $\delta \in [0,1]$ measures how much $\mathcal{E}\{\mathbf{X}^n_{est}(f)\mathbf{X}^{n,H}_{est}(f)\}$ is perturbed by the non-stationarity. The parameter δ can be interpreted as a measure for the amount of non-stationarity. The error (3.54) in the noise correlation matrix leads to a wrong estimate $\mathcal{E}\{\mathbf{X}^s_{est}(f)\mathbf{X}^{s,H}_{est}(f)\}$ of $\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}$, i.e.,

$$\begin{aligned}\mathcal{E}\{\mathbf{X}^s_{est}(f)\mathbf{X}^{s,H}_{est}(f)\} &= \mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} \\ &\quad +\delta\left(\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} - \mathcal{E}\{\tilde{\mathbf{X}}^n(f)\tilde{\mathbf{X}}^{n,H}(f)\}\right), \quad (3.55)\end{aligned}$$

resulting in

$$\begin{aligned}\mathbf{W}(f) &= \mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}^{-1}\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\}\mathbf{e_1} + \boldsymbol{\Delta}\mathbf{W}(f), \\ \Delta\mathbf{W}(f) &= \delta\cdot\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}^{-1} \\ &\quad \cdot\left(\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} - \mathcal{E}\{\tilde{\mathbf{X}}^n(f)\tilde{\mathbf{X}}^{n,H}(f)\}\right)\mathbf{e}_1. \quad (3.56)\end{aligned}$$

Example 3.1 As an example, **Figure 3.16** plots the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the MWF, the GSC and the NSC-GSC with $\eta = -10$ dB, as a function of δ with $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ and $\mathcal{E}\{\tilde{\mathbf{X}}^n(f)\tilde{\mathbf{X}}^{n,H}(f)\}$ computed using

$$\begin{aligned}P^n_{X_kX_l}(f) &= \frac{P^s_X(f)}{1+\beta(f)}\left[0.8e^{j2\pi f\cos\theta\frac{d_l-d_k}{c}}\right. \\ &\quad \left.+0.2\,\text{sinc}(2\pi f(d_l-d_k)/c)\right], \quad (3.57) \\ P^n_{\tilde{X}_k\tilde{X}_l}(f) &= \frac{\Delta\alpha(f)P^s_X(f)}{1+\beta(f)\Delta\alpha(f)}\left[0.8e^{j2\pi f\cos(\theta+\Delta\theta)\frac{d_l-d_k}{c}}\right. \\ &\quad \left.+0.2\,\text{sinc}(2\pi f(d_l-d_k)/c)\right], \quad (3.58)\end{aligned}$$

where $\Delta\alpha(f) = P^n_{\tilde{X}}(f)/P^n_X(f)$ represents the difference in PSD between the new and the previous noise scenario and $\Delta\theta$ the difference in angle of the directional noise component. **Figure 3.16(a) and (b)** depict the results for $\theta = 180°$, $\beta(f) = -30$ dB with $\Delta\theta = 15°$ and $\Delta\alpha(f) = 0$ dB and $\Delta\alpha(f) = 2$ dB, respectively. As expected, the MWF is more sensitive to a wrongly estimated noise correlation matrix than the GSC and the NSC-GSC because of the error in $\mathcal{E}\{\mathbf{X}^s_{est}(f)\mathbf{X}^{s,H}_{est}(f)\}$ (see (3.55)). Note that the GSC will perform

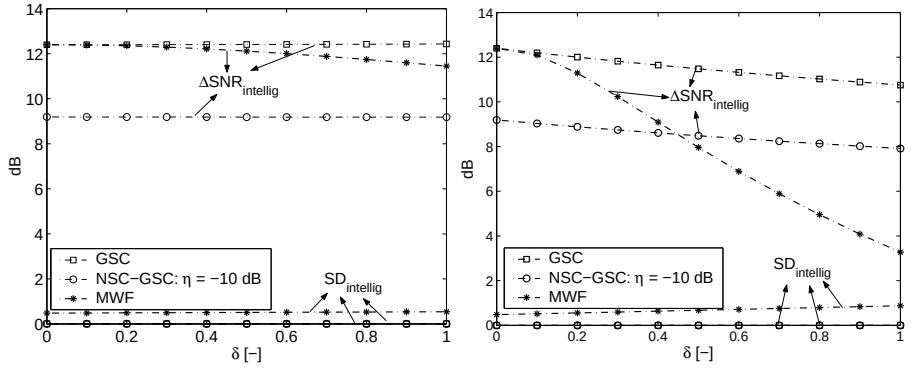


Figure 3.16: Influence of inaccurate estimation of the noise correlation matrix on the performance of the MWF, the GSC and the NSC-GSC ($\eta = -10$ dB). (a) $\Delta\alpha = 0$ dB; (b) $\Delta\alpha = 2$ dB.

worse than in Figure 3.16 if other signal model errors such as mismatch, occur. In particular, a change in PSD, i.e., $\Delta\alpha(f) \neq 0$ dB, causes a large error in $\mathcal{E}\{\mathbf{X}^s_{est}(f)\mathbf{X}^{s,H}_{est}(f)\}$, and hence, a significant performance degradation to the MWF. Especially the noise reduction performance $\Delta\text{SNR}_{\text{intellig}}$ of the MWF is affected. The increase in distortion $\text{SD}_{\text{intellig}}$ is small for this scenario: the deviation $\Delta\mathbf{W}(f)$ in (3.56) from the ideal MWF corresponds to an MWF that suppresses the desired speech signal $\mathbf{X}^s(f)$ so that the distortion $\text{SD}_{\text{intellig}}$ is limited. In contrast to microphone mismatch for the GSC, the degradation in the performance of the MWF due to inaccurate estimation of the second order statistics was not found to be worse at low f (or for a high noise sensitivity $\Psi(f)$) than at high f (or for a low sensitivity $\Psi(f)$). △

Example 3.1 illustrates that the MWF may lose its benefit over the NSC-GSC in non-stationary noise scenarios due to a wrong estimation of the second order statistics. The loss in performance of the MWF will depend on the non-stationarity of the noise scenario and the implementation of the MWF (e.g., the forgetting factor that is used in the update formula of the speech + noise and noise correlation matrix and the speech detection algorithm).

Validation through experimental results

Below, we experimentally verify the effect of a change in noise scenario on the MWF and the NSC-GSC.

As an illustration, **Figure 3.17** depicts the speech cancellation and noise suppression by the MWF and the NSC-GSC with $\eta \approx -20$ dB and $\lambda = \lambda_{s+n} =$

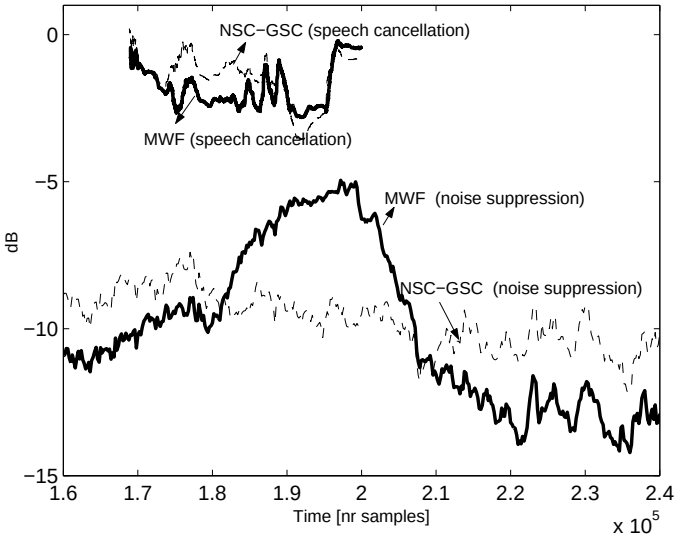


Figure 3.17: Influence of change in noise scenario on the speech cancellation and noise suppression by the NSC-GSC (with $\eta \approx -20$ dB) and the MWF.

$\lambda_n = 0.99995$ for a sudden change in position and power of a stationary speech-weighted noise source. The speech-weighted noise signal was taken from the HINT-database [185]. The position of the noise source changes at sample 1.8×10^5 from 180° to 150° and its power increases by 1.5 dB. Between sample 1.68×10^5 and 2×10^5, speech + noise is present; the other samples contain noise-only. The speech cancellation/noise suppression are measured as the ratio of the output speech/noise power over the speech/noise power at the first microphone. As theoretically predicted in Section 3.5.2, especially the noise suppression of the MWF is degraded by the change: during speech + noise periods, the noise correlation matrix is not updated, which results in a wrong estimate of the speech correlation matrix by the MWF. The speech cancellation by the MWF is less affected.

3.6 Sensitivity to speech detection errors on the performance

Both the GSC and the MWF rely on a speech detection algorithm. Since the MWF does not require any other a priori information, the reliance on the speech detection is expected to be crucial to achieve the potentially better performance. In this section, we analyze the average effect of speech detection errors on the performance of the MWF, the GSC and the NSC-GSC. As will

be detailed below, this can be done by modifying the correlation matrices $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}(f)^H\}$ and $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}(f)^{n,H}\}$ in the formulas (3.12) and (3.25) for the ANC of the GSC and the MWF. To simplify the analysis, we assume that the second order statistics of the noise are stationary.

Again, the theoretical results are illustrated for a 3-microphone linear endfire array with microphone interspacings $d_2 - d_1 = 0.01$ m and $d_3 - d_2 = 0.015$ m and the multiple noise source scenario with five localized noise sources at 75°, 150°, 180°, 240°, 285° and $\beta(f) = -30$ dB.

3.6.1 Multi-channel Wiener Filter

Speech + noise samples being erroneously detected as noise-only samples

Suppose that $(\delta \times 100)\%$ of the samples used by the MWF to estimate the noise correlation matrix $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ are actually speech + noise samples. Assuming that the average PSDs $P_X^s(f)$ of the speech in the $(\delta \times 100)\%$ wrongly detected speech + noise samples and the speech in the correctly detected speech + noise samples are the same[6], then the estimated noise correlation matrix $\mathcal{E}\{\mathbf{X}_{est}^n(f)\mathbf{X}_{est}^{n,H}(f)\}$ equals

$$\begin{aligned}\mathcal{E}\{\mathbf{X}_{est}^n(f)\mathbf{X}_{est}^{n,H}(f)\} &= \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\} + \delta\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^{s,H}(f)\} \\ &= P_X^s(f)\left(\alpha(f)\mathbf{\Gamma}^n(f) + \delta\mathbf{\Gamma}^{\boldsymbol{s}}(f)\right) \qquad (3.59)\end{aligned}$$

with $1/\alpha(f)$ the input SNR ratio (cf. (3.29)). Using (3.59) in (3.25) and assuming a homogeneous speech and noise field, we find that the MWF $\mathbf{W}(f)$ changes into

$$\boxed{\mathbf{W}(f) = \frac{(1-\delta)P_X^s(f)}{P_X^s(f) + P_X^n(f)}\mathbf{\Gamma}^{-1}(f)\mathbf{\Gamma}^s(f)\mathbf{e}_1,} \qquad (3.60)$$

with $\mathbf{\Gamma}(f)$ and $\mathbf{\Gamma}^s(f)$ the coherence functions in case of a perfect speech detection.

Hence, the PTFs $G^s(f)$ and $G^n(f)$ of the MWF are both scaled by $(1-\delta)^2$. *For*[7] $\delta < 1$*, the SNR improvement* $G^s(f)/G^n(f)$ *at frequency* f *remains unchanged, but the speech distortion increases with increasing speech detection error rate* δ. In practice, the average speech PSD $P_X^s(f)$ of the correctly and wrongly detected speech + noise samples may differ, resulting in a frequency dependent $\delta(f)$ and attenuation/distortion (see also footnote 6), with an average effect as

[6] In practice, the average PSDs $P_X^s(f)$ can be different. This corresponds to replacing δ by a frequency dependent $\delta(f) = \delta.\kappa(f)$, where $\kappa(f)$ is the ratio of the average speech PSD of the wrongly detected speech + noise samples to the average PSD of the correctly detected speech + noise samples (cf. [242]).

[7] Note that $\delta = 1$ implies that all noise samples are detected as speech + noise samples and all speech + noise samples are detected as noise-only samples.

derived above. The distortion that is caused by the erroneous speech detection, is independent of the input SNR and the noise scenario. Consequently, the sensitivity of the MWF to erroneous speech detection is not proportional to the noise sensitivity $\Psi(f)$.

Noise samples being erroneously detected as speech + noise samples

A similar reasoning can be applied when $(\delta \times 100)\%$ of the samples used to estimate the speech + noise correlation matrix $\mathcal{E}\{\mathbf{X}(f)\mathbf{X}^H(f)\}$ are actually *noise-only* samples. The estimate $\mathcal{E}\{\mathbf{X}_{est}(f)\mathbf{X}_{est}^H(f)\}$ then becomes[8]

$$\begin{aligned}\mathcal{E}\{\mathbf{X}_{est}(f)\mathbf{X}_{est}^H(f)\} &= (1-\delta)\mathcal{E}\{\mathbf{X}(f)\mathbf{X}(f)\} + \delta\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^n(f)\} \\ &= (1-\delta)\mathcal{E}\{\mathbf{X}^s(f)\mathbf{X}^s(f)\} + \mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^n(f)\} \quad (3.61)\end{aligned}$$

so that $\mathbf{W}(f)$ equals

$$\boxed{\mathbf{W}(f) = \frac{P_X^s(f)}{P_X^s(f) + \frac{1}{1-\delta}P^n(f)}\tilde{\mathbf{\Gamma}}^{-1}(f)\mathbf{\Gamma}^{\boldsymbol{s}}(f)\mathbf{e}_1,} \quad (3.62)$$

with

$$\tilde{\mathbf{\Gamma}}(f) = \frac{\mathbf{\Gamma}^s(f) + \tilde{\alpha}(f)\mathbf{\Gamma}^n(f)}{1+\tilde{\alpha}(f)}; \qquad \tilde{\alpha}(f) = \frac{\alpha(f)}{1-\delta}. \quad (3.63)$$

The MWF acts as if the input SNR $1/\alpha(f)$ is reduced by a factor $1/(1-\delta)$. As a consequence, the filter will pay more attention to noise reduction than to speech distortion, so that the speech signal is attenuated more than in case of a perfect speech detection. For an excessive error rate $(\delta \times 100\%) = 50\%$, the effect corresponds to a decrease in input SNR of only 3 dB, so that the *effect of noise detected as speech + noise is found to be limited.* As an illustration, **Figure 3.18** depicts the performance of the MWF, i.e., $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$, as a function of δ for two external noise-to-signal ratios $\alpha(f)$, i.e., $\alpha(f) = 0$ dB and $\alpha(f) = -6$ dB. In free field, the improvement $\Delta\text{SNR}_{\text{intellig}}$ is not affected by the speech detection errors, while the speech distortion $\text{SD}_{\text{intellig}}$ only slightly grows with increasing δ.

In conclusion, the MWF is especially affected when speech + noise is detected as noise-only. In that case, the noise reduction performance of the MWF is preserved, but the speech distortion increases with increasing speech detection error rate.

3.6.2 Generalized Sidelobe Canceller

If the blocking matrix $\mathbf{B}(f)$ of the GSC would be perfect so that the noise references do not contain any speech, the performance of the GSC would be

[8] Here, δ is f-independent, since the second order statistics of the noise are assumed to be stationary.

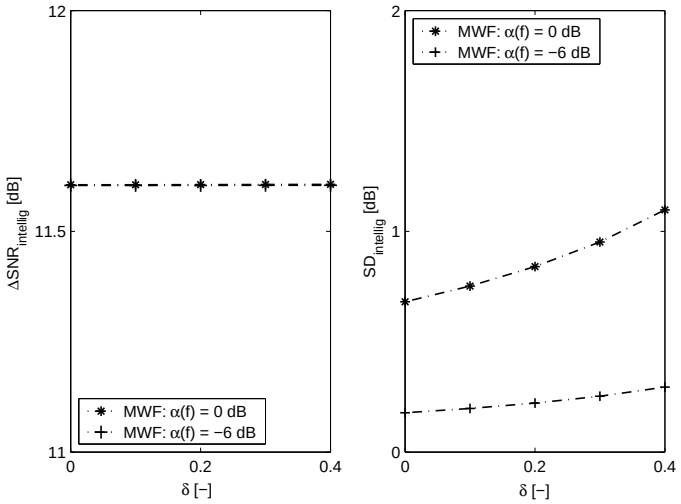


Figure 3.18: Theoretically optimal performance ($\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$) of the MWF, when ($\delta \times 100$) percent of the samples used to compute the speech + noise correlation matrix are actually noise-only samples.

independent of the speech detection. However, in practice, speech leakage into the noise references occurs. To then prevent speech cancellation at high input SNR [199], the ANC is adapted during periods of noise-only [74, 258, 103] and/or a constraint on the noise sensitivity of the GSC is applied [103, 29, 96]. In Section 3.4, we have shown that the latter is crucial for small-sized arrays to guarantee the performance of the GSC, even when the ANC is only adapted during noise-only periods (cf. Section 3.4). Still, in the event of signal leakage, and whenever speech + noise is detected as noise, the ANC may partially cancel the desired speech signal. On the other hand, whenever noise is detected as speech, the performance of the GSC will, apart from a slower convergence rate, not degrade, since the ANC is not adapted during speech + noise periods.

Suppose that $\delta \times 100\%$ of the samples used by the ANC to compute the noise correlation matrix are actually speech + noise samples. The formulas of the ANC filters are then obtained by replacing $\mathcal{E}\{\mathbf{X}^n(f)\mathbf{X}^{n,H}(f)\}$ in (3.12) with (3.59), resulting in

$$\mathbf{W}_a(f) = \left(\mathbf{B}^H(f)\tilde{\mathbf{\Gamma}}^n(f)\mathbf{B}(f)\right)^{-1}\mathbf{B}^H(f)\tilde{\mathbf{\Gamma}}^n(f)\mathbf{A}(f) \tag{3.64}$$

with

$$\tilde{\mathbf{\Gamma}}^n(f) = \mathbf{\Gamma}^n(f) + \frac{\delta}{\alpha(f)}\mathbf{\Gamma}^s(f). \tag{3.65}$$

The impact of an erroneous speech detection on the GSC depends on the quality of the noise references $Y_l(f) = \mathbf{B}^H(l,:)\mathbf{X}(f)$, $l = 1, ..., M-1$: the larger the

ratio of speech leakage $Y_l^s(f)$ to noise $Y_l^n(f)$ in the noise references, the larger the impact. From (3.64)-(3.65), it follows that the higher the SNR, i.e. the lower $\alpha(f)$, the more the ANC filters and hence the performance will be affected when model errors are present.

The ANC $\mathbf{W}_a(f)$ of the NSC-GSC is obtained by replacing $\mathbf{\Gamma}^n(f)$ in (3.65) by (3.43).

3.6.3 Comparison between GSC, NSC-GSC and MWF

Figure 3.19 depicts the impact of speech detection errors on the performance of the GSC, the NSC-GSC with $\eta = -10$ dB and the MWF when $\delta \times 100$ % of the samples used to compute the ANC and the MWF are actually speech + noise samples. Two input SNRs $1/\alpha(f)$ are depicted, i.e., $\alpha(f) = 0$ dB (dashed-dotted lines) and $\alpha(f) = -6$ dB (dotted lines). In the simulations, the second microphone has a small gain mismatch υ of 1.5 dB.

In the presence of signal model errors, the GSC is more sensitive to speech detection errors than the MWF. In addition to increasing robustness to model errors [29], the noise sensitivity constraint of the NSC-GSC reduces the drastic impact of speech detection errors on the GSC, especially at high input SNR $1/\alpha(f)$. This can be understood as follows. If the model errors are not too large, the noise reference of the GSC contains less speech than the speech reference. To cancel the speech effectively the ANC filters have to amplify the speech in the noise reference, resulting in a high noise sensitivity. A noise sensitivity constraint limits the filter amplitudes of the ANC and hence the speech cancellation.

The MWF additionally distorts the speech by $10\log_{10}(1-\delta)^2$ but conserves the improvement $\Delta\text{SNR}_{\text{intellig}}$. In contrast to the GSC, this additional distortion (i.e. $10\log(1-\delta)^2$), is independent of the input SNR. Compared to the NSC-GSC with $\eta = -10$ dB, the speech signal is attenuated more, especially for large δ. For error rates $\delta < 0.2$, the distortion is limited, while more noise is reduced.

3.6.4 Experimental validation

In this section, we verify the theoretical results based on real recordings with the BTE. To predict the average effect of speech detection errors, the desired and noise signals are stationary with the long-term PSD of speech. The signals were taken from the HINT-database [185]. The first half of the samples contain speech + noise, the other half noise-only. In [33] , the effect of speech detection errors on the MWF has been analyzed experimentally by means of real speech signals, leading to similar results.

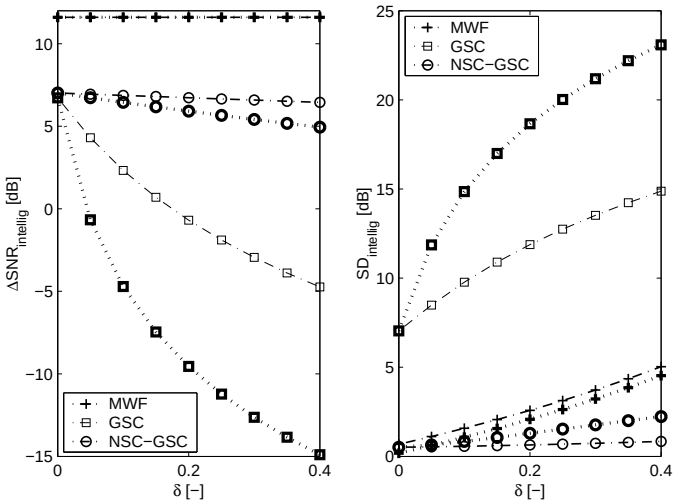


Figure 3.19: Theoretical performance ($\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$) of the MWF, the GSC and the NSC-GSC with $\eta = -10$ dB when $(\delta \times 100)\%$ of the samples used to compute the noise correlation matrix are actually speech + noise samples. Two input SNRs $1/\alpha(f)$ are depicted: $\alpha(f) = 0$ dB (dashed-dotted lines) and $\alpha(f) = -6$ dB (dotted lines).

Noise samples wrongly detected as speech + noise samples

Figure 3.20 depicts the performance of the MWF when $(\delta \times 100)\%$ of the samples used to compute the speech + noise correlation matrix are actually noise-only samples. Two input noise-to-signal ratios $\alpha(f)$ are considered: $\alpha(f) = 0$ dB and $\alpha(f) = -6$ dB. The behavior of the curves is quite well predicted by Figure 3.18: an increased error rate δ only leads to a small increase in distortion $\text{SD}_{\text{intellig}}$. The increase in distortion $\text{SD}_{\text{intellig}}$ grows with increasing $\alpha(f)$. The noise reduction performance $\Delta\text{SNR}_{\text{intellig}}$ is not significantly affected by the speech detection errors. In contrast to the results of Figure 3.18, the improvement $\Delta\text{SNR}_{\text{intellig}}$ drops for decreasing α. This difference is due to the presence of reverberation in the speech signal. The Wiener filter trades off noise reduction and cancellation of the reverberant part in the speech signal. For large $\alpha(f)$, the MWF will pay more attention to noise reduction. The improvement $\Delta\text{SNR}_{\text{intellig}}$ also slightly increases as a function of δ since a larger δ corresponds to an increase of $\alpha(f)$ (cf. (3.62)-3.63) in Section 3.6.1).

Speech + noise samples wrongly detected as noise-only samples

Figure 3.21 shows the performance of the GSC, the NSC-GSC (with $\eta = -10$ dB) and the MWF when $(\delta \times 100)\%$ of the samples used to compute the noise correlation matrix, are actually speech + noise samples. In the experi-

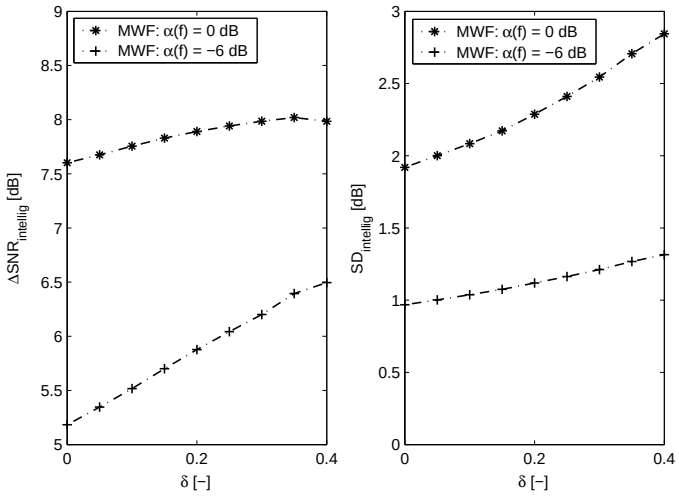


Figure 3.20: $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the MWF, when $(\delta \times 100)\%$ of the samples used to compute the speech + noise correlation matrix are actually noise-only samples.

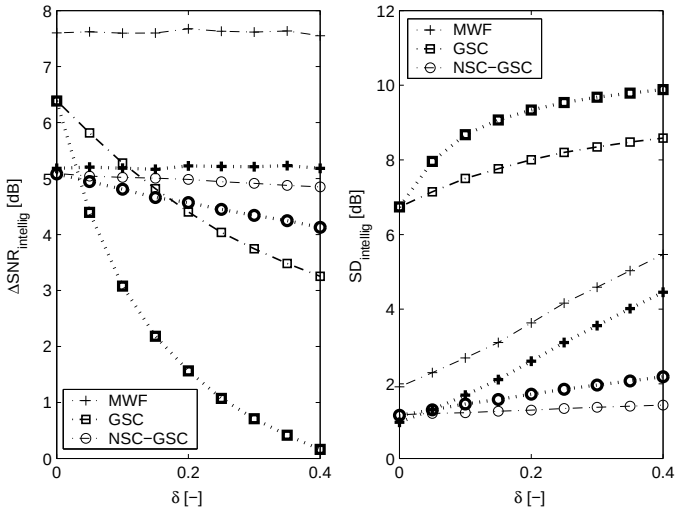


Figure 3.21: $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the MWF, the GSC and the NSC-GSC with $\eta = -10$ dB when $(\delta \times 100)\%$ of the samples used to compute the noise correlation matrix are actually speech + noise samples. Two input SNRs $1/\alpha(f)$ are depicted: $\alpha(f) = 0$ dB (dashed-dotted lines) and $\alpha(f) = -6$ dB (dotted lines).

ments, a gain deviation of 1.5 dB has been applied to the second microphone. Although differences occur between Figure 3.21 and Figure 3.19 due to reverberation and a different $\beta(f)$, the experimental results are quite well predicted by the theory. In the presence of signal model errors, the MWF is less sensitive to speech detection errors than the GSC. The improvement $\Delta\text{SNR}_{\text{intellig}}$ of the MWF is (again) hardly affected by erroneous speech detection, but $\text{SD}_{\text{intellig}}$ increases on average with $10\log_{10}(1-\delta)^2$. For error rates $\delta < 0.2$, the distortion by the MWF is limited, while more noise is reduced compared to the NSC-GSC with $\eta = -10$ dB, especially for large $\alpha(f)$.

In conclusion, theoretical and experimental analysis showed that the noise reduction performance of the MWF is, in contrast to the GSC and the NSC-GSC, not degraded by speech detection errors, but the speech gets distorted when speech + noise samples are wrongly detected as noise-only. The effect of noise samples detected as speech + noise, on the other hand, is negligible. For a reasonable speech detection error rate of 20% or less, the distortion by the MWF is limited and hence the MWF preserves its potential benefit over the GSC, even when the latter is supplied with a robustness constraint.

3.7 Conclusions

In this chapter, we presented a theoretical framework to assess both the broadband performance of the GSC, the NSC-GSC and the MWF and their robustness to different types of model errors for different noise scenarios. Both errors in the assumed desired signal model and inaccurate estimation of the second order statistics have been considered. The robustness has been analyzed for small-sized arrays as used in hearing aids and in terms of speech intelligibility weighted performance measures. The theoretical results have been verified through experiments with a three-microphone BTE.

Theoretical analysis of the noise sensitivity showed that sensitivity to errors in the assumed signal model grows with an increasing complexity of the noise scenario, with decreasing $f \cdot d$ and increasing number of microphones per array length. We have shown that because of the high noise sensitivity of small-sized arrays, the number of useful microphones in a commercial BTE is limited to three, provided that the noise reduction algorithm is sufficiently robust. Among the different signal model errors, microphone gain mismatch is particularly harmful to GSC, even if the ANC is adapted during noise-only. As a consequence, a noise sensitivity constraint is essential, which goes at the expense of less noise reduction. The MWF on the other hand, is not degraded by signal model errors, so that more noise is reduced compared to a robust NSC-GSC. Because the MWF does not require a noise sensitivity constraint, the performance benefit of a third microphone in a BTE can be fully exploited.

Since the MWF does not require a priori information about the desired signal model, it is expected to be more sensitive to inaccurate estimation of the second order statistics caused by, e.g., non-stationarities in the noise or speech detection errors. Experiments demonstrated that the MWF preserves its benefit in robustness and performance over the GSC and the NSC-GSC for spectrally short-time non-stationary noise, provided that the spatial and the *long-term* spectral characteristics of the noise vary relatively slowly in time (on the order of seconds). In addition, it has been shown both theoretically and experimentally that the noise reduction performance of the MWF is, in contrast to the GSC and the NSC-GSC, not degraded by speech detection errors, but that the speech gets distorted when speech + noise samples are wrongly detected as noise-only. For a reasonable speech detection error rate of 20% or less, the distortion by the MWF is limited, while more noise is reduced compared to the NSC-GSC.

For all types of model errors that have been studied, the experimental results confirmed the theoretically predicted ones. Hence, the analytical framework indeed offers an easy way to predict the impact of specific model errors on the performance of the noise reduction algorithms.

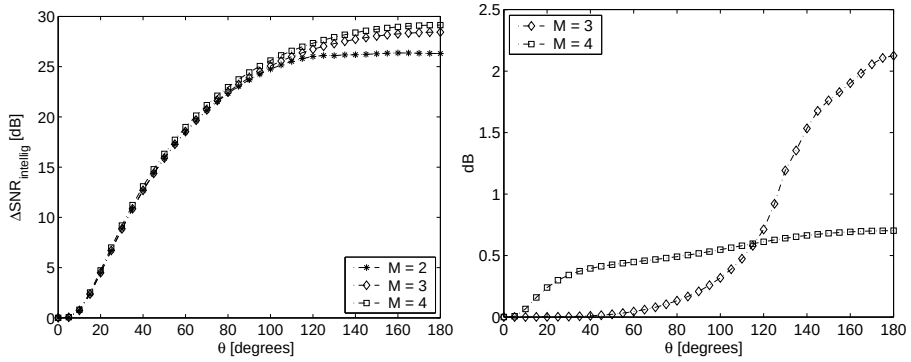
3.A Appendix: Influence of the number of microphones

In the case of a single noise source, two microphones theoretically (i.e., in the absence of reverberation) suffice to cancel one noise source effectively, provided that no spatial aliasing occurs (i.e., $d < \frac{2c}{f}$ for $f \leq \frac{f_s}{2}$). As a result, the benefit of additional microphones in a BTE will be limited in the absence of reverberation.

To illustrate this, **Figure 3.A.22(a)** depicts the theoretical improvement $\Delta\text{SNR}_{\text{intellig}}$ of the GSC and MWF that is obtained with 2, 3 and 4 microphones for a single noise source with angle θ. Remember from Section 3.2.2 that in the absence of signal model errors and for a single speech source, the performance of the GSC and the MWF coincide. The increase in $\Delta\text{SNR}_{\text{intellig}}$ that is obtained with an additional third ($M = 3$) and fourth microphone ($M = 4$) is shown in **Figure 3.A.22(b)**. For $-100° < \theta < 100°$, the benefit of an additional third and fourth microphone is indeed negligible. For angles around 180°, the gain of an extra third microphone is larger, which is due to spatial aliasing. For $M = 2$, spatial aliasing occurs at frequencies $f \geq \frac{c}{2A_l} = 5667$ Hz. E.g., for $f = \frac{c}{2A_l}$, the angles 0° and 180° can not be distinguished from each other because for $A_l = d = \frac{c}{2.f}$

$$2\pi f \frac{d}{c}\cos(0) \quad = \quad 2\pi f \frac{d}{c}\cos\pi + 2\pi.$$

This explains the larger benefit of an extra third microphone for θ around 180°.



(a) (b)

Figure 3.A.22: Influence of the number of microphones M on $\Delta\text{SNR}_{\text{intellig}}$ of the GSC and MWF for a localized noise source at θ; $\beta(f) = -30$ dB. (a) $\Delta\text{SNR}_{\text{intellig}}$ for $M = 2$, $M = 3$ and $M = 4$; (b) Additional improvement in $\Delta\text{SNR}_{\text{intellig}}$ thanks to an additional 3rd and 4th microphone for $\beta = -30$ dB .

Chapter 4

Spatially pre-processed speech distortion weighted MWF

IN this chapter, we establish a generalized noise reduction scheme, called the spatially pre-processed speech distortion weighted multi-channel Wiener filter (SP-SDW-MWF), that encompasses the GSC and the MWF as extreme cases. In addition, the scheme allows for in-between solutions such as the speech distortion regularized GSC (SDR-GSC). The SDR-GSC or more generally the SP-SDW-MWF adds robustness to the GSC by taking speech distortion explicitly into account in the design criterion of the adaptive stage. Compared to the widely studied GSC with quadratic inequality constraint (cf. Section 2.3.3), the SP-SDW-MWF achieves a better noise reduction performance, for a given maximum speech distortion level.

In Section 4.1, the MMSE criterion of the MWF is generalized to allow for a trade-off between speech distortion and noise reduction. We refer to this generalization as speech distortion weighted MWF (SDW-MWF). In addition, an equivalent optimization criterion for the SDW-MWF is described and possible recursive implementations are briefly discussed.

In Section 4.2, we establish a common framework for the GSC and the SDW-MWF. We show that both algorithms can be integrated into one generalized scheme that consists of a fixed, spatial pre-processor and an adaptive SDW-MWF stage, hence the name *Spatially Pre-processed Speech Distortion Weighted MWF*. The SP-SDW-MWF allows to understand the differences and links between the GSC and the (SDW-)MWF more thoroughly. In addition, in-

between solutions are found where, like in the GSC, no filtering is applied on the speech reference. These solutions can be interpreted as a Speech Distortion Regularized GSC: the ANC design criterion is then supplemented with a regularization term that limits speech distortion due to signal model errors. In the literature, the robustness of the GSC against errors in the assumed signal model is often increased by imposing a QIC on the ANC filters [29, 93, 96, 103, 255], at the expense of noise reduction.

In Section 4.3, we show that the SDR-GSC, in contrast to the QIC-GSC, guarantees robustness at large signal model errors while less affecting the noise reduction performance in the presence of small model errors. In addition, the extra filtering of the speech reference signal in the SP-SDW-MWF further improves the performance that is achieved by the SDR-GSC and the QIC-GSC in the presence of signal model errors. We prove that in the absence of model errors and for infinite filter lengths, the SP-SDW-MWF corresponds to a cascade of an SDR-GSC with a speech distortion weighted single-channel Wiener filter. In contrast to the SDR-GSC and the QIC-GSC, its performance is shown not to degrade due to microphone mismatch.

4.1 Speech distortion weighted MWF

In this section, the MMSE criterion of the MWF is generalized to allow for a trade-off between noise reduction and speech distortion. We refer to this generalization as speech distortion weighted MWF (SDW-MWF). In addition, we formulate an equivalent optimization criterion for the SDW-MWF and briefly discuss possible recursive implementations.

4.1.1 Trade-off speech distortion versus noise reduction

Recall from Section 2.4.1 that the MWF $\mathbf{w}$ provides an MMSE estimate of the speech component $x_l^s[k-\Delta]$ in the l-th microphone signal, i.e.,[1]

$$\mathbf{w} = \arg\min_{\mathbf{w}} \mathcal{E}\left\{\left|x_l^s[k-\Delta] - \mathbf{w}^H\mathbf{x}[k]\right|^2\right\}, \tag{4.1}$$

$$= \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\}^{-1}\mathcal{E}\{\mathbf{x}^s[k]x_l^{s,*}[k-\Delta]\}. \tag{4.2}$$

The residual error energy of the MWF equals

$$\mathcal{E}\{|e[k]|^2\} = \mathcal{E}\{\left|x_l^s[k-\Delta] - \mathbf{w}^H\mathbf{x}[k]\right|^2\}, \tag{4.3}$$

[1] In a time-domain implementation, the filter $\mathbf{w}$ and its input signals $x_l[k]$ are real. Hence, $\mathbf{w}_{1:M-1}^H = \mathbf{w}_{1:M-1}^T$ and $\mathbf{x}^H[k] = \mathbf{x}^T[k]$. In the sequel, the time-domain formulas are generalized to complex input signals so that they can also be applied in a subband implementation.

and can be decomposed into

$$\underbrace{\mathcal{E}\{\left|x_l^s[k-\Delta]-\mathbf{w}^H\mathbf{x}^s[k]\right|^2\}}_{\varepsilon_s^2}+\underbrace{\mathcal{E}\{\left|\mathbf{w}^H\mathbf{x}^n[k]\right|^2\}}_{\varepsilon_n^2} \tag{4.4}$$

where ε_s^2 equals the speech distortion energy and ε_n^2 the residual noise energy.

Hence, the design criterion (4.1) of the MWF can be easily generalized to allow for a trade-off between speech distortion and noise reduction, by incorporating a weighting factor μ [36] with $\mu \in [0,\infty]$

$$\boxed{\mathbf{w}=\arg\min_{\mathbf{w}}\mathcal{E}\{\left|x_l^s[k-\Delta]-\mathbf{w}^H\mathbf{x}^s[k]\right|^2\}+\mu\mathcal{E}\{\left|\mathbf{w}^H\mathbf{x}^n[k]\right|^2\}.} \tag{4.5}$$

The solution of (4.5) is given by

$$\boxed{\mathbf{w}=\mathcal{E}\{\mathbf{x}^s[k]\mathbf{x}^{s,H}[k]+\mu\mathbf{x}^n[k]\mathbf{x}^{n,H}[k]\}^{-1}\mathcal{E}\{\mathbf{x}^s[k]x_l^{s,*}[k-\Delta]\},} \tag{4.6}$$

which corresponds to the Wiener filter (4.2) with an adjustable input noise level. The filter (4.6) is an extension of the single-channel time domain constrained estimator proposed in [54], which minimizes the speech distortion ε_s^2 while keeping the residual noise energy ε_n^2 below some threshold $T=\alpha\mathcal{E}\{\mathbf{x}^{n,H}[k]\mathbf{x}^n[k]\}$:

$$\min_{\mathbf{w}}\varepsilon_s^2 \quad \text{subject to } \varepsilon_n^2 \leq \alpha\mathcal{E}\{\mathbf{x}^{n,H}[k]\mathbf{x}^n[k]\} \tag{4.7}$$

where $0\leq\alpha\leq 1$ and μ is the Lagrange-multiplier [36].

The factor $\mu\in[0,\infty]$ trades off speech distortion versus noise reduction, hence, the name SDW-MWF: the smaller μ, the smaller the resulting speech distortion. If $\mu=1$, the MMSE criterion (4.1)-(4.2) is obtained. If $\mu>1$, the residual noise level will be reduced at the expense of increased speech distortion. By setting μ to ∞, all emphasis is put on noise reduction and speech distortion is completely ignored, resulting in $\mathbf{w}=0$. Setting μ to 0, on the other hand, results in $\mathbf{w}=\mathbf{e}_{(l-1)L+\Delta+1}$ where $\mathbf{e}_l$ is the l-th canonical $ML\times 1$-dimensional vector

$$\mathbf{e}_l=\begin{bmatrix}0 & \cdots & 0 & \underbrace{1}_{\text{position } l} & 0 & \cdots & 0\end{bmatrix}^T, \tag{4.8}$$

and hence, no noise reduction.

4.1.2 Equivalent optimization criterion

The SDW-MWF (4.6) provides an estimate of the (delayed) speech component in the l-th microphone signal. An equivalent SDW-MWF approach, whose structure shows some resemblance with an ANC, consists in estimating a delayed version of the (unknown) noise component $x_l^n[k-\Delta]$. The estimate $z[k]$

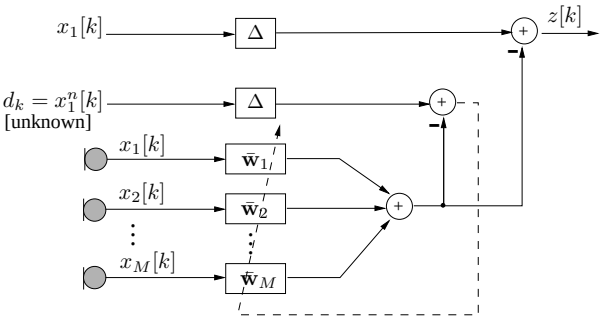


Figure 4.1: Equivalent SDW-MWF approach.

of the speech component $x_l^s[k-\Delta]$ is then obtained by subtracting the estimate $\bar{\mathbf{w}}^H\mathbf{x}[k]$ of $x_l^n[k-\Delta]$ from the delayed l-th microphone signal $x_l[k-\Delta]$, i.e.,

$$z[k] = x_l[k-\Delta] - \bar{\mathbf{w}}^H[k]\mathbf{x}[k] = (\mathbf{e}_{(l-1)L+\Delta+1} - \bar{\mathbf{w}})^H\mathbf{x}[k]. \tag{4.9}$$

This is illustrated in **Figure 4.1** for the MWF $\bar{\mathbf{w}}$ with $x_l^n[k-\Delta] = x_1^n[k-\Delta]$.

The optimization criterion now becomes

$$\bar{\mathbf{w}}[k] = \arg\min_{\bar{\mathbf{w}}} \mathcal{E}\left\{\left|x_l^n[k-\Delta] - \bar{\mathbf{w}}^H\mathbf{x}[k]\right|^2\right\}, \tag{4.10}$$

for the MWF $\bar{\mathbf{w}}[k]$ or, more generally,

$$\boxed{\bar{\mathbf{w}}[k] = \arg\min_{\bar{\mathbf{w}}} \mathcal{E}\left\{\left|x_l^n[k-\Delta] - \bar{\mathbf{w}}^H[k]\mathbf{x}^n[k]\right|^2 + \frac{1}{\mu}\left|\bar{\mathbf{w}}^H[k]\mathbf{x}^s[k]\right|^2\right\}} \tag{4.11}$$

for the SDW-MWF, leading to

$$\boxed{\bar{\mathbf{w}} = \left(\mathcal{E}\left\{\mathbf{x}^n[k]\mathbf{x}^{n,H}[k]\right\} + \frac{1}{\mu}\mathcal{E}\left\{\mathbf{x}^s[k]\mathbf{x}^{s,H}[k]\right\}\right)^{-1}\mathcal{E}\{\mathbf{x}^n[k]x_l^{n,*}[k-\Delta]\}.} \tag{4.12}$$

Using (4.6) and (4.12), it can be easily shown that

$$\boxed{\mathbf{w}[k] + \bar{\mathbf{w}}[k] = \mathbf{e}_{(l-1)L+\Delta+1}.} \tag{4.13}$$

This demonstrates that the two approaches indeed lead to exactly the same speech signal estimate $z[k]$. Again, μ trades off between noise reduction and speech distortion. For $\mu = \infty$, $\bar{\mathbf{w}}$ equals $\mathbf{e}_{(l-1)L+\Delta+1}$, while for $\mu = 0$, $\bar{\mathbf{w}} = \mathbf{0}$.

4.1.3 Implementation of SDW-MWF

Because of double-talk of speech and noise, the speech correlation matrix $\mathcal{E}\left\{\mathbf{x}^s[k]\mathbf{x}^{s,H}[k]\right\}$ is not available in practice. Like the MWF, the SDW-MWF estimates $\mathcal{E}\left\{\mathbf{x}^s[k]\mathbf{x}^{s,H}[k]\right\}$ as (cf. Section 2.4.1)

$$\mathcal{E}\{\mathbf{x}^s[k]\mathbf{x}^{s,H}[k]\} = \mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\} - \mathcal{E}\{\mathbf{x}^n[k]\mathbf{x}^{n,H}[k]\}, \tag{4.14}$$

resulting in

$$\begin{aligned}\mathbf{w}[k] &= \left(\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^H[k]\} + (\mu-1)\mathcal{E}\{\mathbf{x}^n[k]\mathbf{x}^{n,H}[k]\}\right)^{-1} \\ &\quad \cdot \left(\mathcal{E}\{\mathbf{x}[k]x_l^*[k-\Delta]\} - \mathcal{E}\{\mathbf{x}^n[k]x_l^{n,*}[k-\Delta]\}\right)\end{aligned} \tag{4.15}$$

and[2]

$$\bar{\mathbf{w}}[k] = \left(\frac{1}{\mu}\mathcal{E}\{\mathbf{x}\mathbf{x}^H\} + (1-\frac{1}{\mu})\mathcal{E}\{\mathbf{x}^n\mathbf{x}^{n,H}\}\right)^{-1}\mathcal{E}\{\mathbf{x}^n[k]x_l^{n,*}[k-\Delta]\}. \tag{4.16}$$

The recursive GSVD and QRD based implementations of the MWF (cf. Section 2.4.2) can be extended so that they apply to the SDW-MWF with $\mathbf{w}[k]$ or $\bar{\mathbf{w}}[k]$ [36, 212]. In the *QRD based implementations*, the computational complexity in Table 2.2 and Table 2.3 is then doubled and storage of noise samples in a circular noise buffer is required. In the (more complex) *GSVD based implementation*, the trade-off factor μ can be easily incorporated in the filter update equation as is illustrated in Example 4.1 below, with only a minor increase in complexity (i.e., $\mathcal{O}(LM)$) and without requiring additional storage of noise samples in a noise buffer. Again, a subband implementation results in a reduced complexity and a better intelligibility weighted performance.

Example 4.1 Using (2.112)-(2.114) of Section 2.4.2 in (4.16), it can be easily shown that the GSVD based SDW-MWF $\bar{\mathbf{w}}[k]$ correspond to the i-th column with $i = (l-1)L + \Delta + 1$ of the filter matrix $\bar{\mathbf{W}}_k$, defined as

$$\bar{\mathbf{W}}_k^* \simeq \mathbf{Q}^{-H}\text{diag}\left\{\min\left(\frac{\frac{p}{q}\frac{\eta_i^2}{\sigma_i^2}}{\frac{1}{\mu}+(1-\frac{1}{\mu})\frac{p}{q}\frac{\eta_i^2}{\sigma_i^2}}, 1\right)\right\}\mathbf{Q}. \tag{4.17}$$

△

In [57], a *stochastic gradient implementation* has been proposed for the SDW-MWF $\mathbf{w}$ (4.15) with $\mu > 1$, using circular noise and speech + noise buffers. This algorithm however relies on some independence assumptions that are not necessarily satisfied, so that the performance is seriously degraded (cf. Appendix 5.A).In Chapter 5, we will derive alternative stochastic gradient implementations for the SDW-MWF $\bar{\mathbf{w}}$ (4.16) that have a better performance.

[2]In the sequel, the time index k in $\mathbf{x}[k]$ is omitted for the sake of conciseness, unless it enhances readability.

Remark 4.1 Instead of estimating $\mathcal{E}\{\mathbf{x}^s[k]\mathbf{x}^{s,H}[k]\}$ online using (4.14), the correlation matrix $\mathcal{E}\{\mathbf{x}^s[k]\mathbf{x}^{s,H}[k]\}$ could be estimated from clean speech recordings measured during an initial calibration phase (cf. the MWF with calibration described in Section 2.4)[189]. Additional recordings of the speech source signal allow to produce an estimate of the non-reverberant speech source signal instead of an estimate of the reverberant speech component in one of the microphone signals. In [189], an LMS based algorithm for this MWF has been developed. However, since the room acoustics, the position of desired speaker and microphone characteristics may change over time, frequent re-calibration is required, making this approach cumbersome and expensive. □

4.2 Spatially pre-processed SDW-MWF

In this section, we show how the GSC and the (SDW-)MWF can be integrated into one generalized noise reduction scheme, called the spatially pre-processed SDW-MWF. In addition, we demonstrate that this generalized scheme allows for in-between solutions that can be interpreted as a speech distortion regularized GSC (SDR-GSC).

4.2.1 Concept

The structure of the equivalent SDW-MWF $\bar{\mathbf{w}}$ (cf. Figure 4.1) strongly resembles to the structure of an ANC. Replacing the ANC in the GSC with the SDW-MWF $\bar{\mathbf{w}}$ results in the generalized noise reduction scheme depicted in **Figure 4.2**, which consists of a spatial pre-processor, i.e., a fixed beamformer $\mathbf{a} \in \mathbb{R}^{ML_A \times 1}$ and a blocking matrix $\mathbf{B} \in \mathbb{R}^{ML_B \times M-1}$, and an SDW-MWF. In the sequel, we will refer to this generalized scheme as the *spatially pre-processed speech distortion weighted multi-channel Wiener filter (SP-SDW-MWF).*

Spatial pre-processor

The input signals to the SDW-MWF are now spatially pre-processed microphone signals and a delayed version of the speech component in the speech reference $y_0[k]$ is estimated. Like in the GSC (cf. Section 2.3.1), the fixed beamformer $\mathbf{a}$ creates a speech reference $y_0[k]$ by steering a beam towards the front (i.e., 0°) and the blocking matrix $\mathbf{B}$ creates $M-1$ noise references $y_l[k]$, $l = 1, \cdots, M-1$, by steering zeroes towards the front. During periods of speech, the references $y_i[k]$ consist of speech + noise, i.e., $y_i[k] = y_i^s[k] + y_i^n[k]$, $i = 0, ..., M-1$. During periods of noise, only the noise component $y_i^n[k]$ is observed.

To preserve the robustness advantage of the MWF, the fixed beamformer $\mathbf{a}$ should be robust against possible signal model errors. In the sequel a DS

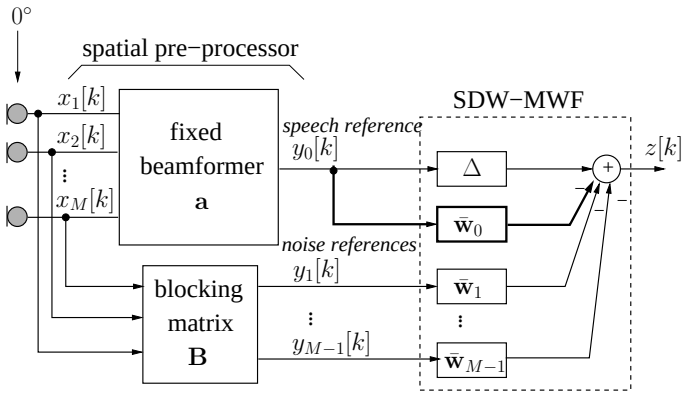


Figure 4.2: Spatially pre-processed SDW-MWF.

beamformer is used. As mentioned before in Section 2.2.3, this beamformer is sufficiently robust when used for small-sized arrays. A further optimized robust fixed beamformer can be designed using the techniques outlined in Section 2.2.5. The noise references are constructed by pairwise subtracting the microphone signals after they have been time-aligned for 0° (cf. Example 2.1 in Section 2.3.1). Further optimized noise references can be created with a spatial filter designed blocking matrix (cf. Section 2.3.3) and/or with the design methods for fixed robust broadband beamformers (cf. Section 2.2.5).

SDW-MWF

Let the SDW-MWF be denoted as the stacked filter vector $\bar{\mathbf{w}}_{0:M-1}[k]$

$$\bar{\mathbf{w}}_{0:M-1}[k] = \left[\begin{array}{cccc} \bar{\mathbf{w}}_0^H[k] & \bar{\mathbf{w}}_1^H[k] & \cdots & \bar{\mathbf{w}}_{M-1}^H[k] \end{array}\right]^H, \tag{4.18}$$

$$\bar{\mathbf{w}}_l[k] = \left[\begin{array}{cccc} \bar{w}_{l,0}[k] & \bar{w}_{l,1}[k] & \cdots & \bar{w}_{l,L-1}[k] \end{array}\right]^T, \tag{4.19}$$

where the subscript $0:M-1$ refers to the input channels of the SDW-MWF. If, like in the GSC, no filtering $\bar{\mathbf{w}}_0[k]$ is applied to the speech reference, the SDW-MWF will be referred to as $\bar{\mathbf{w}}_{1:M-1} \in \mathbb{C}^{(M-1)L\times 1}$, while the more general SDW-MWF with filter $\bar{\mathbf{w}}_0[k]$ will be denoted as $\bar{\mathbf{w}}_{0:M-1}[k] \in \mathbb{C}^{ML\times 1}$. Similarly, define the ML-dimensional stacked data vector $\mathbf{y}_{0:M-1}[k]$ as[3],

$$\mathbf{y}_{0:M-1}[k] = \left[\begin{array}{cccc} \mathbf{y}_0^H[k] & \mathbf{y}_1^H[k] & \ldots & \mathbf{y}_{M-1}^H[k] \end{array}\right]^H, \tag{4.20}$$

$$\mathbf{y}_l[k] = \left[\begin{array}{cccc} y_l[k] & y_l[k-1] & \ldots & y_l[k-L+1] \end{array}\right]^T. \tag{4.21}$$

[3]In the sequel, the time index k in $\mathbf{y}_{0:M-1}[k]$ is omitted for the sake of conciseness, unless it enhances readability.

The SDW-MWF $\bar{\mathbf{w}}_{0:M-1}[k]$, which now provides an estimate of a delayed version $y_0^n[k-\Delta]$ of the noise component in the speech reference, is obtained as the filter that minimizes the cost function $J(\bar{\mathbf{w}}_{0:M-1}[k])$ (cf. (4.11) of Section 4.1.2)

$$\boxed{\begin{aligned} J(\bar{\mathbf{w}}_{0:M-1}[k]) &= \underbrace{\mathcal{E}\{\left|y_0^n[k-\Delta] - \bar{\mathbf{w}}_{0:M-1}^H[k]\mathbf{y}_{0:M-1}^n[k]\right|^2\}}_{\varepsilon_n^2} \\ &\quad + \frac{1}{\mu}\underbrace{\mathcal{E}\{\left|\bar{\mathbf{w}}_{0:M-1}^H[k]\mathbf{y}_{0:M-1}^s[k]\right|^2\}}_{\varepsilon_s^2}, \end{aligned}} \tag{4.22}$$

resulting in

$$\boxed{\begin{aligned} \bar{\mathbf{w}}_{0:M-1}[k] = &\left(\frac{1}{\mu}\mathcal{E}\{\mathbf{y}_{0:M-1}^s\mathbf{y}_{0:M-1}^{s,H}\} + \mathcal{E}\{\mathbf{y}_{0:M-1}^n\mathbf{y}_{0:M-1}^{n,H}\}\right)^{-1} \\ &\cdot \mathcal{E}\{\mathbf{y}_{0:M-1}^n[k]y_0^{n,*}[k-\Delta]\}. \end{aligned}} \tag{4.23}$$

Again, the parameter μ trades off between distortion of the speech reference $y_0^s[k]$ and noise reduction. For $\boxed{\mu = 0}$, $\bar{\mathbf{w}}_{0:M-1} = \mathbf{0}$ and so $z[k]$ is equal to the output of the fixed beamformer $\mathbf{a}[k]$ delayed by Δ samples. In noise scenarios with very low SNR, e.g., -10 dB, a fixed beamformer output may be preferred: for low SNR, the error rate of the speech detection mechanism may be indeed too high, resulting in unacceptable speech distortion by the adaptive filter (cf. Section 3.6). Adaptivity can then easily be reduced or excluded in the SP-SDW-MWF by decreasing μ to 0. Alternatively, adaptivity can be limited by applying an QIC $\bar{\mathbf{w}}_{0:M-1}[k]^H\bar{\mathbf{w}}_{0:M-1}[k] \leq \zeta^2$ (cf. Section 2.3.3) to $\bar{\mathbf{w}}_{0:M-1}[k]$.

Remark 4.2 Note that when the fixed beamformer $\mathbf{a}$ and the blocking matrix $\mathbf{B}$ are set to

$$\mathbf{a} = \begin{bmatrix} 1 & 0 & ... & 0 \end{bmatrix}^H \in \mathbb{R}^{M\times 1} \tag{4.24}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}^H \in \mathbb{R}^{M\times M-1}, \tag{4.25}$$

we obtain the original SDW-MWF that operates on the received microphone signals $x_l[k]$, $l = 1, ..., M$. It should, however, be noted that estimating the (delayed) noise component $y_o^n[k-\Delta]$ in the output of the fixed beamformer $\mathbf{a}$ (e.g., a DS beamformer) instead of the noise component $x_l^n[k-\Delta]$ in the l-th microphone signal increases the robustness of the SDW-MWF against microphone failures. □

Remark 4.3 The adaptive filter $\bar{\mathbf{w}}_{0:M-1}[k]$ (4.23) can be implemented recursively using the (subband) GSVD or QRD based techniques for the SDW-MWF (cf. Section 4.1.3). The GSVD based implementation can however only be applied when $\bar{\mathbf{w}}_0 \neq \mathbf{0}$. In Chapter 5, cheaper stochastic gradient based algorithms will be developed for the SP-SDW-MWF. □

In the next sections, the different parameter settings of the SP-SDW-MWF are discussed. Depending on the setting of the trade-off parameter μ and the presence or absence of the filter $\bar{\mathbf{w}}_0$ on the speech reference, the GSC, the (SDW-)MWF as well as in-between solutions may be obtained. We distinguish between two cases, i.e., the case where no filter $\bar{\mathbf{w}}_0$ is applied to the speech reference (i.e., the filter length L_0 of $\bar{\mathbf{w}}_0$ equals 0) and the case where an additional filter $\bar{\mathbf{w}}_0$ is used ($L_0 \neq 0$).

4.2.2 Speech distortion regularized GSC

First, consider *the case without* $\bar{\mathbf{w}}_0$, i.e., $L_0 = 0$. The solution for $\bar{\mathbf{w}}_{1:M-1}$ in (4.22) then reduces to

$$\boxed{\arg \min_{\bar{\mathbf{w}}_{1:M-1}} \underbrace{\mathcal{E}\{\left|y_0^n[k-\Delta] - \bar{\mathbf{w}}_{1:M-1}^H \mathbf{y}_{1:M-1}^n[k]\right|^2\}}_{\varepsilon_n^2} + \frac{1}{\mu} \underbrace{\mathcal{E}\{\left|\bar{\mathbf{w}}_{1:M-1}^H \mathbf{y}_{1:M-1}^s[k]\right|^2\}}_{\varepsilon_s^2},} \tag{4.26}$$

leading to

$$\boxed{\begin{aligned} \bar{\mathbf{w}}_{1:M-1} \quad = \quad & \left(\frac{1}{\mu}\mathcal{E}\{\mathbf{y}_{1:M-1}^s \mathbf{y}_{1:M-1}^{s,H}\} + \mathcal{E}\{\mathbf{y}_{1:M-1}^n \mathbf{y}_{1:M-1}^{n,H}\}\right)^{-1} \\ & \cdot \mathcal{E}\{\mathbf{y}_{1:M-1}^n[k] y_0^{n,*}[k-\Delta]\}, \end{aligned}} \tag{4.27}$$

where ε_s^2 is the speech distortion energy and ε_n^2 the residual noise energy.

Compared to the optimization criterion (2.77) of the ANC $\bar{\mathbf{w}}_a$ in the GSC (cf. Section 2.3.1), a regularization term

$$\frac{1}{\mu}\mathcal{E}\{\left|\bar{\mathbf{w}}_{1:M-1}^H[k]\mathbf{y}_{1:M-1}^s[k]\right|^2\} \tag{4.28}$$

has been added. This regularization term limits the amount of speech distortion that is caused by the filter $\bar{\mathbf{w}}_{1:M-1}[k]$ when speech leaks into the noise references, i.e., $y_l^s[k] \neq 0$, $l = 1, ..., M-1$. In the sequel, we therefore refer to the SP-SDW-MWF without filter $\bar{\mathbf{w}}_0[k]$ as the *Speech Distortion Regularized GSC (SDR-GSC)*. The smaller μ, the smaller the resulting amount of speech distortion will be. For $\boxed{\mu = 0}$, all emphasis is put on speech distortion and so $\bar{\mathbf{w}}_{1:M-1} = \mathbf{0}$ and $z[k]$ is equal to the output of the fixed beamformer $\mathbf{a}$ delayed by Δ samples. For $\boxed{\mu = \infty}$, all emphasis is put on noise reduction and speech

distortion is not taken into account, which corresponds to the GSC-solution. Hence, the SDR-GSC encompasses the GSC as a special case.

The regularization term (4.28) with $\frac{1}{\mu} \neq 0$ adds robustness to the GSC, while not affecting the noise reduction performance in the absence of speech leakage:

- In the *absence of speech leakage, i.e.,* $y_l^s[k] = 0$, $l = 1, ..., M-1$, the regularization term (4.28) equals 0 for all μ and for all $\bar{\mathbf{w}}_{1:M-1}[k]$ and hence the residual noise energy ε_n^2 is effectively minimized. In other words, in the absence of speech leakage, the GSC solution is obtained.

- In the *presence of speech leakage*, i.e., $y_l^s[k] \neq 0$, $l = 1, ..., M-1$, speech distortion is explicitly taken into account in the optimization criterion (4.26) for the adaptive filter $\bar{\mathbf{w}}_{1:M-1}[k]$, limiting speech distortion while reducing noise. The larger the amount of speech leakage, the more attention is paid to speech distortion.

To limit speech distortion alternatively, a QIC is often imposed on the adaptive filter $\bar{\mathbf{w}}_{1:M-1}$ (cf. (2.92) in Section 2.3.3). In contrast to the SDR-GSC, the QIC acts irrespective of the amount of speech leakage $\mathbf{y}_{1:M-1}^s[k]$ that is present. The constraint value ζ^2 in (2.92) has to be chosen based on the largest model errors that may occur. As a consequence, noise reduction performance is compromised even when no or very small model errors are present. Hence, the QIC is more conservative than the SDR-GSC. The experimental results in Section 4.3 confirm this.

Remark 4.4 For $L_0 = 0$, it is readily seen that the relation (4.13) does not hold, i.e., $\bar{\mathbf{w}}_{1:M-1} + \mathbf{w}_{1:M-1} \neq \mathbf{e}_{\Delta+1}$ where

$$\begin{aligned} \mathbf{w}_{1:M-1} &= \left(\mathcal{E}\{\mathbf{y}_{1:M-1}^s \mathbf{y}_{1:M-1}^{s,H}\} + \mu\mathcal{E}\{\mathbf{y}_{1:M-1}^n \mathbf{y}_{1:M-1}^{n,H}\}\right)^{-1} \\ &\quad \cdot \mathcal{E}\{\mathbf{y}_{1:M-1}^s[k] y_0^{s,*}[k-\Delta]\}. \end{aligned} \tag{4.29}$$

This is because the speech component $\mathbf{y}_{1:M-1}^s[k]$ in the input data vector $\mathbf{y}_{1:M-1}[k]$ to the adaptive filter $\mathbf{w}_{1:M-1}$ does not contain the speech signal $y_0^s[k-\Delta]$ that is being estimated. □

4.2.3 SP-SDW-MWF with filter $\bar{\mathbf{w}}_0$ on the speech reference

Since the SDW-MWF (4.22)-(4.23) takes speech distortion explicitly into account in its optimization criterion, a filter $\bar{\mathbf{w}}_0$ on the speech reference $y_0[k]$ can be added (cf. Figure 4.2). The SDW-MWF then equals (4.23). Again, μ trades

off speech distortion and noise reduction. For $\boxed{\mu = \infty}$, speech distortion ϵ_d^2 is completely ignored so that the solution becomes

$$\bar{\mathbf{w}}_{0:M-1} = \begin{bmatrix} \bar{\mathbf{w}}_0 \\ \bar{\mathbf{w}}_{1:M-1} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{\Delta+1} \\ \mathbf{0} \end{bmatrix}, \tag{4.30}$$

which results in a trivial zero output signal $z[k]$. For $\boxed{\mu = 1}$, we obtain an MWF.

In addition, we can make the following statements.

Lemma 4.1

- *In the* absence of speech leakage*, i.e., $y_l^s[k] = 0$ for $l = 1, ..., M-1$, and for infinitely long filters $\bar{\mathbf{w}}_l$, $l = 0, ..., M-1$, the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ corresponds to a* cascade of an SDR-GSC and an SDW Single-channel Wiener (SDW-SWF) post-processor *[56, 162, 220].*
- *In the* presence of speech leakage, *the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ tries to preserve its performance: the SP-SDW-MWF then contains extra filtering operations that compensate for the performance degradation of the SDR-GSC with SDW-SWF due to speech leakage (see Figure 4.3).*

Proof : In the case of infinite filter lengths, the SP-SDW-MWF and its optimization criterion can be represented in the frequency domain[4]. For simplicity, but without loss of generality, we assume $\Delta = 0$.

Let $\bar{\mathbf{W}}_{0:M-1}(f)$ be defined as[5]

$$\bar{\mathbf{W}}_{0:M-1}(f) \;=\; \begin{bmatrix} \bar{W}_0^*(f) & \bar{\mathbf{W}}_{1:M-1}^H(f) \end{bmatrix}^H. \tag{4.31}$$

Then, the frequency-domain SP-SDW-MWF $\bar{\mathbf{W}}_{0:M-1}(f)$ equals:

$$\begin{aligned} \bar{\mathbf{W}}_{0:M-1}(f) = \arg \min_{\bar{\mathbf{W}}_{0:M-1}} \; & \mathcal{E}\left\{ \left| \begin{bmatrix} \left(1 - \bar{W}_0^*\right) & -\bar{\mathbf{W}}_{1:M-1}^H \end{bmatrix} \begin{bmatrix} Y_0^n \\ \mathbf{Y}_{1:M-1}^n \end{bmatrix} \right|^2 \right\} \\ & + \frac{1}{\mu} \mathcal{E}\left\{ \left| \begin{bmatrix} \bar{W}_0^* & \bar{\mathbf{W}}_{1:M-1}^H \end{bmatrix} \begin{bmatrix} Y_0^s \\ \mathbf{Y}_{1:M-1}^s \end{bmatrix} \right|^2 \right\} \end{aligned} \tag{4.32}$$

Decompose $\bar{\mathbf{W}}_{1:M-1}(f)$ as

$$\bar{\mathbf{W}}_{1:M-1}(f) = \left(1 - \bar{W}_0(f)\right) \bar{\mathbf{W}}_d(f) \tag{4.33}$$

[4]The frequency parameter f is often omitted in the sequel for the sake of conciseness.

[5]To simplify the expression for $\bar{W}_0(f)$, the filter $\bar{W}_0^*(f)$ instead of $\bar{W}_0(f)$ is defined here as the DTFT of $\bar{\mathbf{w}}_0^*[k]$.

with $\bar{W}_0(f)$ the single-channel filter applied to the speech reference and $\bar{\mathbf{W}}_d(f)$ a multi-channel filter and define an intermediate output $V(f)$ (see also Figure 4.3) as

$$V(f) = Y_0(f) - \bar{\mathbf{W}}_d^H(f)\mathbf{Y}_{1:M-1}(f). \tag{4.34}$$

Then, the cost function that is minimized by $\bar{\mathbf{W}}_{0:M-1}(f)$ in (4.32) can be re-written as

$$J(\bar{W}_0, \bar{\mathbf{W}}_d) = \mathcal{E}\left\{\left|\left(1-\bar{W}_0^*\right)V^n\right|^2\right\} + \frac{1}{\mu}\mathcal{E}\left\{\left|\bar{W}_0^* V^s + \bar{\mathbf{W}}_d^H \mathbf{Y}_{1:M-1}^s\right|^2\right\}. \tag{4.35}$$

From $\frac{\partial}{\partial \bar{W}_0} J(\bar{W}_0, \bar{\mathbf{W}}_d) = 0$, we find

$$\begin{aligned}\bar{W}_0(f) &= \underbrace{\left(\mathcal{E}\{V^n V^{n,*}\} + \frac{1}{\mu}\mathcal{E}\{V^s V^{s,*}\}\right)^{-1}\mathcal{E}\{V^n V^{n,*}\}}_{\bar{W}_{0,1}(f)} \\ &\quad + \underbrace{\left(\mu\mathcal{E}\{V^n V^{n,*}\} + \mathcal{E}\{V^s V^{s,*}\}\right)^{-1}\left(-\mathcal{E}\{V^s \mathbf{Y}_{1:M-1}^{s,H}\bar{\mathbf{W}}_d\}\right)}_{\bar{W}_{0,2}(f)}. \end{aligned}\tag{4.36}$$

This *single-channel filter* $\bar{W}_0(f)$ thus consists of two terms.

- The first term $\bar{W}_{0,1}(f)$ estimates the noise component $V^n(f)$ in the intermediate output $V(f)$. The filter $1-\bar{W}_{0,1}(f)$ then corresponds to an SDW-SWF that estimates the speech component $V^s(f)$ in the intermediate output $V(f)$.
- The second term $\bar{W}_{0,2}(f)$ estimates the speech leakage filtered by $-\bar{\mathbf{W}}_d(f)$, i.e., $-\bar{\mathbf{W}}_d^H\mathbf{Y}_{1:M-1}^s$. The speech component in the intermediate output $V(f)$ equals $V^s(f) = Y_0^s - \bar{\mathbf{W}}_d^H\mathbf{Y}_{1:M-1}^s$. The filter $\bar{W}_{0,2}(f)$ thus tries to compensate for the distortion $-\bar{\mathbf{W}}_d^H\mathbf{Y}_{1:M-1}^s$ by adding an estimate of $\bar{\mathbf{W}}_d^H\mathbf{Y}_{1:M-1}^s$ to the output of the SDW-SWF. In the absence of speech leakage (i.e., $\mathbf{Y}_{1:M-1}^s = \mathbf{0}$), the filter vanishes, i.e., $\bar{W}_{0,2}(f) = 0$.

From $\frac{\partial}{\partial \bar{\mathbf{W}}_d} J(\bar{W}_0, \bar{\mathbf{W}}_d) = 0$, we obtain the following solution for $\bar{\mathbf{W}}_d(f)$:

$$\begin{aligned}\bar{\mathbf{W}}_d &= \underbrace{\left(\mathcal{E}\{\mathbf{Y}_{1:M-1}^n \mathbf{Y}_{1:M-1}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}_{1:M-1}^s \mathbf{Y}_{1:M-1}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{Y}_{1:M-1}^n Y_0^{n,*}\}}_{\bar{\mathbf{W}}_{d,1}(f)} \\ &\quad \underbrace{-\left(\mu\mathcal{E}\{\mathbf{Y}_{1:M-1}^n \mathbf{Y}_{1:M-1}^{n,H}\} + \mathcal{E}\{\mathbf{Y}_{1:M-1}^s \mathbf{Y}_{1:M-1}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{Y}_{1:M-1}^s Y_0^{s,*}\frac{\bar{W}_0}{1-\bar{W}_0}\}}_{\bar{\mathbf{W}}_{d,2}(f)}. \end{aligned}\tag{4.37}$$

This *multi-channel filter* $\bar{\mathbf{W}}_d(f)$ consists of two terms.

- The first term $\bar{\mathbf{W}}_{d,1}(f)$ corresponds to the SDR-GSC (cf. (4.27)) and estimates the noise component $Y_0^n(f)$ at the output of the fixed beamformer $\mathbf{A}(f)$.
- The second term $\bar{\mathbf{W}}_{d,2}(f)$ tries to compensate for the speech distortion $-\bar{W}_0^*(f)Y_0^s(f)$ caused by $\bar{W}_0(f)$ by adding an estimate of $\frac{\bar{W}_0^*(f)}{1-\bar{W}_0^*(f)}Y_0^s(f)$ to the output of the SDR-GSC. Note that this corresponds to adding an estimate of $\bar{W}_0^*(f)Y_0^s(f)$ to the output $Z(f)$ of the SP-SDW-MWF. In the absence of speech leakage (i.e., $\mathbf{Y}_{1:M-1}^s = \mathbf{0}$), the filter vanishes, i.e., $\bar{\mathbf{W}}_{d,2}(f) = \mathbf{0}$.

Figure 4.3 graphically illustrates the solution for $\bar{\mathbf{W}}_d(f)$ and $\bar{W}_0(f)$ for $\Delta = 0$. In the *absence of speech leakage*, the filters $\bar{W}_{0,2}(f)$ and $\bar{\mathbf{W}}_{d,2}(f)$ equal zero, hence, the SP-SDW-MWF corresponds to an SDR-GSC (or GSC) cascaded with an SDW-SWF post-processor. This statement can also be proven in the time-domain, assuming sufficiently long filters $\bar{\mathbf{w}}_{0:M-1}$ (see Lemma 4.A.1 in Appendix 4.A). The SDW-SWF increases noise suppression at the expense of speech distortion, especially for large μ and at frequencies where the SNR of the intermediate output $V(f)$ is low. In the *presence of speech leakage,* the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ tries to preserve its performance: the SP-SDW-MWF then contains extra filtering operations (i.e., $\bar{W}_{0,2}(f)$ and $\bar{\mathbf{W}}_{d,2}(f)$) that compensate for the performance degradation of the SDR-GSC with SDW-SWF due to speech leakage. □

Lemma 4.2 demonstrates that the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ corresponds to a GSC with an SDW-SWF post-processor if there is no speech leakage. For $\mu = 1$ and *in the absence of speech leakage*, the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ theoretically equals the cascade of the GSC with a single-channel Wiener post-filter [56, 162, 220] (cf. Section 2.4). In practice, the performance of both schemes in the absence of speech leakage may be different because of differences in implementation. The (SDW-)MWF is typically computed using estimates of the *long-term* correlation matrix of the speech and the noise (cf. Section 2.4.2), whereas the single-channel post-filter [56, 162, 220] exploits estimates of the *short-term* PSD of the speech and the noise. The computation of the single-channel post-filter [56, 162, 220, 282, 283] is often based on the assumption of spatially uncorrelated noise, resulting in suboptimal performance for spatially correlated noise fields. The SDW-MWF on the other hand does not make any assumptions about the spatial coherence of the noise field.

In the presence of speech leakage, the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ theoretically outperforms the cascade of the GSC with a single-channel Wiener post-filter. In contrast to the GSC with single-channel post-filter, the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ contains extra filtering operations that compensate for the performance degradation due to speech leakage.

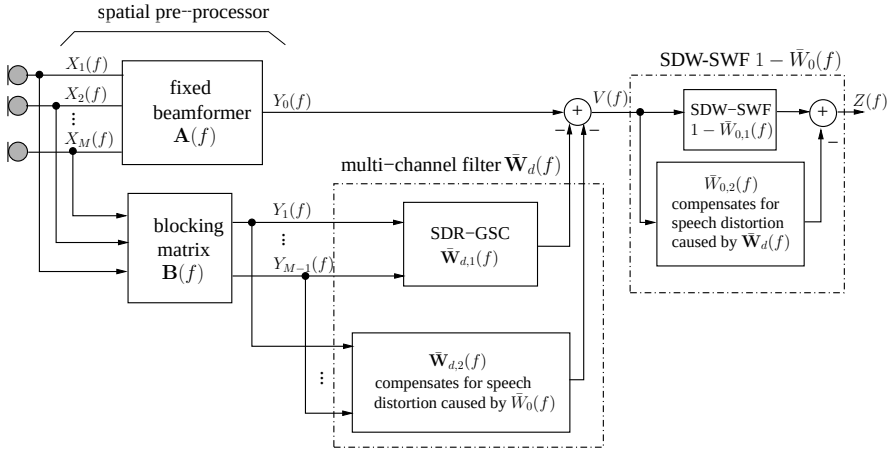


Figure 4.3: Decomposition of SP-SDW-MWF with $\bar{W}_0(f)$ in a multi-channel filter $\bar{\mathbf{W}}_d(f)$ and a single-channel post-processor $1-\bar{W}_0(f)$.

Lemma 4.2 *For infinite filter lengths, the performance of the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ is not affected by microphone mismatch as long as the desired speech component $y_0^s[k]$ at the output of the fixed beamformer $\mathbf{a}$ remains unaltered.*

Proof: In the case of infinite filter lengths, the SDW-MWF filter $\bar{\mathbf{w}}_{0:M-1}$ in (4.23) can be represented in the frequency domain as[6]

$$\bar{\mathbf{W}}(f) = \left(\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{Y}^n Y_0^{n,*}\mathrm{e}^{j2\pi f\Delta}\}. \tag{4.38}$$

Suppose that the received microphone signals $\tilde{X}_l(f)$, with $l = 1, ..., M$ are perturbed by $\Upsilon_l(f) = \upsilon_l(f)\mathrm{e}^{j\phi_l(f)}$ with $\upsilon_l(f)$ a gain deviation and $\phi_l(f)$ a phase deviation, with respect to the signals $X_l(f)$ received in the absence of mismatch, i.e.,

$$\tilde{\mathbf{X}}(f) = \mathbf{D}(f)\mathbf{X}(f), \tag{4.39}$$

with

$$\tilde{\mathbf{X}}(f) = [\tilde{X}_1(f)\ \cdots\ \tilde{X}_M(f)]^T \tag{4.40}$$

$$\mathbf{D}(f) = \mathrm{diag}\{\Upsilon_1(f), ..., \Upsilon_M(f)\} \tag{4.41}$$

and with $\mathrm{diag}\{\Upsilon_1(f), ..., \Upsilon_M(f)\}$ of full rank, i.e., $\Upsilon_l(f) \neq 0$, $l = 1, \cdots, M$. Then, assuming that the matrix $[\ \mathbf{A}(f)\quad \mathbf{B}(f)\ ]$ is of full rank, the input vector $\tilde{\mathbf{Y}}(f)$ to the adaptive filter $\bar{\mathbf{W}}(f)$ can be described as

$$\tilde{\mathbf{Y}}(f) = \left[\ \mathbf{A}(f)\quad \mathbf{B}(f)\ \right]^H \mathbf{D}(f)\mathbf{X}(f) = \mathbf{T}(f)\mathbf{Y}(f) \tag{4.42}$$

[6]The subscripts $0:M-1$ in $\bar{\mathbf{W}}_{0:M-1}(f)$ and $\mathbf{Y}_{0:M-1}(f)$ are omitted here for conciseness.

where $\mathbf{Y}(f)$ is the input vector in the absence of mismatch and $\mathbf{T}(f) \in \mathbb{C}^{M\times M}$ is a full rank matrix that depends on $\mathbf{A}(f)$, $\mathbf{B}(f)$ and $\Upsilon_l(f)$. Plugging (4.42) into (4.38), we obtain a modified filter

$$\tilde{\mathbf{W}}(f) = \mathbf{T}^{-H}\left(\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\}\mathbf{T}^H(:,1)\mathrm{e}^{j2\pi f\Delta}. \tag{4.43}$$

Hence, the SP-SDW-MWF compensates for the mismatch through the factor $\mathbf{T}^{-H}(f)$. If the desired speech component at the output of the fixed beamformer $\mathbf{A}(f)$ remains unaltered, i.e., if

$$\mathbf{Y}^{s,H}(f)\mathbf{T}^H(:,1) = Y_0^{s,*}(f), \tag{4.44}$$

the output $Z(f)$ of the SP-SDW-MWF is not affected by the mismatch. Indeed, using (4.43), (4.44) and the frequency-domain equivalent of (4.13), i.e.,

$$\begin{aligned}\left(\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-1}&\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\}\\ &= \mathbf{I}_M - \left(\mu\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-1}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\},\end{aligned} \tag{4.45}$$

we find

$$\begin{aligned}Z(f) &= \left(\mathrm{e}^{-j2\pi f\Delta}\mathbf{T}(1,:)\mathbf{Y}(f) - \tilde{\mathbf{W}}^H(f)\tilde{\mathbf{Y}}(f)\right)\\ &= \mathrm{e}^{-j2\pi f\Delta}\mathbf{T}(1,:)\frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\left(\mathcal{E}\{\mathbf{Y}^n\mathbf{Y}^{n,H}\} + \frac{1}{\mu}\mathcal{E}\{\mathbf{Y}^s\mathbf{Y}^{s,H}\}\right)^{-H}\mathbf{Y}(f)\\ &= \mathrm{e}^{-j2\pi f\Delta}Y_0(f) - \bar{\mathbf{W}}^H(f)\mathbf{Y}(f),\end{aligned} \tag{4.46}$$

which equals the output obtained when there is no microphone mismatch. □

It should be noted that, without $\bar{\mathbf{w}}_0$, performance degrades in the presence of speech leakage. Indeed, the SDR-GSC filter $\bar{\mathbf{w}}_{1:M-1}$ should both suppress noise and keep the speech distortion $\varepsilon_s^2 = \mathcal{E}\{|\bar{\mathbf{w}}_{1:M-1}^H\mathbf{y}_{1:M-1}^s|^2\}$ limited at those frequencies where speech leakage occurs (cf. (4.26)): for $\mathbf{y}_{1:M-1}^s[k] = \mathbf{0}$, ε_s^2 equals zero and hence $\bar{\mathbf{w}}_{1:M-1}$ can spend all efforts to noise suppression, whereas for $\mathbf{y}_{1:M-1}^s[k] \neq \mathbf{0}$ only the component $\bar{\mathbf{w}}_n = \bar{\mathbf{w}}_{1:M-1} - \bar{\mathbf{w}}_s$ with $\bar{\mathbf{w}}_n^H\bar{\mathbf{w}}_s = 0$ for all $\bar{\mathbf{w}}_s \in \mathrm{Range}\{\mathcal{E}\{\mathbf{y}_{1:M-1}^s\mathbf{y}_{1:M-1}^{s,H}\}\}$[7], can freely (i.e., without causing speech distortion) suppress noise. The component $\bar{\mathbf{w}}_s$ should be kept small to avoid speech distortion. Hence, at those frequencies where speech leakage occurs, the filter $\bar{\mathbf{w}}_{1:M-1}$ has less degrees of freedom to suppress noise than in the absence of speech leakage. In the absence of spatial aliasing, reverberation and internal

[7] $\mathcal{E}\{\mathbf{y}_{1:M-1}^s\mathbf{y}_{1:M-1}^{s,H}\}$ is rank deficient and hence, $\bar{\mathbf{w}}_n \neq \mathbf{0}$, if the speech leakage does not cover the full frequency spectrum and/or when the number of noise references exceeds one and the direct-to-reverberant ratio of the speech is high.

noise (such as sensor noise) and assuming a single desired speech source, the filter $\bar{\mathbf{w}}_n$ can still suppress $M-2$ localized noise sources coming from directions other than the direction of the desired speech source (cf. Section 3.3.3). Hence, performance degradation will especially occur when the total number of sound sources (i.e., speech and noise sources) exceeds the number of noise references ($= M-1$) and/or in the presence of reverberation. An additional filter $\bar{\mathbf{w}}_0$ on the speech reference then compensates for this performance degradation (cf. Section 4.3).

4.3 Experimental validation

This section illustrates the theoretical results of Section 4.2 through experimental results with the BTE hearing aid (cf. Section 2.1.5). Section 4.3.1 discusses some implementation issues. In Section 4.3.2 and Section 4.3.3, the impact of the different parameter settings of the SP-SDW-MWF on the performance and the sensitivity to signal model errors is evaluated. Finally, a comparison is made with the QIC-GSC in Section 4.3.4.

4.3.1 Implementation issues

The impact of the different parameter settings for μ and $\bar{\mathbf{w}}_0$ on the performance of the SP-SDW-MWF will be illustrated for a five noise source scenario. The five noise sources are positioned at angles 75°, 120°, 180°, 240°, 285° with respect to the desired source at 0°. To get an idea of the average performance based on directivity only, stationary speech and noise signals with the same, average long-term power spectral density are used. The speech and noise signals are different segments of the unmodulated noise signal on track 1 of [106]. The total duration of the input signal is 10 seconds of which 5 seconds contains noise-only and 5 seconds contain both the speech and noise signal. For evaluation purposes, the speech and noise signal have been recorded separately.

In the experiments, the microphones have been calibrated by means of anechoic recordings of a speech-weighted noise signal at 0°, measured while the BTE was mounted on the head. To investigate the effect of the different parameter settings (i.e., μ, $\bar{\mathbf{w}}_0$) on the performance, the filter coefficients have been computed using (4.23) where $\mathcal{E}\{\mathbf{y}^s_{0:M-1}\mathbf{y}^{s,H}_{0:M-1}\}$ is estimated by means of the clean speech contributions of the microphone signals. In practice, $\mathcal{E}\{\mathbf{y}^s_{0:M-1}\mathbf{y}^{s,H}_{0:M-1}\}$ is approximated as (cf. Section 4.1.3)

$$\mathcal{E}\{\mathbf{y}^s_{0:M-1}\mathbf{y}^{s,H}_{0:M-1}\} = \mathcal{E}\{\mathbf{y}_{0:M-1}\mathbf{y}^H_{0:M-1}\} - \mathcal{E}\{\mathbf{y}^n_{0:M-1}\mathbf{y}^{n,H}_{0:M-1}\}, \tag{4.47}$$

where $\mathcal{E}\{\mathbf{y}_{0:M-1}\mathbf{y}^H_{0:M-1}\}$ and $\mathcal{E}\{\mathbf{y}^n_{0:M-1}\mathbf{y}^{n,H}_{0:M-1}\}$ are estimated during periods of speech + noise and periods of noise-only, respectively. Under the assumption

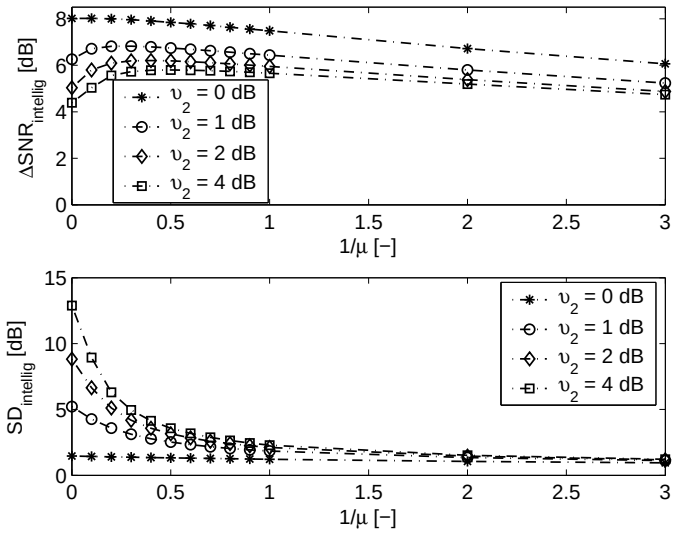


Figure 4.4: Influence of $1/\mu$ on the performance of the SDR-GSC for different gain mismatches υ_2 at the second microphone.

of a perfect speech detection, the effect of approximation (4.47) on the performance was found to be small for the given data set (i.e., differences of at most 0.5 dB in $\Delta\text{SNR}_{\text{intellig}}$). The QIC-GSC is implemented using variable loading RLS [255]. The filter length L per channel equals 96. The microphone signals are pre-whitened prior to processing to improve intelligibility (cf. Remark 2.12 of Section 2.5.1)[150], and the output is accordingly de-whitened .

To assess the sensitivity of the SDR-GSC, the SP-SDW-MWF and the QIC-GSC to errors in the assumed signal model, the influence of microphone mismatch on the performance is illustrated for a gain mismatch $\Upsilon_2 = \upsilon_2$ at the second microphone. Recall from Section 3.4 that among the different possible signal model errors, microphone gain mismatch was found to be especially harmful to the performance of the GSC when used for hearing aid applications.

The broadband performance measures, i.e., the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the distortion $\text{SD}_{\text{intellig}}$ (cf. Section 2.1.4), have been computed with respect to the output of the fixed DS beamformer $\mathbf{a}$ to exclude the effect of the spatial pre-processor.

4.3.2 SDR-GSC

Figure 4.4 plots the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the speech distortion $\text{SD}_{\text{intellig}}$ as a function of $1/\mu$ obtained by the SDR-GSC (i.e., the SP-SDW-

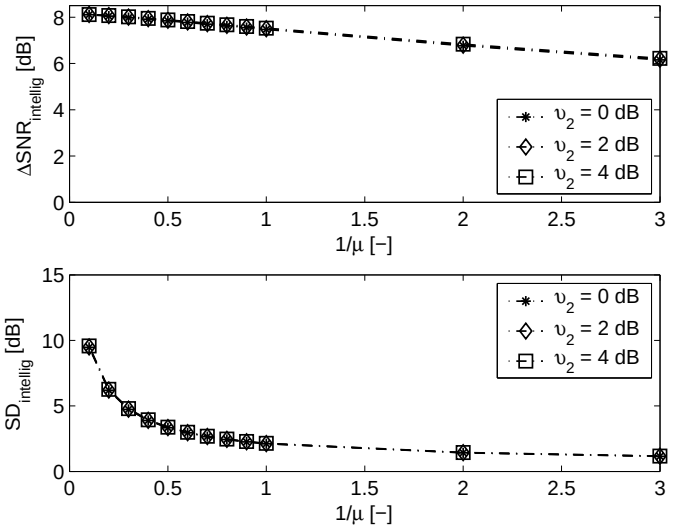


Figure 4.5: Influence of $1/\mu$ on the performance of the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ for different gain mismatches υ_2 at the second microphone.

MWF without filter $\bar{\mathbf{w}}_0$) for different gain mismatches $\Upsilon_2 = \upsilon_2$ at the second microphone.

In the absence of microphone mismatch, the amount of speech leakage into the noise references is limited. Hence, the amount of speech distortion is low for all μ. Since there is still a small amount of speech leakage due to reverberation, the amount of noise reduction and speech distortion slightly decreases for increasing $1/\mu$, especially for $1/\mu > 1$.

In the presence of microphone mismatch, the amount of speech leakage into the noise references grows. For $1/\mu = 0$ (i.e., the GSC), the speech gets significantly distorted. Due to the cancellation of the desired signal, also the improvement $\Delta\text{SNR}_{\text{intellig}}$ degrades. Setting $1/\mu > 0$, improves the performance of the GSC in the presence of model errors without compromising performance in the absence of signal model errors. For the given set-up, a value $1/\mu$ around 0.5 seems appropriate for guaranteeing good performance for a gain mismatch up to 4 dB.

4.3.3 SP-SDW-MWF with $\bar{\mathbf{w}}_0$

Figure 4.5 plots the performance measures $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the SP-SDW-MWF with filter $\bar{\mathbf{w}}_0$. In general, the amount of speech distortion and noise reduction grows for decreasing $1/\mu > 0$. For $1/\mu = 0$, all emphasis is put

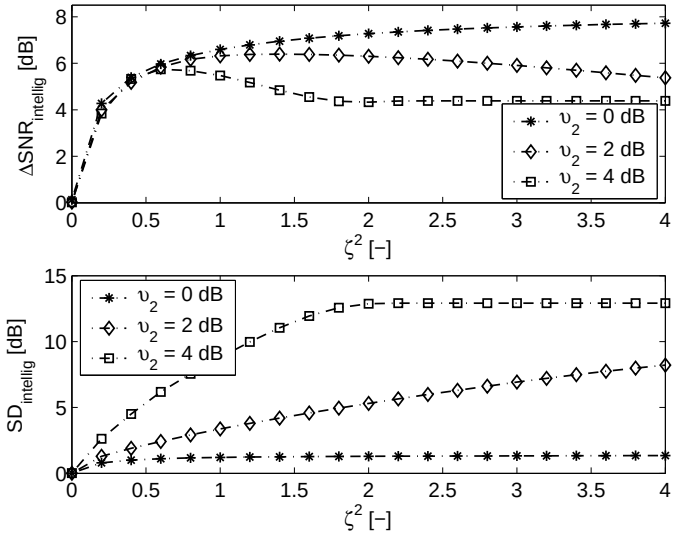


Figure 4.6: $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ for QIC-GSC as a function of ζ^2 for different gain mismatches υ_2 at the second microphone.

on noise reduction. This results in a total cancellation of the speech and the noise signal (cf. Section 4.2.3) and hence degraded performance (for $1/\mu = 0$, $\Delta\text{SNR}_{\text{intellig}} = 5.4$ dB, $\text{SD}_{\text{intellig}} = 39$ dB).

In the absence of microphone mismatch, the SDR-GSC (i.e., $L_0 = 0$) and the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ (i.e., $L_0 \neq 0$) result, except for $1/\mu = 0$, in approximately the same $\Delta\text{SNR}_{\text{intellig}}$ [8], while the distortion for the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ is higher due to the additional single-channel SDW-SWF.

In the presence of microphone mismatch, the performance of the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ does not degrade due to the microphone mismatch, in contrast to the SDR-GSC.

4.3.4 Comparison with QIC-GSC

Figure 4.6 depicts the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the speech distortion $\text{SD}_{\text{intellig}}$, respectively, of the QIC-GSC as a function of ζ^2. Like the SDR-GSC, the QIC increases the robustness of the GSC: in the presence of microphone mismatch, $\text{SD}_{\text{intellig}}$ decreases for decreasing ζ^2. The QIC is independent of the actual amount of speech leakage. As a consequence, distortion grows fast with

[8] For $L_0 \neq 0$, the fullband SNR improvement was larger thanks to the SDW-SWF post-processor (see Section 4.2.3). For other noise sources, e.g., a narrow band noise source, also a better improvement in $\text{SNR}_{\text{intellig}}$ can be achieved by $L_0 \neq 0$ thanks to the SDW-SWF.

increasing microphone mismatch. The constraint value ζ^2 should be chosen so that the maximum allowable speech distortion level is not exceeded for the largest possible model errors (e.g., 5 dB $\text{SD}_{\text{intellig}}$ for a gain mismatch up to 4 dB). This goes at the expense of reduced noise reduction for small model errors (e.g., $\Delta\text{SNR}_{\text{intellig}} = 5.4$ dB for $\zeta^2 = 0.4$). The SDR-GSC on the other hand, keeps the speech distortion limited for all model errors (see Figure 4.4). Emphasis on speech distortion is increased if the amount of speech leakage grows. As a result, a better noise reduction performance is obtained for small model errors (e.g., between 5.7 (i.e., for $\upsilon_2 = 4$ dB) and 8 dB (i.e., for $\upsilon_2 = 0$ dB) for $1/\mu$ between 0.3 and 0.4), while guaranteeing sufficient robustness for large model errors. In addition, Figure 4.5 demonstrates that an additional filter $\bar{\mathbf{w}}_0$ in the SP-SDW-MWF further improves the performance in the presence of microphone mismatch (e.g., for $1/\mu$ between 0.3 and 0.4, the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ achieves an improvement $\Delta\text{SNR}_{\text{intellig}}$ of 8 dB, independent of the amount of gain mismatch υ_2, while the SDR-GSC obtains an improvement $\Delta\text{SNR}_{\text{intellig}}$ of only 5.7 dB for $\upsilon_2 = 4$ dB (cf. Figure 4.4).

4.4 Conclusion

In this chapter, we have shown that the GSC and the MWF can be integrated into one generalized noise reduction scheme, referred to as *spatially pre-processed speech distortion weighted MWF (SP-SDW-MWF),* which consists of a fixed, spatial pre-processor and an adaptive speech distortion weighted MWF. The new scheme encompasses the GSC and the MWF as extreme cases. In addition, it allows for in-between solutions, such as a speech distortion regularized GSC (SDR-GSC). Depending on the setting of a trade-off *parameter* μ and *the presence or absence of the filter* $\bar{\mathbf{w}}_0$ *on the speech reference*, the GSC, the SDR-GSC or an (SDW-)MWF is obtained.

In Section 4.2, the different parameter settings of the SP-SDW-MWF have been interpreted.

- Without $\bar{\mathbf{w}}_0$, the SP-SDW-MWF corresponds to an SDR-GSC: the ANC design criterion is supplemented with a regularization term that limits the speech distortion due to signal model errors. The larger $\frac{1}{\mu}$, the smaller the amount of distortion. For $\frac{1}{\mu} = 0$, distortion is ignored completely, which corresponds to the GSC-solution. The SDR-GSC is then an alternative technique to the QIC-GSC to decrease the sensitivity of the GSC to signal model errors. In contrast to the QIC-GSC, the SDR-GSC shifts emphasis towards speech distortion when the amount of speech leakage grows. As a result, a better noise reduction performance is obtained for small model errors, while guaranteeing robustness against large model errors.

- Since the SP-SDW-MWF takes speech distortion explicitly into account, a filter $\bar{\mathbf{w}}_0$ on the speech reference can be added. It has been shown that in the absence of speech leakage and for infinitely long filter lengths, the SP-SDW-MWF corresponds to a cascade of an SDR-GSC with an SDW-SWF post-processor. In the presence of speech leakage, the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ tries to preserve its performance: compared to an SDR-GSC with SDW-SWF post-processor, the SP-SDW-MWF then contains extra filtering operations that compensate for the performance degradation due to speech leakage. In contrast to the SDR-GSC (and thus also the GSC), performance does not degrade due to microphone mismatch.

In Section 4.3, experimental results with recordings from a BTE hearing aid have illustrated the theoretical results of Section 4.2. The SP-SDW-MWF indeed increases the robustness of the GSC against signal model errors. Comparison with the QIC-GSC demonstrated that the SP-SDW-MWF achieves a better noise reduction performance for a given maximum allowable speech distortion level.

4.A Appendix: SP-SDW-MWF with filter $\bar{\mathbf{w}}_0[k]$

Lemma 4.A.1 *In the absence of speech leakage, i.e.* $y_l^s[k] = 0$, $l = 1, \cdots, M-1$ *and for sufficiently long filter lengths* L, *the SP-SDW-MWF with* $\bar{\mathbf{w}}_0[k] \neq \mathbf{0}$ *corresponds to the cascade of a GSC and a single-channel SDW-SWF post-processor.*

Proof: If $\mathbf{y}_{1:M-1}^s[k] = \mathbf{0}$, the cost function (4.22) of the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ becomes

$$J(\bar{\mathbf{w}}_{0:M-1}) = \underbrace{\mathcal{E}\left\{ \left| \begin{bmatrix} (\mathbf{e}_{\Delta+1} - \bar{\mathbf{w}}_0)^H & -\bar{\mathbf{w}}_{1:M-1}^H \end{bmatrix} \begin{bmatrix} \mathbf{y}_0^n \\ \mathbf{y}_{1:M-1}^n \end{bmatrix} \right|^2 \right\}}_{\varepsilon_n^2} + \frac{1}{\mu} \underbrace{\mathcal{E}\left\{ \left| \bar{\mathbf{w}}_0^H \mathbf{y}_0^s \right|^2 \right\}}_{\varepsilon_s^2}. \tag{4.A.1}$$

Without loss of generality, we assume that $\Delta = \left\lceil \frac{L}{2} \right\rceil - 1$ where L is the filter length per channel. Assume that the filter length L is sufficiently long such that $\bar{\mathbf{w}}_{0:M-1}[k]$ can be decomposed as depicted in **Figure 4.A.7**, i.e.,

$$\bar{\mathbf{w}}_0[k] = \bar{\mathbf{w}}_b[k] \otimes \mathbf{e}_{\Delta_1+1}, \tag{4.A.2}$$

$$\bar{\mathbf{w}}_l[k] = (\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b[k]) \otimes \bar{\mathbf{w}}_{dl}[k], \quad l = 1, ..., M-1 \tag{4.A.3}$$

where $\bar{\mathbf{w}}_b[k] \in \mathbb{C}^{L_b \times 1}$ is a single-channel filter with length L_b,

$$\bar{\mathbf{w}}_d[k] = \begin{bmatrix} \bar{\mathbf{w}}_{d1}^T[k] & \bar{\mathbf{w}}_{d2}^T[k] & \cdots & \bar{\mathbf{w}}_{dM-1}^T[k] \end{bmatrix}^T \in \mathbb{C}^{L_d(M-1)\times 1} \tag{4.A.4}$$

is a multi-channel filter with filter length L_d per channel and $\Delta_1 + \Delta_2 = \Delta$. Define the intermediate output $v[k]$ as the output of the multi-channel filter $\begin{bmatrix} \mathbf{e}_{\Delta_1+1}^H & -\bar{\mathbf{w}}_d^H[k] \end{bmatrix}^H$, i.e.,

$$v[k] = y_0[k - \Delta_1] - \sum_{l=1}^{M-1} \bar{\mathbf{w}}_{dl}^H[k] \mathbf{y}_l[k]. \tag{4.A.5}$$

Then, using $\mathbf{y}_{1:M-1}^s[k] = 0$, the cost function $J(\bar{\mathbf{w}}_{0:M-1})$ can be re-written as

$$J(\bar{\mathbf{w}}_b, \bar{\mathbf{w}}_d) = \mathcal{E}\left\{ \left| (\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b)^H \mathbf{v}^n[k] \right|^2 \right\} + \frac{1}{\mu} \mathcal{E}\left\{ \left| \bar{\mathbf{w}}_b^H \mathbf{v}^s[k] \right|^2 \right\} \tag{4.A.6}$$

where

$$\mathbf{v}^{n,H}[k] = \begin{bmatrix} v^n[k] & v^n[k-1] & \cdots & v^n[k-L_b+1] \end{bmatrix}^H, \tag{4.A.7}$$

$$\mathbf{v}^{s,H}[k] = \begin{bmatrix} v^s[k] & v^s[k-1] & \cdots & v^s[k-L_b+1] \end{bmatrix}^H. \tag{4.A.8}$$

From $\frac{\partial}{\partial \bar{\mathbf{w}}_b} J(\bar{\mathbf{w}}_b, \bar{\mathbf{w}}_d) = 0$, we find

$$\bar{\mathbf{w}}_b[k] = \left(\mathcal{E}\{\mathbf{v}^n \mathbf{v}^{n,H}\} + \frac{1}{\mu} \mathcal{E}\{\mathbf{v}^s \mathbf{v}^{s,H}\} \right)^{-1} \mathcal{E}\{\mathbf{v}^n[k] v^{n,*}[k - \Delta_2]\}. \tag{4.A.9}$$

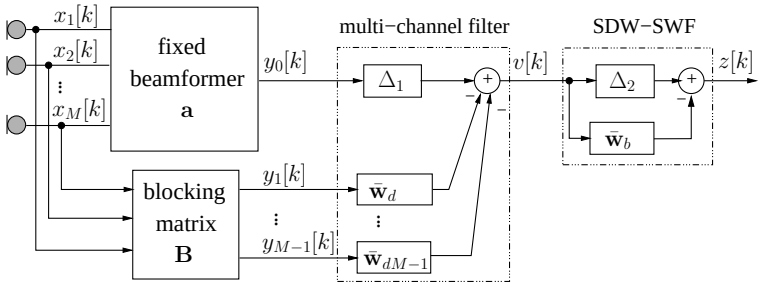


Figure 4.A.7: Decomposition of the SP-SDW-MWF $\bar{\mathbf{w}}_{0:M-1}[k]$ in a multi-channel filter $\bar{\mathbf{w}}_d[k]$ and a single-channel post-processor $\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b[k]$.

Hence, the filter $(\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b[k])$ corresponds to a single channel SDW-SWF that estimates the (delayed) speech component $v^s[k-\Delta_2]$ in the output $v[k]$ of the multi-channel filter $\left[\begin{array}{cc} \mathbf{e}_{\Delta_1+1}^H & -\bar{\mathbf{w}}_d^H[k] \end{array}\right]^H$.
The solution for $\bar{\mathbf{w}}_d$ can be found by setting the derivative $\frac{\partial}{\partial \bar{\mathbf{w}}_d}\varepsilon_n^2$ (cf. (4.A.1)) to zero[9]. Re-writing ε_n^2 as

$$\varepsilon_n^2 = \mathcal{E}\left\{ \left| \left[\begin{array}{cc} \mathbf{e}_{\Delta_1+1}^H & -\bar{\mathbf{w}}_d^H[k] \end{array}\right] \mathbf{y}_{0:M-1,p}^n[k] \right|^2 \right\}, \quad (4.A.10)$$

$$\mathbf{y}_{0:M-1,p}^n[k] = \left[\begin{array}{cccc} \mathbf{y}_{0,p}^{n,H}[k] & \mathbf{y}_{1,p}^{n,H}[k] & \cdots & \mathbf{y}_{M-1,p}^{n,H}[k] \end{array}\right]^H, \quad (4.A.11)$$

with

$$\mathbf{y}_{l,p}^n = \left[\begin{array}{cccc} \boxed{(\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b)^H} & 0 & \cdots & 0 \\ 0 & \boxed{(\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b)^H} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \boxed{(\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b)^H} \end{array}\right] \left[\begin{array}{c} y_l[k] \\ y_l[k-1] \\ \vdots \\ y_l[k-L+1] \end{array}\right], \quad (4.A.12)$$

we find

$$\bar{\mathbf{w}}_d[k] = \mathcal{E}\{\mathbf{y}_{0:M-1,p}^n[k]\mathbf{y}_{0:M-1,p}^{n,H}[k]\}^{-1}\mathcal{E}\{\mathbf{y}_{0:M-1,p}^n[k]y_{0,p}^n[k-\Delta_1]\}. \quad (4.A.13)$$

The filter coefficients $\bar{\mathbf{w}}_d$ correspond to the ANC coefficients of a GSC where the speech and noise references have been pre-filtered with the single -channel SDW-SWF filter $\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b$. For infinite filter lengths, $\bar{\mathbf{w}}_d$ then equals the ANC operating on the original speech and noise references.

In short, in the absence of speech leakage, the SP-SDW-MWF equals a GSC with a single channel SDW-SWF post-processor $\mathbf{e}_{\Delta_2+1} - \bar{\mathbf{w}}_b$. □

[9]Note that for $\mathbf{y}^s = \mathbf{0}$, ε_s^2 in (4.A.1) does not depend on $\bar{\mathbf{w}}_d[k]$.

Chapter 5

Stochastic gradient implementation of the SP-SDW-MWF

IN the previous chapter, we have proposed a generalized noise reduction scheme, called the spatially pre-processed speech distortion weighted MWF (SP-SDW-MWF), that adds robustness to the GSC by taking speech distortion explicitly into account in the design criterion of the adaptive stage. Compared to the widely studied GSC with quadratic inequality constraint (QIC-GSC), the SP-SDW-MWF achieves a better noise reduction performance, for a given maximum speech distortion level. However, in contrast to the GSC and the QIC-GSC (cf. Section 2.3.3), no cheap stochastic gradient based implementation of the SP-SDW-MWF is available yet. Although a subband QRD implementation of the SP-SDW-MWF already has a low computational complexity, the limited processing power in the current digital hearing aids as well as the ongoing miniaturization in the hearing aid industry motivate the development of cheaper stochastic gradient algorithms. In [57], a stochastic gradient implementation has been proposed for the SDW-MWF (4.6), using a circular speech + noise and noise data buffer (cf. Section 4.1.3). This algorithm can be modified so that it applies to the SP-SDW-MWF. However, it relies on some independence assumptions that are not necessarily satisfied, so that the performance is seriously degraded (cf. Appendix 5.A).

In this chapter, we develop an alternative low-cost, stochastic gradient implementation of the SP-SDW-MWF that has a better performance. To speed up convergence and reduce computational complexity, the algorithm is implemented in the frequency-domain. Experimental results with the three-

microphone BTE show that the proposed frequency-domain stochastic gradient (FD-SG) algorithm preserves the benefit of the exact SP-SDW-MWF over the QIC-GSC, while its computational cost is comparable to the LMS based scaled projection algorithm (cf. Section 2.3.3) for the QIC-GSC.

In Section 5.1, a stochastic gradient algorithm is derived from the cost function of the SP-SDW-MWF. This stochastic gradient algorithm however requires an instantaneous estimate of the speech correlation matrix. In practice, this instantaneous estimate cannot be computed, so that approximations have to be made. In Section 5.2, a first solution is proposed that is based on instantaneous estimates of the speech + noise and the noise correlation matrix. To speed-up convergence and reduce computational complexity, the algorithm is implemented in the frequency-domain. In Section 5.3, we suggest that the performance of the proposed FD-SG algorithm in spectrally highly non-stationary noise is improved by applying a low pass filter (LPF) to the regularization term in the filter update equation that limits speech distortion. The FD-SG with LPF algorithm however requires data buffers for calculating the regularization term in the filter update equation, resulting in a large memory usage. We show that the memory usage of the FD-SG algorithm can be decreased dramatically by approximating the regularization term using (diagonal) correlation matrices, making an implementation of the SP-SDW-MWF in commercial hearing aids feasible both in terms of complexity as well as memory usage.

Section 5.4 and Section 5.5 compare the computational complexity and the performance of the proposed frequency-domain stochastic gradient implementations with the NLMS based SPA. Experiments with the three-microphone BTE demonstrate that the FD-SG algorithm with LPF outperforms the FD-SG algorithm, especially in spectrally highly non-stationary noise. In addition, the proposed FD-SG with LPF algorithm indeed preserves the performance benefit of the exact SP-SDW-MWF over the QIC-GSC, while its computational complexity is comparable to the NLMS based SPA. Preliminary results indicate that the FD-SG algorithm using correlation matrices (CM) has about the same performance as the FD-SG algorithm with LPF, making it a good alternative to the SPA for implementation in future hearing aids.

5.1 Towards a stochastic gradient algorithm

Starting from the cost function of the SP-SDW-MWF (cf. Section 4.2), a time-domain stochastic gradient algorithm is derived in Section 5.1.1. However, the resulting algorithm, which will be referred to as *optimal* TD-SG algorithm, requires an estimate of the instantaneous speech correlation matrix. In practice, this information is not available, so that approximations have to be made. Possible approximations of the optimal TD-SG algorithm are summarized in Section 5.1.2 and will be elaborated in Section 5.2 and Section 5.3.

5.1.1 Optimal stochastic gradient algorithm

Recall from Section 4.2 that the cost function $J(\bar{\mathbf{w}}_{0:M-1})$ of the SP-SDW-MWF equals

$$J(\bar{\mathbf{w}}_{0:M-1}) = \underbrace{\mathcal{E}\left\{\left|y_0^n[k-\Delta] - \bar{\mathbf{w}}_{0:M-1}^H \mathbf{y}_{0:M-1}^n[k]\right|^2\right\}}_{\varepsilon_n^2} + \frac{1}{\mu}\underbrace{\mathcal{E}\left\{\left|\bar{\mathbf{w}}_{0:M-1}^H \mathbf{y}_{0:M-1}^s[k]\right|^2\right\}}_{\varepsilon_s^2}, \quad (5.1)$$

where ε_n^2 is the residual noise energy and ε_s^2 the speech distortion energy. A stochastic gradient algorithm approximates the steepest descent algorithm

$$\bar{\mathbf{w}}_{0:M-1}[j] = \bar{\mathbf{w}}_{0:M-1}[j-1] + \frac{\rho}{2}\left(-\frac{\partial J(\bar{\mathbf{w}}_{0:M-1})}{\partial \bar{\mathbf{w}}_{0:M-1}}\right)_{\bar{\mathbf{w}}_{0:M-1}=\bar{\mathbf{w}}_{0:M-1}[j-1]}, \quad (5.2)$$

where j denotes the iteration index and ρ equals the step size, using an instantaneous gradient estimate. Replacing the iteration index j by a time index k and leaving out the expectation operators $\mathcal{E}\{.\}$ as is typically done in the derivation of a stochastic gradient algorithm, we obtain the following update equation[1] for the cost function (5.1) of the SP-SDW-MWF:

$$\boxed{\bar{\mathbf{w}}[k] = \bar{\mathbf{w}}[k-1] + \rho\left\{\mathbf{y}^n[k](y_0^{n,*}[k-\Delta] - \mathbf{y}^{n,H}[k]\bar{\mathbf{w}}[k-1]) - \mathbf{r}[k]\right\},} \quad (5.3)$$

with

$$\boxed{\mathbf{r}[k] = \frac{1}{\mu}\mathbf{y}^s[k]\mathbf{y}^{s,H}[k]\bar{\mathbf{w}}[k-1],} \quad (5.4)$$

where $\bar{\mathbf{w}}[k]$, $\mathbf{y}[k] \in \mathbb{C}^{NL\times 1}$ and N denotes the number of input channels to the adaptive filter ($N = M$ if $\bar{\mathbf{w}}_0$ is present, $N = M-1$ if $\bar{\mathbf{w}}_0$ is absent) . For $\frac{1}{\mu} = 0$ and no filter $\bar{\mathbf{w}}_0$ on the speech reference, equation (5.3) reduces to an LMS type update formula often used in GSC (cf. (2.82) in Section 2.3.1), which is then operated during *noise-only* periods (i.e., when $y_l[k] = y_l^n[k]$, $l = 0, ..., M-1$). The additional term $\mathbf{r}[k]$ in the gradient estimate limits the speech distortion due to possible signal model errors.

Equation (5.4) however requires an instantaneous estimate $\mathbf{y}^s[k]\mathbf{y}^{s,H}[k]$ of the clean speech correlation matrix. However, since the clean speech samples cannot be observed in practice, this instantaneous estimate cannot be computed. Hence, $\mathbf{r}[k]$ will have to be approximated. In Section 5.1.2, different approximations for the true regularization term (5.4) are proposed. In the sequel, the

[1] The update formula can be applied to both the SDR-GSC and the SP-SDW-MWF with $\bar{\mathbf{w}}_0$. In the sequel, the subscripts $0:M-1$ (or $1:M-1$) in the adaptive filter $\bar{\mathbf{w}}_{0:M-1}$ (or $\bar{\mathbf{w}}_{1:M-1}$) and the input vector $\mathbf{y}_{0:M-1}$ (or $\mathbf{y}_{1:M-1}$) are omitted for the sake of conciseness.

TD-SG algorithm (5.3) with the unknown regularization term (5.4) is called *optimal* TD-SG algorithm. In Section 5.5, the *optimal* TD-SG algorithm will be used as a reference to evaluate the performance of the different approximations.

5.1.2 Approximations of the optimal SG algorithm

This section summarizes different approximations for the true regularization term (5.4). The different solutions are discussed more in detail in Section 5.2 and Section 5.3.

TD-SG algorithm

A first solution, which is referred to as the TD-SG algorithm, is to *approximate the instantaneous speech correlation matrix estimate* based on estimates of the instantaneous speech + noise and noise correlation matrices. The instantaneous estimates of the speech + noise and noise correlation matrices are measured at different moments in time and are computed using data buffers. However, because of the non-continuous estimation of the speech + noise and the noise correlation matrix, the estimate of (5.4) will be poor, especially for highly non-stationary noise such as multi-talker babble noise. As a result, the TD-SG algorithm suffers from a large residual excess error (cf. Section 5.2.1).

TD-SG algorithm using correlation matrices

In practical scenarios, the speech and noise signals are often spectrally highly non-stationary, while their long-term spectral and spatial characteristics usually vary more slowly in time. If the long-term characteristics of the speech are slowly time-varying, non-stationary noise can still be effectively suppressed by using the *long-term speech correlation matrix* in the regularization term, i.e., using

$$\boxed{\mathbf{r}[k] = \frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s[k]\mathbf{y}^{s,H}[k]\}\bar{\mathbf{w}}[k-1],} \tag{5.5}$$

in (5.3). Note that the algorithm (5.3) with (5.5) is not a true stochastic gradient algorithm anymore, since the expectation operator in the term ε_s^2 of the cost function (5.1) has been preserved.

Assuming that the *long-term* average correlation matrix of the noise varies slowly in time, a reliable estimate of $\mathcal{E}\{\mathbf{y}^s[k]\mathbf{y}^{s,H}[k]\}$ can still be obtained by long-term averaging (e.g., during one second) of the measured speech + noise and the noise correlation matrices (cf. Section 3.5). However, this would require expensive matrix operations. In Section 5.2.2 and Section 5.3, three solutions are proposed to avoid expensive matrix operations, namely, a frequency-domain implementation of the stochastic gradient (FD-SG) algorithm, a TD-SG or FD-SG algorithm with low pass filtering of the regularization term (TD-SG and

FD-SG algorithm with LPF) and a FD-SG algorithm using correlation matrices (FD-SG algorithm using CM).

FD-SG algorithm

A frequency-domain implementation of the TD-SG (cf. Section 5.2.1) intrinsically averages the measured correlation matrices over L samples, where L is the filter length per channel. In addition, compared to its time-domain equivalent, a reduced computational complexity and a faster convergence is achieved. For spectrally stationary noise, the limited averaging over L samples by the FD-SG algorithm suffices to produce a reliable estimate of the speech correlation matrix (5.5). However, for spectrally highly non-stationary noise, a long-term averaging (i.e., over $K \gg NL$ samples) is required in order to obtain a good performance (cf. Section 5.5).

TD-SG and FD-SG algorithm with low pass filtering (LPF)

In Section 5.3, two cheap approximations are proposed that allow to incorporate a long-term averaging. A first solution, called TD-SG with LPF or FD-SG with LPF, is to low pass filter the regularization term in the TD-SG or FD-SG algorithm that takes speech distortion into account. Experiments in Section 5.5 show that the FD-SG algorithm with LPF outperforms the FD-SG algorithm, especially in spectrally highly non-stationary noise. The FD-SG (and TD-SG) algorithm with LPF however require data buffers for computing the regularization term, resulting in a large memory usage.

FD-SG algorithm using correlation matrices (CM)

In the case of a frequency-domain implementation, an alternative solution to incorporate a long-term averaging is proposed, which leads to a significant memory reduction. In this alternative FD-SG algorithm, the regularization term is approximated using (diagonal) frequency-domain correlation matrices instead of data buffers. In the sequel, this technique is referred to as FD-SG algorithm using CM.

5.2 Stochastic gradient algorithm

In this section, we propose a first approximation of the optimal TD-SG algorithm, which is referred to as TD-SG algorithm. The TD-SG algorithm approximates the instantaneous speech correlation matrix estimate by instantaneous estimates of the speech + noise and the noise correlation matrix. To speed-up convergence and reduce computational complexity, the TD-SG algorithm is implemented in the frequency-domain, as described in Section 5.2.2.

5.2.1 Time-domain stochastic gradient algorithm

In the TD-SG algorithm, NL-dimensional *speech + noise* signal vectors $\mathbf{y}[k]$ are stored in a circular buffer $\mathbf{B}_y \in \mathbb{C}^{NL \times L_{B_y}}$ with L_{B_y} the buffer length per channel, during processing as in [57]. *During periods of noise-only* (i.e., when $y_l[k] = y_l^n[k]$, $l = 0, ..., M-1$), the filter $\bar{\mathbf{w}}[k]$ is updated with (5.3) using the following approximation of the term $\mathbf{r}[k]$:

$$\boxed{\mathbf{r}[k] = \frac{1}{\mu}\left(\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^H[k] - \mathbf{y}^n[k]\mathbf{y}^{n,H}[k]\right)\bar{\mathbf{w}}[k-1],} \tag{5.6}$$

where $\mathbf{y}_{B_y}[k]$ is an NL-dimensional speech + noise vector from the buffer $\mathbf{B}_y$. In the sequel, we refer to (5.3), (5.6) as the TD-SG algorithm.

Additional storage of L-dimensional noise-only vectors $\mathbf{y}_l^n[k]$, $l = 0, \cdots, M-1$ in a second buffer $\mathbf{B}_y^n \in \mathbb{C}^{ML \times L_{B_y^n}}$ with buffer length $L_{B_y^n}$, allows to adapt $\bar{\mathbf{w}}[k]$ also *during periods of speech + noise,* using:

$$\begin{aligned}
\bar{\mathbf{w}}[k] &= \bar{\mathbf{w}}[k-1] + \rho\left\{\mathbf{y}_{B_y^n}[k](y_{B_y^n,0}^*[k-\Delta] - \mathbf{y}_{B_y^n}^H[k]\bar{\mathbf{w}}[k-1]) - \mathbf{r}[k]\right\}, \\
\mathbf{r}[k] &= \frac{1}{\mu}\left(\mathbf{y}[k]\mathbf{y}^H[k] - \mathbf{y}_{B_y^n}[k]\mathbf{y}_{B_y^n}^H[k]\right)\bar{\mathbf{w}}[k-1].
\end{aligned} \tag{5.7}$$

Below, we will, for the sake of conciseness, only consider the update procedure during noise-only of the TD-SG algorithm to explain further refinements and modifications, hence, $\mathbf{y}[k] = \mathbf{y}^n[k]$ and $y_0[k] = y_0^n[k]$. The extension towards updating during speech + noise periods with the use of a second, noise buffer $\mathbf{B}_{B_y^n}$ is straightforward: e.g., equation (5.7) is found by replacing the noise-only input vectors $\mathbf{y}^n[k]$ by $\mathbf{y}_{B_y^n}[k]$ and the speech + noise vectors $\mathbf{y}_{B_y}[k]$ by the input speech + noise vector $\mathbf{y}[k]$ in formulas (5.3)-(5.6).

Remark 5.1 For $\mu > 1$, an alternative stochastic gradient algorithm similar to the implementation in [57] can be derived from the TD-SG algorithm by invoking the independence assumptions that have been used in [57]. Simulations, however, have shown that these independence assumptions result in a significant performance degradation with respect to the stochastic gradient algorithms developed in this chapter, while not significantly reducing the computational complexity (cf. Appendix 5.A). □

Convergence properties

Using[2]

$$\bar{\mathbf{w}}_{opt} = \left(\frac{1}{\mu}\mathcal{E}\{\mathbf{y}_{B_y}\mathbf{y}_{B_y}^H\} + (1-\frac{1}{\mu})\mathcal{E}\{\mathbf{y}^n\mathbf{y}^{n,H}\}\right)^{-1}\mathcal{E}\{\mathbf{y}^n[k]y_0^{n,*}[k-\Delta]\}, \tag{5.8}$$

[2]When the second order statistics of the noise are stationary, i.e., $\mathcal{E}\{\mathbf{y}_{B_y}^n\mathbf{y}_{B_y}^{n,H}\} = \mathcal{E}\{\mathbf{y}^n\mathbf{y}^{n,H}\}$, $\bar{\mathbf{w}}_{opt}$ equals (4.23).

and (5.3), (5.6), it can be shown that

$$\mathcal{E}\{\bar{\mathbf{w}}[k]-\bar{\mathbf{w}}_{opt}\} = \left(\mathbf{I}_{NL}-\rho\mathcal{E}\left\{\frac{1}{\mu}\mathbf{y}_{B_y}\mathbf{y}_{B_y}^H+(1-\frac{1}{\mu})\mathbf{y}^n\mathbf{y}^{n,H}\right\}\right)^k \cdot \mathcal{E}\{\bar{\mathbf{w}}[0]-\bar{\mathbf{w}}_{opt}\}. \quad (5.9)$$

Hence, the TD-SG algorithm (5.3), (5.6) is *convergent in the mean* provided that the step size ρ is smaller than $\frac{2}{\lambda_{\max}}$ with $\lambda_{\max}$ the maximum eigenvalue of $\mathcal{E}\{\frac{1}{\mu}\mathbf{y}_{B_y}\mathbf{y}_{B_y}^H+(1-\frac{1}{\mu})\mathbf{y}^n\mathbf{y}^{n,H}\}$. The similarity of (5.3) with standard LMS lets us presume that setting

$$\rho < \frac{2}{\frac{1}{\mu}L\sum_{l=M-N}^{M-1}\mathcal{E}\{|y_{B_y,l}[k]|^2\}+(1-\frac{1}{\mu})L\sum_{l=M-N}^{M-1}\mathcal{E}\{|y_l^n[k]|^2\}}, \quad (5.10)$$

guarantees *convergence in the mean square*. In the sequel, the step size ρ of (5.3)-(5.6) is therefore normalized as[3]:

$$\rho = \frac{\bar{\rho}}{\frac{1}{\mu}\left|\mathbf{y}_{B_y}^H[k]\mathbf{y}_{B_y}[k]-\mathbf{y}^{n,H}[k]\mathbf{y}^n[k]\right|+\mathbf{y}^{n,H}[k]\mathbf{y}^n[k]+c}, \quad (5.11)$$

where the step size $\bar{\rho}\in[0,2)$ and c is a very small constant. Note that, in practice, the energies $\mathbf{y}_{B_y}^H[k]\mathbf{y}_{B_y}[k]$ and $\mathbf{y}^{n,H}[k]\mathbf{y}^n[k]$ are more efficiently computed using a low pass IIR filter (cf. (2.84) of Section 2.3.1).

Residual excess error

The estimate (5.6) of the regularization term (5.4) is however quite poor due to large differences between the rank-one estimates $\mathbf{y}^n[k]\mathbf{y}^{n,H}[k]$ and $\mathbf{y}^n[k']\mathbf{y}^{n,H}[k']$ of the noise correlation matrix at different time instants k and k'. Consequently, the instantaneous gradient estimate in the TD-SG update equation (5.3)-(5.6) is, compared to the optimal TD-SG algorithm (5.3)-(5.4), additionally perturbed by

$$\frac{\rho}{\mu}\left(\mathbf{y}^n[k]\mathbf{y}^{n,H}[k]-\mathbf{y}_{B_y}^n[k]\mathbf{y}_{B_y}^{n,H}[k]\right)\bar{\mathbf{w}}[k-1], \quad (5.12)$$

for $\mu\neq\infty$. Hence, for $\mu\neq\infty$, the TD-SG algorithm, i.e., (5.3) with (5.6), suffers from a larger residual excess error than the optimal TD-SG algorithm (5.3)-(5.4). The additional excess error grows for decreasing μ, increasing step size ρ and increasing vector length NL of the vector $\mathbf{y}[k]$. It is expected to be especially large for highly non-stationary noise, e.g., multi-talker babble noise. The gradient estimate can be improved by replacing the instantaneous estimate

$$\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^H[k]-\mathbf{y}^n[k]\mathbf{y}^{n,H}[k] \quad (5.13)$$

[3]The absolute value $\left|\mathbf{y}_{B_y}^H\mathbf{y}_{B_y}-\mathbf{y}^{n,H}\mathbf{y}^n\right|$ has been inserted to guarantee a positive valued estimate of the clean speech energy $\mathbf{y}^{s,H}[k]\mathbf{y}^s[k]$.

in (5.6) with a time-average, which is an estimate of the speech correlation matrix $\mathcal{E}\{\mathbf{y}^s[k]\mathbf{y}^{s,H}[k]\}$. This results in

$$\mathbf{r}[k] = \frac{1}{\mu}\frac{1}{K}\left(\sum_{l=k-K+1}^{k}\mathbf{y}_{B_y}[l]\mathbf{y}_{B_y}^H[l] - \sum_{l=k-K+1}^{k}\mathbf{y}^n[l]\mathbf{y}^{n,H}[l]\right)\bar{\mathbf{w}}[k-1]. \quad (5.14)$$

However, this would require expensive matrix operations, unless the stochastic algorithm is performed in subbands (cf. Section 2.5) so that only small-sized matrices have to be computed in each subband (cf. Section 5.3.3).

A frequency-domain implementation implicitly performs this averaging (5.14) over $K = L$ samples, while offering a reduced computational complexity and a faster convergence compared to its time-domain equivalent. Therefore, the TD-SG algorithm is better implemented in the frequency-domain as is described in Section 5.2.2. To improve the performance in spectrally highly non-stationary noise, Section 5.3 proposes cheap approximations of (5.14) that allow for averaging over more than L samples.

5.2.2 Frequency-domain stochastic gradient algorithm

Similar to frequency-domain LMS [219], the frequency-domain stochastic gradient (FD-SG) algorithm is derived from a *block-based implementation of the TD-SG algorithm* with block length $K = L$, by computing the linear convolutions and correlations in the DFT-domain. The block-based TD-SG algorithm (and hence, the FD-SG algorithm) intrinsically performs the averaging (5.14) over $K = L$ samples:

$$\begin{aligned}\bar{\mathbf{w}}[kL] \;&=\; \bar{\mathbf{w}}[(k-1)L] \\ &+\frac{\rho}{L}\left[\sum_{i=0}^{L-1}\mathbf{y}^n[kL+i]\left(y_0^{n,*}[kL+i-\Delta] - \mathbf{y}^{n,H}[kL+i]\bar{\mathbf{w}}[(k-1)L]\right)\right. \\ &\left.-\frac{1}{\mu}\sum_{i=0}^{L-1}\Big(\mathbf{y}_{B_y}[kL+i]\mathbf{y}_{B_y}^H[kL+i] - \mathbf{y}^n[kL+i]\mathbf{y}^{n,H}[kL+i]\Big)\bar{\mathbf{w}}[(k-1)L]\right]. \quad (5.15)\end{aligned}$$

The gradient and hence also (5.13) are averaged over L terms prior to making adjustments to $\bar{\mathbf{w}}$. This goes at the expense of a reduced (i.e., by a factor L) update rate [26, 219]. For stationary signals, the block-based algorithm (5.15) requires an L times larger step size ρ than the non-block based algorithm to obtain the same convergence rate and misadjustment. The block-based algorithm (5.15) is convergent in the mean provided that the step size ρ is smaller than $\frac{2}{\lambda_{\max}}$ with $\lambda_{\max}$ the maximum eigenvalue of $\mathcal{E}\{\frac{1}{\mu}\mathbf{y}_{B_y}\mathbf{y}_{B_y}^H + (1-\frac{1}{\mu})\mathbf{y}^n\mathbf{y}^{n,H}\}$ [26, 219]. As the stability bound for the step size ρ of the block-based algorithm is the same as for the non-block based algorithm, the block-based algorithm is more likely to run unstable when a fast convergence rate is required.

The block-based implementation (5.15) is computationally more efficient, especially for large filter lengths, when it is implemented in the *frequency-domain*: the linear convolutions and correlations are then efficiently realized by FFT algorithms based on overlap-save or overlap-add procedures [219]. In this paper, we concentrate on a frequency-domain implementation based on overlap-save procedures. Let the $2L$-dimensional frequency-domain filters $\bar{\mathbf{W}}_l[k]$ be derived from the corresponding time-domain filters $\bar{\mathbf{w}}_l[k]$, $l = M-N, \cdots, M-1$, as

$$\begin{aligned} \bar{\mathbf{W}}_l[k] &= \left[\begin{array}{cccc} \bar{W}_{l,0}[k] & \bar{W}_{l,1}[k] & \cdots & \bar{W}_{l,2L-1}\ [k] \end{array} \right]^T \\ &= \mathbf{F} \left[\begin{array}{cccc} \bar{\mathbf{w}}_l^H[k] & 0 & \cdots & 0 \end{array} \right]^T, \end{aligned} \tag{5.16}$$

where $\mathbf{F}$ denotes the $2L \times 2L$ DFT matrix, and let during periods of noise, the $2L \times 2L$-dimensional matrices $\mathbf{Y}_l^n[k]$ and $\mathbf{Y}_l[k]$ be derived from $\mathbf{y}_l^n[kL-1]$, $\mathbf{y}_l^n[kL+L-1]$ and $\mathbf{y}_{B_y,l}[kL-1]$, $\mathbf{y}_{B_y,l}[kL+L-1]$ as

$$\mathbf{Y}_l^n[k] = \mathrm{diag}\left\{ \mathbf{F} \left[\begin{array}{cccc} y_l^n[kL-L] & ... & y_l^n[kL+L-1] \end{array} \right]^T \right\}, \tag{5.17}$$

$$\mathbf{Y}_l[k] = \mathrm{diag}\left\{ \mathbf{F} \left[\begin{array}{cccc} y_{B_y,l}[kL-L] & ... & y_{B_y,l}[kL+L-1] \end{array} \right]^T \right\}, \tag{5.18}$$

and

$$\mathbf{d}[k] = \left[\begin{array}{cccc} y_0^n[kL-\Delta] & ... & y_0^n[kL-\Delta+L-1] \end{array} \right]^T. \tag{5.19}$$

Then the FD-SG algorithm corresponding to (5.15) becomes

$$\boxed{\begin{aligned} \bar{\mathbf{W}}_l[k] &= \bar{\mathbf{W}}_l[k-1] + \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\boldsymbol{\rho}[k] \left(\mathbf{Y}_l^{n,H}[k]\mathbf{E}[k] - \mathbf{R}_l[k] \right), \\ \mathbf{R}_l[k] &= \frac{1}{\mu} \left(\mathbf{Y}_l^H[k]\mathbf{E}_y[k] - \mathbf{Y}_l^{n,H}[k]\mathbf{E}_y^n[k] \right), \end{aligned}} \tag{5.20}$$

with

$$\mathbf{E}[k] = \mathbf{F}\mathbf{k}^T \left(\mathbf{d}[k] - \mathbf{k}\mathbf{F}^{-1} \sum_{j=M-N}^{M-1} \mathbf{Y}_j^n[k]\bar{\mathbf{W}}_j[k-1] \right), \tag{5.21}$$

$$\mathbf{E}_y^n[k] = \mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1} \sum_{j=M-N}^{M-1} \mathbf{Y}_j^n[k]\bar{\mathbf{W}}_j[k-1], \tag{5.22}$$

$$\mathbf{E}_y[k] = \mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1} \sum_{j=M-N}^{M-1} \mathbf{Y}_j[k]\bar{\mathbf{W}}_j[k-1], \tag{5.23}$$

where the matrices $\mathbf{g}$ and $\mathbf{k}$ are defined as

$$\mathbf{g} = \left[\begin{array}{cc} \mathbf{I}_L & \mathbf{0}_L \\ \mathbf{0}_L & \mathbf{0}_L \end{array} \right]; \ \mathbf{k} = \left[\begin{array}{cc} \mathbf{0}_L & \mathbf{I}_L \end{array} \right]. \tag{5.24}$$

The time-varying matrix $\boldsymbol{\rho}[k]$ in (5.20) is a diagonal matrix that contains the step sizes $\rho_m[k]$, $m = 0, \cdots, 2L-1$ of the different frequency bins m, i.e.,

$$\boldsymbol{\rho}[k] = \text{diag } \{\rho_0[k], ..., \rho_{2L-1}[k]\}. \tag{5.25}$$

Each step size $\rho_m[k]$ is normalized according to the speech + noise power and the noise power in the m-th frequency bin:

$$\rho_m[k] = \frac{2\bar{\rho}}{L} P_m^{-1}[k], \tag{5.26}$$

$$P_m[k] = \gamma P_m[k-1] + (1-\gamma) \left\{ \sum_{j=M-N}^{M-1} \left|Y_{j,m}^n[k]\right|^2 + \frac{1}{\mu} \left| \sum_{j=M-N}^{M-1} \left(|Y_{j,m}[k]|^2 - \left|Y_{j,m}^n[k]\right|^2 \right) \right| \right\}, \tag{5.27}$$

where $\bar{\rho} \in [0, 2)$, γ is a forgetting factor and $Y_{j,m}^n[k]$, $Y_{j,m}[k]$ are the m-th diagonal element of $\mathbf{Y}_j^n[k]$ and $\mathbf{Y}_j[k]$, respectively.

In the *frequency-domain implementation*, each frequency bin has its own optimal step size $\rho_m[k]$, resulting in faster convergence compared to a time-domain implementation while (upon convergence) not degrading the steady-state excess MSE [219]: the different bins have a more uniform convergence rate and excess error that are less dependent on the eigenvalue spread of the correlation matrix of the input signals and hence, on the power spectra of the input signals. In a *time-domain implementation*, one common step size ρ is (implicitly) used for the different frequency bins. The convergence rate and excess errors in each bin then depend on the power spectra of the input signals: in frequency bins with little power this common step size will be smaller than in the frequency-domain approach, resulting in slower convergence and less excess error in this bin; in frequency bins with large power on the other hand, this common step size will be larger than in the frequency-domain approach, resulting in faster convergence but larger excess error.

Remark 5.2 In the case of signal model errors such as microphone mismatch, the improvement $\Delta\text{SNR}_{\text{intellig}}$ (2.45) of *a time-domain stochastic gradient implementation* is not only determined by the signal model errors themselves but also by their effect on the input power spectra. Signal model errors modify the power spectra of the noise references and hence, the convergence rates and the frequency content of the excess errors. The improvement $\Delta\text{SNR}_{\text{intellig}}$ is determined by the SNR-improvements and hence, the excess errors in each frequency bin. As a result, changes in the power spectra also have an impact on $\Delta\text{SNR}_{\text{intellig}}$. In *a frequency-domain implementation* that optimizes the step size for each frequency bin, changes in the input power spectra caused by signal

model errors have a smaller effect on the convergence rates and the excess errors in the different frequency bins so that differences in $\Delta\text{SNR}_{\text{intellig}}$ are mostly due to the signal model errors themselves and not to the change in power spectra. Hence, frequency-domain implementations are more appropriate to analyze the effect of signal model errors on $\Delta\text{SNR}_{\text{intellig}}$. □

The straightforward extension of (5.17)-(5.27) to include updating during periods of speech + noise (cf. Section 5.2.1) can be found in the tabulated **Algorithm 5.2** of Section 5.3, which summarizes the complete FD-SG algorithm (including the improvement with LPF in Section 5.3.2). Note that N FFT operations can be saved in the FD-SG algorithm, by storing the FFT-transformed speech + noise and noise-only vectors in the buffers[4] $\tilde{\mathbf{B}}_y \in \mathbb{C}^{NL \times L_{\tilde{B}_y}}$ and $\tilde{\mathbf{B}}_y^n \in \mathbb{C}^{NL \times L_{\tilde{B}_y^n}}$, respectively, instead of storing the time-domain vectors: then, the FD-SG algorithm (5.20) requires $(3N+4)$ FFTs of length $2L$. When adaptation is also done during speech + noise periods, also the L-dimensional time-domain vector $\mathbf{y}_0^n[k-\Delta]$ should be stored in an additional buffer $\tilde{\mathbf{B}}_{y_0}^n \in \mathbb{R}^{L \times L_{\tilde{B}_y^n}}$ during periods of noise-only, which for $N = M$, results in an additional storage of $L \cdot L_{\tilde{B}_y^n}$ real Words.

Remark 5.3 In (5.20), a common trade-off parameter μ is used in all frequency bins. Alternatively, a different μ can be used per frequency bin. E.g., for the SDR-GSC, μ could be set to ∞ at those frequencies where the GSC is sufficiently robust, e.g., for small-sized arrays at high frequencies (cf. Section 3.4). In this case, only a few low-frequency components of the regularization terms $\mathbf{R}_l[k]$, $l = M-N, \cdots, M-1$ have to be computed, again reducing the computational complexity. □

5.3 Improvements of the stochastic gradient algorithms

For spectrally stationary noise, the limited (i.e., $K = L$) averaging of (5.13) by the block-based TD-SG algorithm[5] and FD-SG algorithm proposed in Section 5.2.2 may offer a reasonable estimate of the *short-term* speech correlation matrix $\mathcal{E}\{\mathbf{y}^s \mathbf{y}^{s,H}\}$. However, in practical scenarios, the speech and the noise signals are often spectrally highly non-stationary (e.g., multi-talker babble noise) while their *long-term* spectral and spatial characteristics such as the positions of the sources usually vary more slowly in time. For these scenarios,

[4] Since the input signals are real, half of the FFT components are complex-conjugated. Hence, only half of the complex FFT components have to be stored in memory.

[5] A block-based algorithm with a large block length K with $K \gg NL$ would result in a too slow convergence rate.

a reliable estimate of the *long-term* speech correlation matrix $\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$ that captures the spatial rather than the *short-term* spectral characteristics can still be obtained by averaging (5.13) over $K \gg NL$ samples. Spectrally highly non-stationary noise can then still be spatially suppressed by using an estimate of the *long-term* speech correlation matrix $\mathcal{E}\{\mathbf{y}^s[k]\mathbf{y}^{s,H}[k]\}$ in the regularization term $\mathbf{r}[k]$, i.e., by setting $K \gg NL$ in (5.14). The limited averaging ($K = L$) in the FD-SG algorithm does not suffice to make such an estimate.

In this section, we propose cheap methods to incorporate a *long-term averaging* ($K \gg NL$) of (5.14) in the TD-SG algorithm and the FD-SG algorithm. A first solution is to low pass filter the regularization term $\mathbf{r}[k]$ or $\mathbf{R}_l[k]$ in (5.20). This averaging method is first explained for the TD-SG algorithm of Section 5.2.1 and then translated to the FD-SG algorithm. In Section 5.4 and Section 5.5, we will demonstrate that the proposed FD-SG algorithm with low pass filtering (LPF) preserves the benefit of the SP-SDW-MWF over the QIC-GSC, while its computational complexity is comparable to the NLMS based SPA. The FD-SG algorithm with LPF however requires data buffers for computing the regularization term, resulting in a large memory usage. To reduce the memory cost, an alternative solution for long-term averaging in the FD-SG algorithm is proposed in Section 5.3.3. We show that by approximating the long-term averaged regularization term (5.14) in the frequency-domain using diagonal correlation matrices instead of data buffers, the memory usage of the FD-SG algorithm can be decreased so that it fits within the memory constraints of current commercial hearing aids [41].

5.3.1 TD-SG algorithm with low pass filtering

Assume that the *long-term* spectral and spatial characteristics of the noise are quasi-stationary during at least K speech + noise samples and K noise samples. A reliable estimate of the long-term speech correlation matrix $\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$ is then obtained by (5.14) with $K \gg NL$. The estimate (5.14) can be approximated using exponential weighting with a forgetting factor $\tilde{\lambda} < 1$ as

$$\frac{1}{\mu}(1-\tilde{\lambda})\sum_{l=0}^{k}\tilde{\lambda}^{k-l}\left(\mathbf{y}_{B_y}[l]\mathbf{y}_{B_y}^H[l] - \mathbf{y}^n[l]\mathbf{y}^{n,H}[l]\right)\bar{\mathbf{w}}[k-1], \tag{5.28}$$

where $\tilde{\lambda}$ corresponds to an averaging window of about $K = \frac{1}{1-\tilde{\lambda}}$ samples. To avoid expensive matrix computations, we approximate (5.28) by

$$\mathbf{r}[k] = \frac{1}{\mu}(1-\tilde{\lambda})\sum_{l=0}^{k}\tilde{\lambda}^{k-l}\left(\mathbf{y}_{B_y}[l]\mathbf{y}_{B_y}^H[l] - \mathbf{y}^n[l]\mathbf{y}^{n,H}[l]\right)\bar{\mathbf{w}}[l-1]. \tag{5.29}$$

Since the filter coefficients $\bar{\mathbf{w}}[k]$ of a stochastic gradient algorithm vary slowly in time, (5.29) appears to be a good approximation of (5.28), especially for

small step sizes $\bar{\rho}$. The averaging operation (5.29) is recursively implemented by a low pass filtering (LPF) operation, i.e.,

$$\boxed{\mathbf{r}[k] = \tilde{\lambda}\mathbf{r}[k-1] + (1-\tilde{\lambda})\frac{1}{\mu}\left(\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^H[k] - \mathbf{y}^n[k]\mathbf{y}^{n,H}[k]\right)\bar{\mathbf{w}}[k-1].} \quad (5.30)$$

The normalized step size ρ in (5.11) is modified into

$$\rho = \frac{\bar{\rho}}{\zeta[k] + \mathbf{y}^{n,H}[k]\mathbf{y}^n[k] + c}, \quad (5.31)$$

$$\zeta[k] = \tilde{\lambda}\zeta[k-1] + (1-\tilde{\lambda})\frac{1}{\mu}\left|\mathbf{y}_{B_y}^H[k]\mathbf{y}_{B_y}[k] - \mathbf{y}^{n,H}[k]\mathbf{y}^n[k]\right|. \quad (5.32)$$

The complete TD-SG algorithm with low pass filtering is summarized in the tabulated **Algorithm 5.1**. In the sequel, we use the abbreviation LPF to refer to the low pass filtering.

5.3.2 FD-SG algorithm with low pass filtering

Equation (5.30) can be translated to the FD-SG algorithm (5.20) by incorporating the low pass filtering operation in the frequency-domain regularization term $\mathbf{R}_l[k]$, $l = M-N, \ldots, M-1$:

$$\boxed{\mathbf{R}_l[k] = \lambda\mathbf{R}_l[k-1] + (1-\lambda)\frac{1}{\mu}\left(\mathbf{Y}_l^H[k]\mathbf{E}_y[k] - \mathbf{Y}_l^{n,H}[k]\mathbf{E}_y^n[k]\right),} \quad (5.33)$$

where $\mathbf{E}_y^n[k]$ and $\mathbf{E}_y[k]$ are defined in (5.22) and (5.23). Note that without LPF (i.e., $\lambda = 0$), the FD-SG algorithm intrinsically performs an averaging over L samples (cf. Section 5.2.2). Hence, to obtain the same time constant in the averaging operation as in the TD-SG algorithm with LPF, λ should equal $\tilde{\lambda}^L$. The step sizes $\rho_m[k] = \frac{2\bar{\rho}}{LP_m[k]}$ in the different frequency bins are now normalized using:

$$P_m[k] = \gamma P_m[k-1] + (1-\gamma)\left(P_{y,m}^n[k] + P_{y,m}^s[k]\right), \quad (5.34)$$

$$P_{y,m}^n[k] = \sum_{j=M-N}^{M-1}\left|Y_{j,m}^n[k]\right|^2, \quad (5.35)$$

$$P_{y,m}^s[k] = \lambda P_{y,m}^s[k-1] + (1-\lambda)\frac{1}{\mu}\left|\sum_{j=M-N}^{M-1}\left(|Y_{j,m}[k]|^2 - \left|Y_{j,m}^n[k]\right|^2\right)\right|. \quad (5.36)$$

Experiments in Section 5.5 show that the LPF (5.33) indeed improves the performance of the FD-SG algorithm, especially for spectrally highly non-stationary noise. The term $\bar{\mathbf{w}}^H[k-1]\mathbf{r}[k]$ with $\mathbf{r}[k]$ computed as (5.30) and

Algorithm 5.1 TD-SG algorithm with LPF.

Initialization and matrix definitions:

$$\bar{\mathbf{w}}[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{C}^{NL\times 1}$$

Speech+noise buffer $\mathbf{B}_y \in \mathbb{C}^{NL\times L_{B_y}}$; Noise buffer $\mathbf{B}_y^n \in \mathbb{C}^{ML\times L_{B_y^n}}$

For each new input sample (per channel):

If noise detected:

$$\mathbf{y}^n[k] = \mathbf{y}[k] = \begin{bmatrix} \mathbf{y}_{M-N}^H[k] & \mathbf{y}_{M-N+1}^H[k] & \cdots & \mathbf{y}_{M-1}^H & [k] \end{bmatrix}^H$$

$\mathbf{y}_{B_y}[k] = NL$-dimensional speech + noise vector from the buffer $\mathbf{B}_y$

Store input vectors $\mathbf{y}_l[k], l = 0, \ldots, M-1$ in noise buffer $\mathbf{B}_y^n$

Update formula:

$$\bar{\mathbf{w}}[k] = \bar{\mathbf{w}}[k-1] + \rho\left\{\mathbf{y}^n[k](y_0^{n,*}[k-\Delta] - \mathbf{y}^{n,H}[k]\bar{\mathbf{w}}[k-1]) - \mathbf{r}[k]\right\}$$
$$\mathbf{r}[k] = \tilde{\lambda}\mathbf{r}[k-1] + (1-\tilde{\lambda})\frac{1}{\mu}\left(\mathbf{y}_{B_y}[k]\mathbf{y}_{B_y}^H[k] - \mathbf{y}^n[k]\mathbf{y}^{n,H}[k]\right)\bar{\mathbf{w}}[k-1]$$

with:

$$\rho = \frac{\bar{\rho}}{\zeta[k] + \mathbf{y}^{n,H}[k]\mathbf{y}^n[k] + c}$$
$$\zeta[k] = \tilde{\lambda}\zeta[k-1] + (1-\tilde{\lambda})\frac{1}{\mu}\left|\mathbf{y}_{B_y}^H[k]\mathbf{y}_{B_y}[k] - \mathbf{y}^{n,H}[k]\mathbf{y}^n[k]\right|$$

If speech detected:

$$\mathbf{y}[k] = \begin{bmatrix} \mathbf{y}_{M-N}^H[k] & \mathbf{y}_{M-N+1}^H[k] & \cdots & \mathbf{y}_{M-1}^H & [k] \end{bmatrix}^H$$

$\mathbf{y}_{B_y^n}[k] = NL$-dimensional noise vector from the buffer $\mathbf{B}_y^n$

Store input vectors $\mathbf{y}_l[k], l = M-N, \ldots, M-1$ in speech + noise buffer $\mathbf{B}_y$

Update formula:

$$\bar{\mathbf{w}}[k] = \bar{\mathbf{w}}[k-1] + \rho\left\{\mathbf{y}_{B_y^n}[k](y_{B_y^n,0}^*[k-\Delta] - \mathbf{y}_{B_y^n}^H[k]\bar{\mathbf{w}}[k-1]) - \mathbf{r}[k]\right\}$$
$$\mathbf{r}[k] = \tilde{\lambda}\mathbf{r}[k-1] + (1-\tilde{\lambda})\frac{1}{\mu}\left(\mathbf{y}[k]\mathbf{y}^H[k] - \mathbf{y}_{B_y^n}[k]\mathbf{y}_{B_y^n}^H[k]\right)\bar{\mathbf{w}}[k-1]$$

with:

$$\rho = \frac{\bar{\rho}}{\zeta[k] + \mathbf{y}_{B_y^n}^H[k]\mathbf{y}_{B_y^n}[k] + c}$$
$$\zeta[k] = \tilde{\lambda}\zeta[k-1] + (1-\tilde{\lambda})\frac{1}{\mu}\left|\mathbf{y}^H[k]\mathbf{y}[k] - \mathbf{y}_{B_y^n}^H[k]\mathbf{y}_{B_y^n}^n[k]\right|$$

Output signal $z[k]$: $z[k] = y_0[k-\Delta] - \bar{\mathbf{w}}^H[k-1]\mathbf{y}[k]$

hence, also the FD term $\sum_{l=M-N}^{M-1} \bar{\mathbf{W}}_l^H[k-1]\mathbf{F}\mathbf{g}\mathbf{F}^{-1}\mathbf{R}_l[k]$ with $\mathbf{R}_l[k]$ calculated as in (5.33), is an (approximate) measure for the average speech distortion over the last $K = 1/(1-\tilde{\lambda})$ samples. Hence, the FD-SG algorithm with LPF reduces noise while keeping the average distortion low. The LPF reduces fluctuations in the filter weights $\bar{\mathbf{W}}_l[k]$ caused by poor estimates of the speech correlation matrix $\mathcal{E}\{\mathbf{y}^s[k]\mathbf{y}^{s,H}[k]\}$ and/or by the highly non-stationary *short-term* speech spectrum.

Admittedly, due to the long-term averaging by the LPF, only slow changes in the nominal position of the desired speaker can be tracked[6]. However, in hearing aid applications, we assume the nominal position of the desired speaker to be in front of the hearing aid so that the spatial characteristics of the speech are expected to vary slowly in time.

Sudden changes in the noise scenario on the other hand are found to be tracked quite well (i.e., similar to GSC) by the FD-SG algorithm with LPF (with e.g., $\lambda = 0.999$). This is experimentally demonstrated in Section 5.5. During periods of noise-only, the update formula (5.16) is given by

$$\bar{\mathbf{W}}_l[k] = \bar{\mathbf{W}}_l[k-1] + \underbrace{\mathbf{F}\mathbf{g}\mathbf{F}^{-1}\boldsymbol{\rho}[k]\mathbf{Y}_l^{n,H}[k]\mathbf{E}[k]}_{\delta\bar{\mathbf{W}}_{l,1}[k]} - \underbrace{\mathbf{F}\mathbf{g}\mathbf{F}^{-1}\boldsymbol{\rho}[k]\mathbf{R}_l[k]}_{\delta\bar{\mathbf{W}}_{l,2}[k]}. \qquad (5.37)$$

A sudden change in noise scenario has an immediate impact on the term $\delta\bar{\mathbf{W}}_{l,1}[k]$: like for GSC, the noise in the new scenario is suppressed with a convergence time determined by the step size matrix $\boldsymbol{\rho}[k]$. The amount of noise suppression and speech distortion is effectively determined by the term $\delta\bar{\mathbf{W}}_{l,2}[k]$, which converges more slowly than $\delta\bar{\mathbf{W}}_{l,1}[k]$ due to the averaging (5.33): initially, the distortion measured before the change determines $\mathbf{R}_l[k]$ and hence, the amount of noise suppression and speech distortion. Experiments show that also in the transient phase after the change in noise scenario, the speech distortion of the FD-SG implementation with LPF of the SP-SDW-MWF (with $\frac{1}{\mu} > 0$) is small compared to the speech distortion of the GSC (i.e., the SDR-GSC with $\frac{1}{\mu} = 0$).

The complete FD-SG algorithm with LPF is summarized in the tabulated **Algorithm 5.2**. Compared to the FD-SG algorithm of Section 5.2.2, the FD-SG algorithm with LPF requires one extra $2L-$point FFT and about $14NL$ extra operations per L samples and additional storage of an $NL \times 1$ complex data vector.

[6]Note that this also holds for the QRD and GSVD based implementations, where typically rather large exponential window lengths are used in the computation of the speech + noise and noise correlation matrix (cf. Section 2.4.2).

Algorithm 5.2 FD-SG algorithm with LPF based on overlap-save.

Initialization and matrix definitions:

$\bar{\mathbf{W}}_l[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T$, $l = M-N, ..., M-1$; $\quad P_m[0] = c_m$, $m = 0, ..., 2L-1$

Speech+noise buffer $\tilde{\mathbf{B}}_y \in \mathbb{C}^{NL \times L_{\tilde{B}_y}}$; Noise buffers $\tilde{\mathbf{B}}_y^n \in \mathbb{C}^{NL \times L_{\tilde{B}_y^n}}$, $\tilde{\mathbf{B}}_{y_0}^n \in \mathbb{R}^{L \times L_{\tilde{B}_y^n}}$

For each new block of L input samples (per channel):

If noise detected:

$$\mathbf{d}[k] = \begin{bmatrix} y_0[kL-\Delta] & \cdots & y_0[kL-\Delta+L-1] \end{bmatrix}^T$$

$$\mathbf{Y}_l^n[k] = \text{diag}\left\{ \mathbf{F} \begin{bmatrix} y_l[kL-L] & \cdots & y_l[kL+L-1] \end{bmatrix}^T \right\}$$

Store input data $\mathbf{Y}_l^n[k]$, $\mathbf{d}[k]$ in noise buffers $\tilde{\mathbf{B}}_y^n$, $\tilde{\mathbf{B}}_{y_0}^n$

Create $\mathbf{Y}_l[k]$ from data in speech+noise buffer $\tilde{\mathbf{B}}_y$

If speech detected:

$$\mathbf{Y}_l[k] = \text{diag}\left\{ \mathbf{F} \begin{bmatrix} y_l[kL-L] & \cdots & y_l[kL+L-1] \end{bmatrix}^T \right\}$$

Store input data $\mathbf{Y}_l[k]$ in speech + noise buffer $\tilde{\mathbf{B}}_y$

Create $\mathbf{Y}_l^n[k]$, $\mathbf{d}[k]$ from data in noise buffers $\tilde{\mathbf{B}}_y^n$, $\tilde{\mathbf{B}}_{y_0}^n$

Update formula:

$$\bar{\mathbf{W}}_l[k] = \bar{\mathbf{W}}_l[k-1] + \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\boldsymbol{\rho}[k]\left\{ \mathbf{Y}_l^{n,H}[k]\mathbf{E}[k] - \mathbf{R}_l[k] \right\}$$

$$\mathbf{R}_l[k] = \lambda \mathbf{R}_l[k-1] + (1-\lambda)\frac{1}{\mu}\left(\mathbf{Y}_l^H[k]\mathbf{E}_y[k] - \mathbf{Y}_l^{n,H}[k]\mathbf{E}_y^n[k] \right)$$

with

$$\mathbf{E}_y^n[k] = \mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j^n[k]\bar{\mathbf{W}}_j[k-1] = \mathbf{F}\mathbf{k}^T\mathbf{e}_y^n[k]$$

$$\mathbf{E}[k] = \mathbf{F}\mathbf{k}^T\left(\mathbf{d}[k] - \mathbf{e}_y^n[k] \right) = \mathbf{F}\mathbf{k}^T\mathbf{e}[k]$$

$$\mathbf{E}_y[k] = \mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j[k]\bar{\mathbf{W}}_j[k-1] = \mathbf{F}\mathbf{k}^T\mathbf{e}_y[k]$$

and

$$\boldsymbol{\rho}[k] = \frac{2\bar{\rho}}{L}\text{diag}\ \left\{ P_0^{-1}[k], ..., P_{2L-1}^{-1}[k] \right\}$$

$$P_m[k] = \gamma P_m[k-1] + (1-\gamma)\left(P_{y,m}^n[k] + P_{y,m}^s[k] \right)$$

$$P_{y,m}^n[k] = \sum_{j=M-N}^{M-1}\left| Y_{j,m}^n[k] \right|^2$$

$$P_{y,m}^s[k] \quad = \quad \lambda P_{y,m}^s[k-1] + (1-\lambda)\frac{1}{\mu}\left| \sum_{j=M-N}^{M-1}\left(|Y_{j,m}[k]|^2 - \left| Y_{j,m}^n[k] \right|^2 \right) \right|$$

Output $\mathbf{z}[k] = \begin{bmatrix} z[kL] & \cdots & z[kL+L-1] \end{bmatrix}^T$*:*

$\mathbf{y}_0[k] = \begin{bmatrix} y_0[kL-\Delta] & \cdots & y_0[kL-\Delta+L-1] \end{bmatrix}^T$

- If noise-only detected: $\mathbf{z}[k] = \mathbf{y}_0[k] - \mathbf{e}_y^n[k]$
- If speech + noise detected: $\mathbf{z}[k] = \mathbf{y}_0[k] - \mathbf{e}_y[k]$

5.3.3 FD-SG algorithm using correlation matrices

In Section 5.4, we will show that in terms of computational complexity, the FD-SG algorithm with LPF is feasible for implementation in commercial hearing aids. In current hearing aids, not only the processing power but also the memory space is limited. E.g., the latest digital commercial BTE hearing aids typically have a data memory of about 2 to 4 kByte (or 1 to 2 kWords of 16 bit) [129, 215]. The stochastic gradient algorithms of the previous sections require a speech + noise data buffer and/or a noise data buffer and hence, a memory of at least[7] $2LN \cdot L_{\tilde{B}_y}$ real Words. In order to achieve a good performance during longer periods of noise-only (or speech + noise), typical values for $2L \cdot L_{\tilde{B}_y}$ (or $2L \cdot L_{\tilde{B}_y^n}$) are between 10000 and 20000, leading to a memory usage of $10N \ldots 20N$ kWords.

In this section, we show that the use of data buffers can be avoided in the FD-SG algorithm by approximating the long-term averaged regularization term (5.14) or (5.28) using (diagonal) correlation matrices instead of using the approximation (5.29) [41]. Instead of data buffers, only a (diagonal) speech+noise and noise correlation matrix then have to be stored.

The starting point for the derivation of the FD-SG algorithm using correlation matrices consists of the frequency-domain equivalent of the long-term averaged regularization term (5.28). This results in the following modifications for the computation of the regularization term $\mathbf{R}_l[k]$ and the power estimate $P_m^s[k]$ in the step size matrix $\boldsymbol{\rho}[k]$:

$$\mathbf{R}_l[k] = \frac{1}{\mu} \sum_{j=M-N}^{M} \left(\mathbf{S}_y^{lj}[k] - \mathbf{S}_y^{n,lj}[k] \right) \bar{\mathbf{W}}_l[k-1], \tag{5.38}$$

$$P_{y,m}^s[k] = \frac{1}{\mu} \left| \sum_{j=M-N}^{M} \left(S_{y,m}^{jj}[k] - S_{y,m}^{n,jj}[k] \right) \right| \tag{5.39}$$

with

$$\mathbf{S}_y^{lj}[k] = \lambda \mathbf{S}_y^{lj}[k-1] + (1-\lambda) \mathbf{Y}_l^H[k] \mathbf{F}\mathbf{k}^T \mathbf{k}\mathbf{F}^{-1} \mathbf{Y}_j[k], \tag{5.40}$$

$$\mathbf{S}_y^{n,lj}[k] = \lambda \mathbf{S}_y^{n,lj}[k-1] + (1-\lambda) \mathbf{Y}_l^{n,H}[k] \mathbf{F}\mathbf{k}^T \mathbf{k}\mathbf{F}^{-1} \mathbf{Y}_j^n[k], \tag{5.41}$$

where the speech + noise and noise correlation matrices $\mathbf{S}_y^{lj}[k]$ and $\mathbf{S}_y^{n,lj}[k]$, $l, j = M-N, \ldots, M-1$ are updated during periods of speech + noise and noise-only, respectively[8], and where, like in the FD-SG with LPF, the exponential weighting factor $\lambda < 1$ determines the averaging window length K in

[7] If adaptation is only performed during periods of noise-only, only a speech + noise data buffer has to be stored.

[8] When using correlation matrices, filter adaptation can only take place during periods of noise-only, since during speech + noise periods the desired signal vector $\mathbf{d}[k]$ in Algorithm 5.2 cannot be constructed from the noise buffer $\tilde{\mathbf{B}}_{y_0}^n$ anymore.

Table 5.1: Total memory cost of FD-SG algorithm with LPF and FD-SG algorithm using CM for $M = 3$, $L = 32$, $L_{\tilde{B}_y} = L_{\tilde{B}_y^n} = L_{\tilde{B}_y^n} = \frac{10000}{2L}$. (a) $N = M$; (b) $N = M - 1$.

Algorithm	*Memory*	*kWords*
FD-SG with LPF:		
adapted during noise-only	$2NL \cdot L_{\tilde{B}_y} + 6NL + 8L$	$20.64^{(a)}$, $30.83^{(b)}$
continuously adapted	$2NL(L_{\tilde{B}_y} + L_{\tilde{B}_y^n}) + L \cdot L_{\tilde{B}_{y_0}^n}$ $+4NL + 7L$	$45,51^{(a)}$, $65,61^{(b)}$
FD-SG using CM	$2N^2L + 6NL + 7L$	$0.86^{(a)}$, $1.38^{(b)}$

(5.14) [41]. Like in a (fullband) TD-SG algorithm, these modifications require expensive matrix operations with a complexity of $\mathcal{O}(N^2L^2)$, while not necessarily reducing the memory cost (storage of the correlation matrices still requires $(NL)^2$ real Words).

As already indicated in Section 5.2.1, the memory usage and computational complexity of (5.38)-(5.41) become feasible when only small-sized matrices have to be computed and stored (e.g., in a subband implementation). Since[9] $\mathbf{F}\mathbf{k}^T\mathbf{k}\mathbf{F}^{-1} \approx \frac{1}{2}\mathbf{I}_{2L}$, [8, 18], the frequency-domain correlation matrices $\mathbf{S}_y^{lj}[k]$ and $\mathbf{S}_y^{n,lj}[k]$ can be approximated by the diagonal matrices [41]

$$\mathbf{S}_y^{lj}[k] = \lambda\mathbf{S}_y^{lj}[k-1] + (1-\lambda)\mathbf{Y}_l^H[k]\mathbf{Y}_j[k]/2, \tag{5.42}$$

$$\mathbf{S}_y^{n,lj}[k] = \lambda\mathbf{S}_y^{n,lj}[k-1] + (1-\lambda)\mathbf{Y}_l^{n,H}[k]\mathbf{Y}_j^n[k]/2, \tag{5.43}$$

reducing the memory usage for $\mathbf{S}_y^{lj}[k]$ and $\mathbf{S}_y^{n,lj}[k]$ to $2LN^2$ real Words[10] (e.g., 0.6 kWords for $L = 32$ and $N = 3$). The implementation of this FD-SG algorithm, which will be referred to as FD-SG algorithm using correlation matrices (CM), is summarized in the tabulated **Algorithm 5.3**. The total memory cost of the FD-SG algorithm using CM and the FD-SG algorithm with LPF is compared in **Table 5.1**, showing that for a small number of microphones M (e.g., $M = 3$), the FD-SG algorithm using CM has a significantly lower memory cost than the FD-SG algorithm with LPF. For the computational complexity we refer to Section 5.4.

Preliminary experiments demonstrate that the FD-SG algorithm using CM, has about the same performance as the FD-SG algorithm with LPF [41]. In further research, the convergence and tracking performance of both algorithms

[9] This is a common approximation in the development of frequency-domain acoustic echo cancellation algorithms.

[10] Since $\mathbf{S}_y^{lj}[k] = \mathbf{S}_y^{jl,H}[k]$ and half of the diagonal elements of $\mathbf{S}_k^{lj}[k]$ are complex-conjugated, only $2LN^2$ real Words have to be stored in memory.

Algorithm 5.3 FD-SG algorithm using (diagonal) correlation matrices.

Initialization and matrix definitions:

$$\bar{\mathbf{W}}_l[0] = \left[\begin{array}{ccc} 0 & \cdots & 0 \end{array}\right]^T,\ l = M-N,\ ...,\ M-1; \quad P_m[0] = c_m,\ m = 0,\ ...,\ 2L-1$$

For each new block of L input samples (per channel):

$$\mathbf{Y}_l[k] = \text{diag}\left\{\mathbf{F}\left[\begin{array}{ccc} y_l[kL-L] & \cdots & y_l[kL+L-1] \end{array}\right]^T\right\}$$

$$\mathbf{d}[k] = \left[\begin{array}{ccc} y_0[kL-\Delta] & \cdots & y_0[kL-\Delta+L-1] \end{array}\right]^T$$

$$\mathbf{e}[k] = \mathbf{d}[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j[k]\bar{\mathbf{W}}_j[k-1]; \quad \mathbf{E}[k] = \mathbf{F}\mathbf{k}^T\mathbf{e}[k]$$

If noise detected: $\mathbf{Y}_l^n[k] = \mathbf{Y}_l[k]$

$$\mathbf{S}_y^{n,lj}[k] = \lambda\mathbf{S}_y^{n,lj}[k-1] + (1-\lambda)\mathbf{Y}_l^{n,H}[k]\mathbf{Y}_j^n[k]/2$$

If speech detected:

$$\mathbf{S}_y^{lj}[k] = \lambda\mathbf{S}_y^{lj}[k-1] + (1-\lambda)\mathbf{Y}_l^H[k]\mathbf{Y}_j[k]/2$$

Update formula (only during noise-only periods):

$$\bar{\mathbf{W}}_l[k] = \bar{\mathbf{W}}_l[k-1] + \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\boldsymbol{\rho}[k]\left\{\mathbf{Y}_l^{n,H}[k]\mathbf{E}_y^n[k] - \mathbf{R}_l[k]\right\}$$

$$\mathbf{R}_l[k] = \frac{1}{\mu}\sum_{j=M-N}^{M}\left(\mathbf{S}_y^{lj}[k] - \mathbf{S}_y^{n,lj}[k]\right)\bar{\mathbf{W}}_l[k-1]$$

with $\mathbf{E}_y^n[k] = \mathbf{E}[k]$ and

$$\boldsymbol{\rho}[k] = \frac{2\bar{\rho}}{L}\text{diag}\ \left\{P_0^{-1}[k],\ ...,\ P_{2L-1}^{-1}[k]\right\}$$

$$P_m[k] = \gamma P_m[k-1] + (1-\gamma)\left(P_{y,m}^n[k] + P_{y,m}^s[k]\right)$$

$$P_{y,m}^n[k] = \sum_{j=M-N}^{M-1}\left|Y_{j,m}^n[k]\right|^2$$

$$P_{y,m}^s[k] \quad = \quad \frac{1}{\mu}\left|\sum_{j=M-N}^{M-1}\left(S_{y,m}^{jj}[k] - S_{y,m}^{n,jj}[k]\right)\right|$$

Output $\mathbf{z}[k] = \left[\begin{array}{ccc} z[kL] & \cdots & z[kL+L-1] \end{array}\right]^T$*:* $\mathbf{z}[k] = \mathbf{e}[k]$

will be compared. Like in the FD-SG algorithm with LPF and the NLMS based GSC, a change in noise scenario during periods of noise-only has an immediate impact on the filter coefficients $\bar{\mathbf{W}}_l[k]$, so that the tracking performance is expected to be similar.

5.4 Computational complexity

Table 5.2 summarizes the computational complexity expressed in the number of operations (counted as the number of real multiplications, additions, divisions, square roots and absolute values) of the TD-SG and FD-SG implementations of the SP-SDW-MWF[11]. We assume that one complex multiplication is equivalent to 4 real multiplications and 2 real additions. A $2L$-point FFT of a real input vector requires $4L \log_2 2L$ real operations (assuming the radix-2 FFT algorithm). Comparison is made with standard (TD as well as FD) NLMS ANC (2.82) and the NLMS SPA [29, 103] (2.95) described in Section 2.3. The update equation of the FD-NLMS ANC corresponds to the update equation (5.20) of the FD-SG algorithm with the number of input channels N set to $M-1$ and the trade-off parameter $1/\mu$ set to 0. The NLMS based SPA is translated to the frequency-domain by the following equations:

$$\|\mathbf{w}_a[k]\|_2^2 = \mathbf{w}_a^T[k]\mathbf{w}_a[k] = \frac{1}{2L}\sum_{l=1}^{M-1}\mathbf{W}_{al}^H[k]\mathbf{W}_{al}[k]; \tag{5.44}$$

$$\text{If } \|\mathbf{w}_a[k]\|_2^2 \geq \zeta^2 : \mathbf{W}_{al}[k] \leftarrow \zeta\frac{\mathbf{W}_{al}[k]}{\|\mathbf{w}_a[k]\|_2}, \, l = 1, \cdots, M-1. \tag{5.45}$$

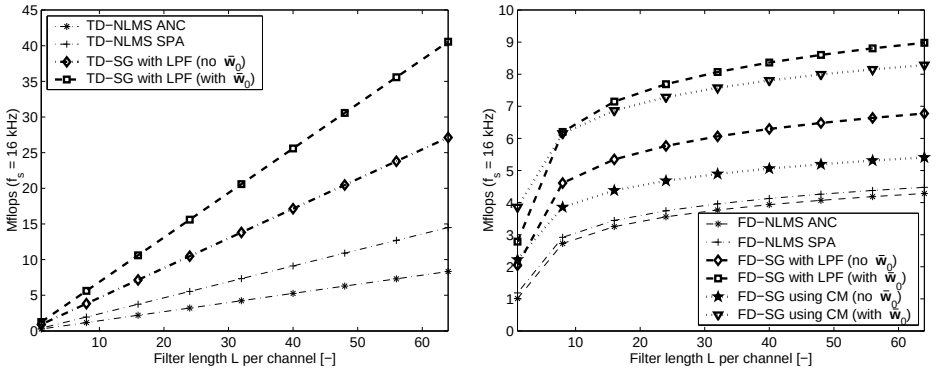
Table 5.2 indicates that the TD-SG SDR-GSC (i.e., the SP-SDW-MWF without filter $\bar{\mathbf{w}}_0$ and hence, $N = M-1$) and the NLMS SPA are about twice as complex as the standard NLMS ANC. When applying an LPF to the regularization term, the TD-SG SDR-GSC algorithm has about three times the computational complexity of the NLMS ANC. The SP-SDW-MWF with extra filter $\bar{\mathbf{w}}_0$ on the speech reference is a bit more complex. The increase in complexity of the frequency-domain implementations is smaller.

As an illustration, **Figure 5.1(a) and (b)** plot the computational complexity (expressed in Mflops) of the TD-SG and FD-SG algorithm with LPF as a function of L for $M = 3$ and a sampling frequency $f_s = 16$ kHz. Comparison is made with the NLMS ANC and the NLMS SPA. Figure 5.1(a) plots the complexity of the time-domain algorithms while Figure 5.1(b) depicts the complexity of the frequency-domain algorithms. For $L > 8$, the frequency-domain

[11] In the calculation of the computational complexity of the time-domain algorithms, we assumed that the normalized step size is efficiently computed using an IIR low pass filter (cf. (2.84) of Section 2.3.1)).

Table 5.2: Computational complexity of the TD-SG algorithm and FD-SG algorithm implementations of the SP-SDW-MWF (expressed as the number of operations per sample).

Algorithm		*update formula*	*normalized step size*
TD	NLMS ANC	$4(M-1)L+1$	$2M+2$
	NLMS SPA	$7(M-1)L+2$	$2M+2$
	SG	$8NL+4$	$4N+10$
	SG with LPF	$13NL+2$	$4N+12$
FD	NLMS ANC	$16M-15+(12M-4)\log_2 2L$	$4M-1$
	NLMS SPA	$22M-21+(12M-4)\log_2 2L$	$4M-1$
	SG	$32N+2+(12N+16)\log_2 2L$	$8N+6$
	SG with LPF	$46N+(12N+20)\log_2 2L$	$8N+8$
	SG using CM	$14N^2+20N+1+(12N+8)\log_2 2L$	$6N+3$



(a) (b)

Figure 5.1: Computational complexity (expressed in Mflops) of (a) TD-SG with LPF; (b) FD -SG with LPF and FD-SG using CM as a function of filter length L per channel; $M = 3$. For comparison, the computational complexity of the standard NLMS ANC and SPA are depicted too.

implementations result in a significantly lower computational cost compared to their time-domain equivalents. For small M, the computational complexity of the FD-NLMS SPA is not significantly different from the cost of the FD-NLMS ANC. The computational cost of the FD-SG algorithm with LPF is comparable to the complexity of the FD-NLMS SPA. In Figure 5.1(b), also the FD-SG algorithm using CM is depicted, showing that for small M, its computational complexity is even lower than the FD-SG algorithm with LPF. Given that a DSP typically performs 1 multiply-accumulate (MAC) per cycle and that almost all operations in the stochastic gradient algorithms are additions or multiplications, the number of MAC per second (and hence, the minimum number of MIPS) equals about half the number of flops depicted in Table 5.2 and Figure 5.1. Hence, we may conclude that in terms of computational complexity, the FD-SG algorithm with LPF and the FD-SG algorithm using CM are both feasible for implementation in commercial hearing aids.

Remark 5.4 In Table 5.2 and Figure 5.1, the computational complexity of time-domain and frequency-domain NLMS ANC and NLMS SPA represents the computational complexity when the adaptive filter is only updated during noise-only periods. If the adaptive filter is also updated during speech + noise periods using data from a noise buffer, the time-domain implementations require $2(M-1)L$ additional operations per sample and the frequency-domain implementations require 2 additional FFT and $(8L(M-1))$ operations per L samples to compute the output.
The computational complexity of NLMS SPA in Table 5.2 and Figure 5.1 represents the worst-case complexity, i.e., assuming that the ANC $\mathbf{w}_a$ has to be scaled down after each update. □

5.5 Experimental results

In this section, we assess the performance of the proposed stochastic gradient algorithms based on experimental results with the three-microphone BTE. In the experiments, we concentrate on the FD-SG algorithm and the improved FD-SG algorithm with LPF. We analyze the impact of the low pass filter on the FD-SG algorithm and compare the performance of the FD-SG algorithm with LPF and the FD-NLMS SPA. Preliminary results in [41] show that the FD-SG algorithm using CM has about the same steady-state performance as the FD-SG algorithm with LPF for a given exponential weighting factor λ. As a result, most of the experimental results obtained here with the FD-SG algorithm with LPF (e.g., the comparison with FD-NLMS SPA) are also representative of the FD-SG algorithm using CM.

5.5.1 Implementation issues

In the experiments, the desired speech source consists of sentences spoken by a male speaker from the HINT-database [185]. Both the speech and the noise signal have a level of 70 dB SPL at the center of the head. The microphones have been calibrated by means of anechoic recordings of a stationary speech-weighted noise signal at 0°, measured while the microphone array was mounted on the head.

The performance of the FD-SG algorithms is evaluated for a filter length $L = 32$ taps per channel, $\rho = 0.8$ and $\gamma = 0.95$. For these values of ρ and L, a good convergence behavior and noise reduction performance has been obtained: for larger L, the performance did not improve significantly. As indicated in Section 5.2, the approximation (5.6) of the regularization term in the FD-SG algorithm causes an additional excess error that increases with $\bar{\rho}$ and L, hence, $\bar{\rho}$ and L should not be chosen too large. The microphone signals are pre-whitened prior to processing to improve the intelligibility-weighted performance (cf. Remark 2.12 of Section 2.5.1) [150], and the output is accordingly de-whitened. A DS beamformer is used as a fixed beamformer, since it is robust to model errors in the case of small microphone interspacings. The blocking matrix $\mathbf{B}$ pairwise subtracts the time aligned calibrated microphone signals. In the experiments, the speech + noise and noise periods have been marked manually.

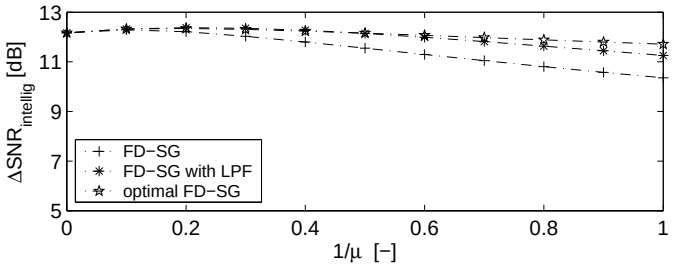
To exclude the effect of the spatial pre-processor, the performance measures $\Delta\text{SNR}_{\text{intellig}}$ (2.45) and $\text{SD}_{\text{intellig}}$ (2.49) are calculated with respect to the output of the fixed beamformer. In some experiments, a microphone gain mismatch $\upsilon_2 = 4$ dB is applied to the second microphone to illustrate the sensitivity of the algorithms to errors in the assumed signal model.

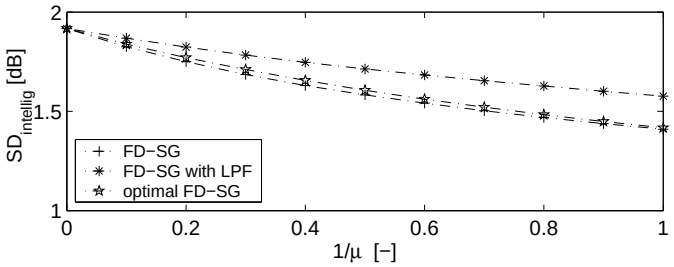
5.5.2 Comparison of the FD-SG and the FD-SG algorithm with LPF

In this section, the performance of the FD-SG (i.e., Algorithm 5.2 with $\lambda = 0$) and the FD-SG with LPF (i.e., Algorithm 5.2 with $\lambda \neq 0$) implementation of the SP-SDW-MWF are compared.

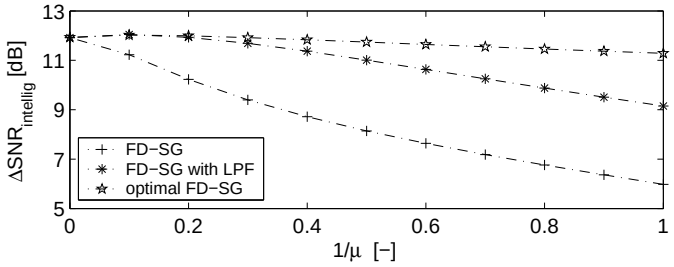
In **Figure 5.2 (a) and (b)**, we compare the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the distortion $\text{SD}_{\text{intellig}}$ of the FD-SG and the FD-SG with LPF (with exponential weighting factor $\lambda = 0.999$) implementation of the SDR-GSC (i.e., SP-SDW-MWF without $\bar{\mathbf{w}}_0$) for different values of the trade-off parameter[12] $1/\mu$. Figure 5.2(a) depicts the performance for a stationary (i.e., speech-weighted noise) noise source at 90°, while Figure 5.2(b) plots the performance measures

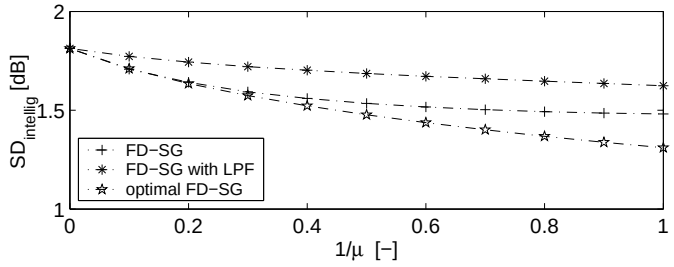
[12]Note that the SDR-GSC reduces to the GSC if $1/\mu = 0$, since then, $\mathbf{R}_i[k] = 0$ (see also (5.20)).





(a)





(b)

Figure 5.2: Performance of the FD-SG and FD-SG with LPF ($\lambda = 0.999$) implementation of the SDR-GSC compared to the performance of the optimal FD-SG implementation for different values of $1/\mu$; (a) Stationary speech-weighted noise at 90° ; (b) Multi-talker babble noise at 90°.

for spectrally non-stationary (i.e., multi-talker babble) noise at 90°. In this experiment, no model errors such as microphone mismatch have been applied (i.e., $\upsilon_2 = 0$ dB). To analyze the impact of the approximation (5.6) of the unknown regularization term (5.4) in the FD-SG algorithms, the performance of the FD implementation of the optimal TD-SG algorithm (5.3)-(5.4), which uses the clean speech, is depicted too.

As indicated in Section 5.2, the approximation (5.6) in the FD-SG algorithm indeed results in a performance degradation with respect to the optimal FD-SG algorithm, especially for large $\frac{1}{\mu}$ and spectrally highly non-stationary noise. For stationary speech-weighted noise, the FD-SG algorithm does not suffer a lot from the approximation. For spectrally highly non-stationary noise, such as multi-talker babble, the limited averaging of approximation (5.6) in the FD implementation (cf. Section 5.2.2) does not suffice to preserve the large noise reduction performance of the optimal FD-SG algorithm. The performance of the FD-SG algorithm in highly non-stationary noise is significantly improved by applying a low pass filter (5.33) with $\lambda = 0.999$ to the regularization term. Alternatively, the performance of the FD-SG algorithm could be improved by decreasing the step size $\bar{\rho}$. Decreasing the step size, however, compromises the convergence speed.

Figure 5.2 demonstrates that the noise reduction performance of the FD-SG SDR-GSC in multi-talker babble noise is significantly improved by applying a low pass filter to the regularization term $\mathbf{R}_i[k]$. In Figure 5.2, the exponential weighting λ of the LPF was set to 0.999, which corresponds to an averaging over $\frac{1}{1-\lambda} = 1000$ frames (1 frame corresponds to 32 samples or 2 ms). In **Figure 5.3**, we study the influence of the forgetting factor λ on the performance of the SDR-GSC and the SP-SDW-MWF with $\mathbf{w}_0$ ($\frac{1}{\mu} = 0.5$) for the babble noise scenario. Figure 5.3 plots $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ obtained with the FD-SG with LPF implementation of the SDR-GSC and SP-SDW-MWF with $\bar{\mathbf{w}}_0$ as a function of $\frac{1}{1-\lambda}$. Performance clearly improves for increasing[13] λ, especially for λ up to 0.999. The SP-SDW-MWF with $\bar{\mathbf{w}}_0$ suffers from a larger excess error and hence, worse $\Delta\text{SNR}_{\text{intellig}}$ compared to the SDR-GSC. This is due to the larger dimensions of $\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$ and the relatively large contribution of the speech reference in $\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$.

The LPF reduces fluctuations in the filter coefficients caused by poor estimates of the speech correlation matrix $\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$ and/or the highly non-stationary *short-term* speech spectrum. Below, we analyze the influence of the LPF on the tracking performance of the FD-SG algorithm. As an illustration, **Figure 5.4 (a) and (b)** compare the tracking performance of the FD-SG algorithm with LPF ($\lambda = 0.999$) and the FD-SG algorithm without LPF (i.e., $\lambda = 0$), when

[13]Note that similar results have been obtained for the matrix-based MWF in Section 3.5.1 (cf. Figure 3.15), where it has been shown that the speech + noise correlation matrix and the noise correlation matrix have to be averaged over at least 0.5 s in order to obtain a good performance in spectrally non-stationary babble noise.

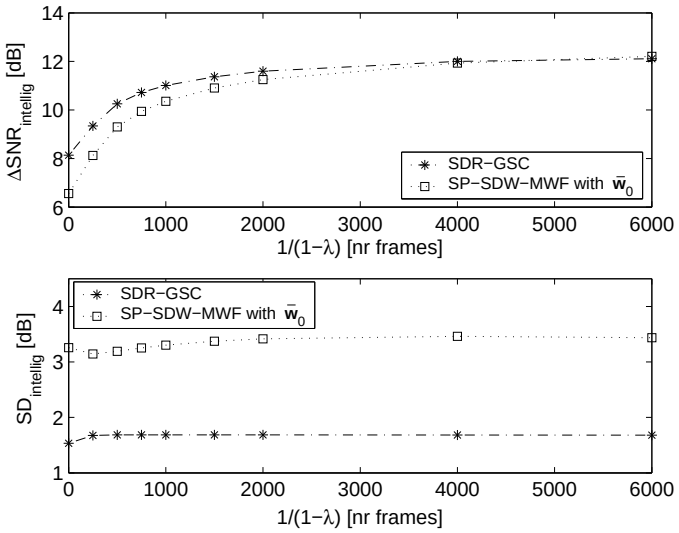
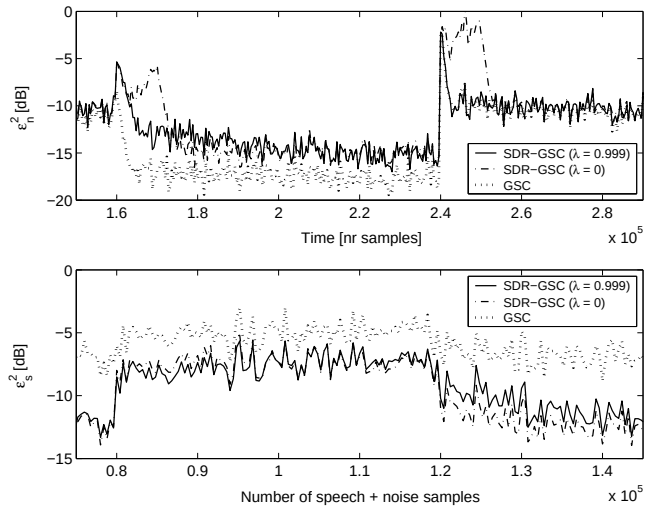


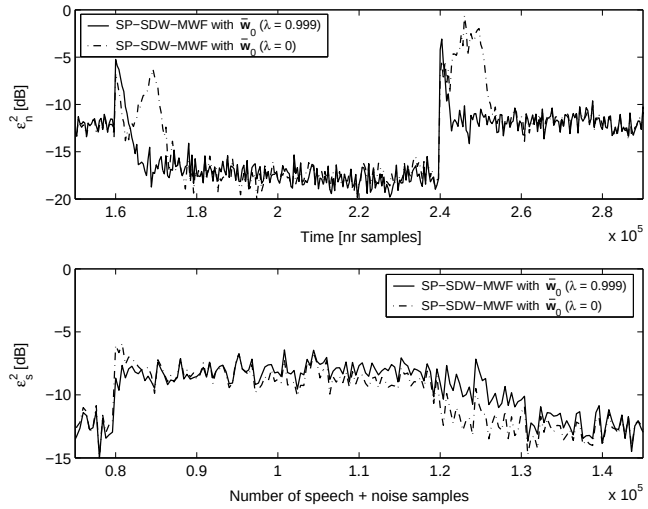
Figure 5.3: Performance of the FD-SG with LPF implementation of the SDR-GSC and the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ ($\frac{1}{\mu} = 0.5$) as a function of $1/(1-\lambda)$ where λ is the exponential weighting factor of the LPF. Babble noise at 90°.

the noise source position changes from 150° to 60° and from 60° to 150°. The changes in noise position occur during periods of noise-only at sample number 1.6×10^5 and sample number 2.4×10^5, respectively. Figure 5.4(a) depicts the results for the SDR-GSC (with $1/\mu = 0.5$), Figure 5.4(b) considers the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ (with $1/\mu = 0.5$). For comparison, the result of the GSC (i.e., the SDR-GSC with $1/\mu = 0$) is also shown in Figure 5.4(a). In the experiment, a gain mismatch v_2 of 4 dB was applied to the second microphone. To avoid fast changes in the residual noise energy ε_n^2 and speech distortion energy ε_s^2 due to highly non-stationary short-term speech and noise spectra, the desired and interfering source in this experiment are stationary, speech-weighted noise: half of the samples contains speech + noise, the others contain noise-only. The upper figures depict the residual noise energy ε_n^2 (defined in (5.1)) as a function of time (expressed in number of samples at $f_s = 16$ kHz), the lower figures plot the speech distortion energy ε_s^2 during speech + noise periods as a function of the number of speech + noise samples. The speech and noise energy ε_s^2 and ε_n^2 have been normalized with the speech and noise energy in the speech reference $y_0[k]$, respectively.

When a change in noise scenario occurs, the FD-SG with LPF implementation of the SDR-GSC and the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ suppress the noise with a convergence time similar to the FD-NLMS GSC, while the speech distortion ε_s^2 is smaller than for GSC, also in the transient phase after the change. In



(a)



(b)

Figure 5.4: Tracking performance of the FD-SG with LPF based SP-SDW-MWF (with $\lambda = 0.999$) compared to the GSC and the FD-SG implementation with $\lambda = 0$. The upper figures show the normalized residual noise energy ε_n^2 as a function of time. The lower figures plot the normalized speech distortion energy ε_s^2 as a function of the number of speech + noise samples. The noise source position changes from 150° to 60° at sample number 1.6×10^5 and from 60° to 150° at sample number 2.4×10^5. (a) SDR-GSC ($\frac{1}{\mu} = 0.5$) and GSC; (b) SP-SDW-MWF with $\bar{\mathbf{w}}_0$ ($\frac{1}{\mu} = 0.5$).

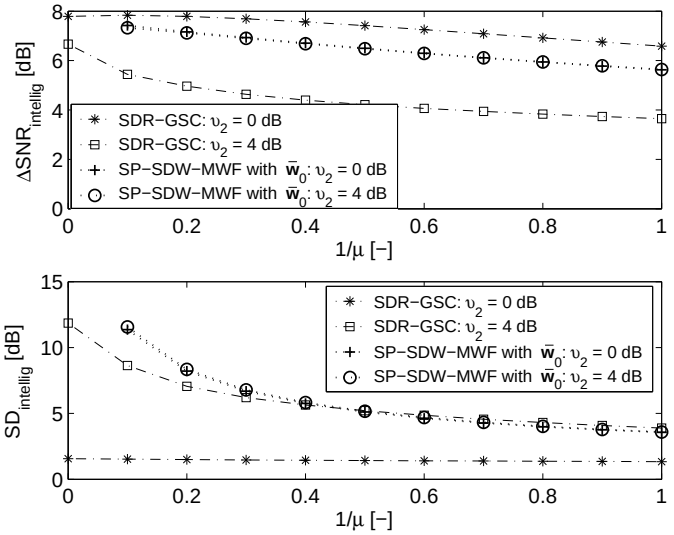


Figure 5.5: Performance of FD-SG SP-SDW-MWF with LPF ($\lambda = 0.999$) in a multiple noise source scenario.

addition, the LPF improves the noise reduction performance of the FD-SG algorithm in the transient phase. After the change, the instantaneous estimate (5.6) of the true regularization term (5.4) is poor: e.g., from sample number 2.4×10^5 until 2.5×10^5, only noise is observed so that the data in the speech + noise buffer are not immediately updated. As a result, the FD-SG implementation with $\lambda = 0$ suffers from a poor noise reduction performance. In the FD-SG algorithm with LPF, the regularization term is averaged over time by the LPF, which reduces the negative impact of inaccurate estimates of the instantaneous speech correlation matrix on the regularization term and hence, on the performance.

Remark 5.5 Because of the LPF, a wrong speech correlation estimate at time instant k has an influence on future regularization terms that decays only exponentially. This explains the further, slight decrease in the residual noise energy ε_n^2 in Figure 5.4(a) from sample number 1.7×10^5 until 1.9×10^5 after the initial convergence. □

5.5.3 Comparison with FD-NLMS based SPA

The above experiments demonstrate that the FD-SG algorithm with LPF outperforms the FD-SG algorithm and offers a good tracking performance. In this section, we investigate whether this FD-SG with LPF implementation of

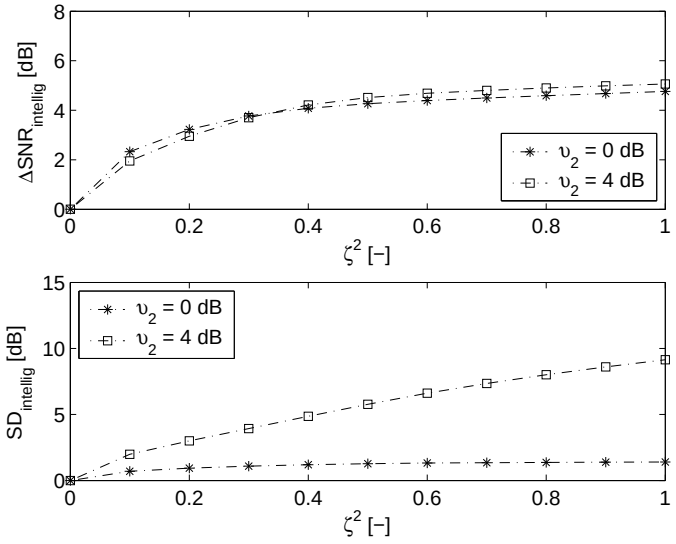


Figure 5.6: Performance of FD-NLMS based SPA in a multiple noise source scenario.

the SP-SDW-MWF preserves the benefit of the exact SP-SDW-MWF over the QIC-GSC by comparing it with the FD-NLMS based SPA. For a fair comparison, the FD-NLMS based SPA is, like the FD-SG algorithm, also adapted during speech + noise periods using data from a noise buffer.

Figure 5.5 and **Figure 5.6** compare the performance of the FD-SG SP-SDW-MWF with LPF ($\lambda = 0.999$) and the FD-NLMS based SPA (5.44)-(5.45) in a multiple noise source scenario. The noise scenario consists of 3 multi-talker babble noise sources positioned at angles 75°, 180°, 240° with respect to the desired source at 0°. To assess the sensitivity of the algorithms against errors in the assumed signal model, the influence of a gain mismatch $\upsilon_2 = 4$ dB of the second microphone on the performance is depicted too. In **Figure 5.5**, the improvement $\Delta\text{SNR}_{\text{intellig}}$ and the distortion $\text{SD}_{\text{intellig}}$ of the SDR-GSC and SP-SDW-MWF with $\bar{\mathbf{w}}_0$ is depicted as a function of the trade-off factor $\frac{1}{\mu}$. **Figure 5.6** shows the results of the FD-NLMS based SPA of (5.44)-(5.45) for different QIC values ζ^2. The FD-NLMS based GSC corresponds to the FD-SG SDR-GSC with $\frac{1}{\mu} = 0$ or the FD-NLMS based SPA with $\zeta^2 = \infty$.

In this noise scenario, the GSC still offers good noise suppression for a mismatch of 4 dB, at the expense of a large distortion. Both the SPA and the stochastic gradient based SP-SDW-MWF increase the robustness of the GSC: distortion decreases with increasing $\frac{1}{\mu}$ and decreasing ζ^2. The SPA is more conservative than the SDR-GSC: the constraint value ζ^2 should be chosen so

that the maximum allowable speech distortion is not exceeded for the largest model error, e.g., 5 dB $\text{SD}_{\text{intellig}}$ for a gain mismatch up to 4 dB. This goes at the expense of smaller noise reduction in case of smaller model errors (e.g., $\Delta\text{SNR}_{\text{intellig}} = 4\,\text{dB}$ for $\zeta^2 = 0.4$). The SDR-GSC on the other hand only puts emphasis on speech distortion when truly needed, i.e., when the amount of speech leakage is large, so that a better noise reduction is obtained for small model errors (e.g., $\Delta\text{SNR}_{\text{intellig}}$ between 4 dB and 7.4 dB for $\frac{1}{\mu} = 0.5$). The SP-SDW-MWF offers more noise suppression at even larger model errors: the performance of the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ is less affected by microphone mismatch than the performance of the SDR-GSC. In the absence of model errors, the SP-SDW-MWF with $\bar{\mathbf{w}}_0$ achieves a slightly worse performance than the SDR-GSC. With $\bar{\mathbf{w}}_0$, the estimate of $\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$ is less accurate due to the larger dimensions of $\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$ and the large contribution of the speech reference in $\frac{1}{\mu}\mathcal{E}\{\mathbf{y}^s\mathbf{y}^{s,H}\}$ (see also Figure 5.3).

In short, the proposed FD-SG with LPF implementation of the SP-SDW-MWF preserves the benefit of the exact SP-SDW-MWF over the QIC-GSC. Preliminary experiments in [41] indicate only limited performance differences (i.e., smaller than 0.5 dB) between the FD-SG using CM and the FD-SG with LPF, demonstrating that also the FD-SG using CM preserves this benefit.

5.6 Conclusions

In this chapter, we proposed a frequency-domain stochastic gradient implementation for the SP-SDW-MWF and compared its performance to the NLMS based SPA [29] for QIC-GSC. Starting from the cost function of the SP-SDW-MWF, an TD-SG algorithm has been derived in Section 5.2. To speed-up convergence and reduce computational complexity, the algorithm has been implemented in the frequency-domain (FD-SG).

Experiments in Section 5.5 indicated that the FD-SG algorithm suffers from a reasonably large excess error when the noise is spectrally non-stationary. In Section 5.3 and Section 5.5, we have demonstrated that for spatially slowly time-varying noise scenarios, the excess error of the proposed FD-SG algorithm is reduced by low pass filtering the part of the gradient estimate that limits speech distortion so that spectrally highly non-stationary noise can still be successfully suppressed. The low pass filtering reduces fluctuations in the filter coefficients caused by poor estimates of the short-term speech correlation matrix and/or by the highly non-stationary short-term speech spectrum, while providing good (i.e., similar to GSC) tracking of a change in the noise scenario. The FD-SG with LPF algorithm requires data buffers, which results in a too large memory cost. In Section 5.3, we have shown that by approximating the regularization term using (diagonal) correlation matrices, the memory cost of

the FD-SG algorithm with LPF can be decreased so that it fits within the memory constraints of commercial hearing aids.

Experiments in Section 5.5 demonstrated that the proposed FD-SG preserves the benefit of the SP-SDW-MWF over QIC-GSC, while its computational complexity is comparable to the NLMS based SPA. Preliminary experiments in [41] indicated that the FD-SG using correlation matrices has about the same performance as the FD-SG algorithm with LPF, making it a good alternative to the SPA for implementation in commercial hearing aids.

5.A Appendix: Approximate stochastic gradient algorithm

In [57], a stochastic gradient based SDW-MWF using speech + noise and noise buffers has been proposed, which, like the MWF technique of [36], provides an estimate of the speech component in one of the received microphone signals. This stochastic gradient SDW-MWF can be modified so that it applies to the structure of the SP-SDW-MWF. Besides the approximation (5.6) of the true regularization term (5.4), the stochastic gradient implementation of [57] relies on some additional, independence assumptions. Applying these independence assumptions to the TD-SG algorithm and FD-SG algorithm proposed in Section 5.2, results in an alternative stochastic gradient implementation of the SP-SDW-MWF similar to [57]. Because of the additional assumptions, we will in the sequel refer to this technique as approximate stochastic gradient (ASG) algorithm. The derivation of the ASG algorithm from the TD-SG algorithm is done in the time-domain. In Section 5.A.2, the resulting TD-ASG implementation is then translated to a FD-ASG algorithm. In Section 5.A.3, we evaluate the impact of the independence assumptions on the performance: it is shown that the FD-ASG algorithm results in a seriously degraded performance with respect to the FD-SG algorithm and the FD-SG algorithm with LPF, while not significantly reducing the computational complexity.

5.A.1 TD-ASG algorithm

In the derivation of the stochastic gradient based SDW-MWF proposed in [57], the authors assume that besides speech and noise, also the noise component $\mathbf{y}^n_{B_y}[k]$ in the speech + noise buffer $\mathbf{B}_y$ and the noise $\mathbf{y}^{n,H}[k]$ received at time instant k during a noise-only period are statistically independent, i.e.,

$$\mathcal{E}\{\mathbf{y}_{B_y}[k]\mathbf{y}^{n,H}[k]\} = \mathcal{E}\{\mathbf{y}^s_{B_y}[k]\mathbf{y}^{n,H}[k]\} + \mathcal{E}\{\mathbf{y}^n_{B_y}[k]\mathbf{y}^{n,H}[k]\} = 0, \qquad (5.A.1)$$

and, similarly, that during speech + noise periods, $\mathcal{E}\{\mathbf{y}_{B^n_y}[k]\mathbf{y}^H[k]\} = 0$. Applying these independence assumptions to the instantaneous gradient estimate in the TD-SG update equation (5.3) with (5.6), corresponds to

$$\sqrt{\frac{1}{\mu-1}}\mathbf{y}_{B_y}[k]y_0^{n,*}[k-\Delta] = 0, \qquad (5.A.2)$$

$$\sqrt{\frac{1}{\mu}\left(1-\frac{1}{\mu}\right)}\left(\mathbf{y}^n[k]\mathbf{y}^H_{B_y}[k] + \mathbf{y}_{B_y}[k]\mathbf{y}^{n,H}[k]\right) = 0, \qquad (5.A.3)$$

and results in an alternative stochastic gradient algorithm for the SP-SDW-MWF that is similar to [57]:

$$\boxed{\bar{\mathbf{w}}[k] = \bar{\mathbf{w}}[k-1] + \frac{\bar{\rho}}{\mathbf{u}^H[k]\mathbf{u}[k]+c}\mathbf{u}[k](d^*[k] - \mathbf{u}^H[k]\bar{\mathbf{w}}[k-1])} \qquad (5.A.4)$$

where the step size $\bar{\rho} \in [0, 2)$ and where

$$d[k] = y_0^n[k-\Delta]\frac{1}{\sqrt{1-\frac{1}{\mu}}}; \qquad \mathbf{u}[k] = \sqrt{1-\tfrac{1}{\mu}}\mathbf{y}^n[k] + \sqrt{\tfrac{1}{\mu}}\mathbf{y}_{B_y}[k] \quad (5.A.5)$$

during *periods of noise-only* (i.e., $\mathbf{y}[k] = \mathbf{y}^n[k]$). During *periods of speech + noise* (i.e., $\mathbf{y}[k] = \mathbf{y}^s[k] + \mathbf{y}^n[k]$), $d[k]$ and $\mathbf{u}[k]$ in (5.A.4) are set to

$$d[k] = y_{B_y^n,0}[k-\Delta]\frac{1}{\sqrt{1-\frac{1}{\mu}}}; \quad \mathbf{u}[k] = \sqrt{1-\frac{1}{\mu}}\mathbf{y}_{B_y^n}[k] + \sqrt{\frac{1}{\mu}}\mathbf{y}[k]. \quad (5.A.6)$$

The TD-ASG algorithm is summarized in the tabulated **Algorithm 5.4**.

5.A.2 FD-ASG algorithm

To reduce complexity and speed up convergence, the ASG algorithm is, like the SG algorithm of Section 5.2, implemented in the frequency-domain:

$$\boxed{\begin{aligned} \bar{\mathbf{W}}_l[k] &= \bar{\mathbf{W}}_l[k-1] + \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\boldsymbol{\rho}[k]\mathbf{U}_l^H[k]\mathbf{E}[k], \\ \mathbf{E}[k] &= \mathbf{F}\mathbf{k}^T\left(\mathbf{d}[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{U}_j[k]\bar{\mathbf{W}}_j[k-1]\right), \end{aligned}} \quad (5.A.7)$$

where $\mathbf{U}_l[k]$, $l = M-N, \cdots M-1$ is derived from $\mathbf{Y}_l^n[k]$ and $\mathbf{Y}_l[k]$ (cf. (5.17)-(5.18) in Section 5.2.2) as:

$$\mathbf{U}_l[k] = \sqrt{1-\frac{1}{\mu}}\mathbf{Y}_l^n[k] + \sqrt{\frac{1}{\mu}}\mathbf{Y}_l[k],\ l = M-N,\ ...,\ M-1, \quad (5.A.8)$$

and $\mathbf{d}[k]$ equals

$$\mathbf{d}[k] = \frac{1}{\sqrt{1-\frac{1}{\mu}}}\left[\ y_0[kL-\Delta] \quad \cdots \quad y_0[kL-\Delta+L-1]\ \right]^T, \quad (5.A.9)$$

during periods of noise, while during periods of speech + noise, $\mathbf{d}[k]$ is constructed from noise data $y_{\tilde{B}_{y_0}^n}[k]$ in the buffer $\tilde{\mathbf{B}}_{y_0}^n$ as

$$\mathbf{d}[k] = \frac{1}{\sqrt{1-\frac{1}{\mu}}}\left[\ y_{\tilde{B}_{y_0}^n}[kL-\Delta] \quad \cdots \quad y_{\tilde{B}_{y_0}^n}[kL-\Delta+L-1]\ \right]^T. \quad (5.A.10)$$

The step size matrix $\boldsymbol{\rho}[k]$ is set to

$$\boldsymbol{\rho}[k] = \frac{2\bar{\rho}}{L}\mathrm{diag}\{P_0^{-1}[k], \cdots, P_{2L-1}^{-1}[k]\}, \quad (5.A.11)$$

Algorithm 5.4 TD-ASG algorithm.

Initialization and matrix definitions:

$$\bar{\mathbf{w}}[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{C}^{NL\times 1}$$

Speech+noise buffer $\mathbf{B}_y \in \mathbb{C}^{NL\times L_{B_y}}$; Noise buffer $\mathbf{B}_y^n \in \mathbb{C}^{ML\times L_{B_y^n}}$

For each new input sample (per channel):

If noise detected:

$$\mathbf{y}^n[k] = \mathbf{y}[k] = \begin{bmatrix} \mathbf{y}_{M-N}^H[k] & \mathbf{y}_{M-N+1}^H[k] & \cdots & \mathbf{y}_{M-1}^H\ [k] \end{bmatrix}^H$$

$\mathbf{y}_{B_y}[k] = NL$-dimensional speech + noise vector from the buffer $\mathbf{B}_y$

Store input vectors $\mathbf{y}_l[k], l = 0, \ldots, M-1$ in noise buffer $\mathbf{B}_y^n$

$$d[k] = y_0^n[k-\Delta]\frac{1}{\sqrt{1-\frac{1}{\mu}}}; \quad \mathbf{u}[k] = \sqrt{1-\frac{1}{\mu}}\mathbf{y}^n[k] + \sqrt{\frac{1}{\mu}}\mathbf{y}_{B_y}[k]$$

If speech detected:

$$\mathbf{y}[k] = \begin{bmatrix} \mathbf{y}_{M-N}^H[k] & \mathbf{y}_{M-N+1}^H[k] & \cdots & \mathbf{y}_{M-1}^H\ [k] \end{bmatrix}^H$$

$\mathbf{y}_{B_y^n}[k] = NL$-dimensional noise vector from the buffer $\mathbf{B}_y^n$

Store input vectors $\mathbf{y}_l[k], l = M-N, \ldots, M-1$ in speech + noise buffer $\mathbf{B}_y$

$$d[k] = y_{B_y^n,0}[k-\Delta]\frac{1}{\sqrt{1-\frac{1}{\mu}}}; \quad \mathbf{u}[k] = \sqrt{1-\frac{1}{\mu}}\mathbf{y}_{B_y^n}[k] + \sqrt{\frac{1}{\mu}}\mathbf{y}[k]$$

Update formula:

$$\bar{\mathbf{w}}[k] = \bar{\mathbf{w}}[k-1] + \rho\mathbf{u}[k](d^*[k] - \mathbf{u}^H[k]\bar{\mathbf{w}}[k-1])$$

with:

$$\rho = \frac{\bar{\rho}}{\mathbf{u}^H[k]\mathbf{u}[k] + c}$$

Output signal $z[k]$: $z[k] = y_0[k-\Delta] - \bar{\mathbf{w}}^H[k-1]\mathbf{y}[k]$

Table 5.3: Computational complexity of the TD-ASG and FD-ASG (expressed as the number of operations per sample).

Algorithm	**update formula**	**adaptation of step size**
TD-ASG	$(8NL+3)$	$(2N+4)$
FD-ASG	$(28N+1)+(12N+12)\log_2 2L$	$(4N+3)$

$$P_m[k] \quad = \quad \gamma P_m[k-1] + (1-\gamma) \sum_{l=M-N}^{M-1} |U_{l,m}[k]|^2 \,. \qquad (5.A.12)$$

The output $\mathbf{z}[k]$ of the FD-ASG algorithm is computed in the same way as the output $\mathbf{z}[k]$ of the FD-SG algorithm with LPF (cf. Algorithm 5.2). The FD-ASG algorithm is summarized in the tabulated **Algorithm 5.5**.

In Section 5.A.3, the performance of the FD-SG algorithm is compared with the FD-ASG algorithm. The TD-ASG and FD-ASG algorithm assume that besides speech and noise, also noise samples at different time instants are statistically independent. In practice, (5.A.2) and (5.A.3) do not hold, especially for large $\sqrt{\frac{1}{\mu-1}}$ and/or $\sqrt{\frac{1}{\mu}\left(1-\frac{1}{\mu}\right)}$, i.e., for $\mu \to 1$. Hence, compared to the proposed FD-SG algorithm, performance is expected to be worse. In contrast to the FD-SG algorithm, the FD-ASG algorithm can not be applied for $\mu \leq 1$. It should also be noted that the improvements that are proposed in Section 5.3 to reduce the excess error of the FD-SG algorithm, are not applicable to the FD-ASG algorithm. Finally, compared to the FD-SG algorithm no significant complexity reduction is achieved by the FD-ASG algorithm: as indicated in **Table 5.3**, the FD-ASG algorithm only requires about $4\log_2 2L + 8N$ operations less per sample than the FD-SG algorithm (cf. Table 5.2 in Section 5.4).

5.A.3 Influence of the independence assumptions on the performance

In **Figure 5.A.7 (a) and (b)**, we compare the performance in terms of $\Delta\text{SNR}_{\text{intellig}}$ and $\text{SD}_{\text{intellig}}$ of the FD-ASG with the FD-SG and the FD-SG with LPF (with exponential weighting factor $\lambda = 0.999$) implementation of the SDR-GSC for different values of the trade-off parameter $1/\mu$. Figure 5.A.7(a) depicts the performance for a stationary (i.e., speech-weighted noise) noise source at 90°, while Figure 5.A.7(b) plots the performance measures for spectrally non-stationary (i.e., multi-talker babble) noise at 90°. For both noise scenarios, the FD-SG algorithm and the FD-SG algorithm with LPF significantly outperform the FD-ASG algorithm, especially for $\frac{1}{\mu} \to 1$, demonstrating that the independence assumptions made in the FD-ASG have a significant impact on the performance.

Algorithm 5.5 FD-ASG algorithm (based on overlap-save).

Initialization and matrix definitions:

$\bar{\mathbf{W}}_l[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T$, $l = M-N, ..., M-1$; $P_m[0] = c_m$, $m = 0, ..., 2L-1$

Speech+noise buffer $\tilde{\mathbf{B}}_y \in \mathbb{C}^{NL \times L_{\tilde{B}_y}}$; Noise buffers $\tilde{\mathbf{B}}_y^n \in \mathbb{C}^{NL \times L_{\tilde{B}_y^n}}$, $\tilde{\mathbf{B}}_{y_0}^n \in \mathbb{R}^{L \times L_{\tilde{B}_y^n}}$

For each new block of L input samples (per channel):

If noise detected:

$$\mathbf{y}_0[k] = \begin{bmatrix} y_0[kL-\Delta] & \cdots & y_0[kL-\Delta+L-1] \end{bmatrix}^T ; \ \mathbf{d}[k] = \frac{1}{\sqrt{1-\frac{1}{\mu}}}\mathbf{y}_0[k]$$

$$\mathbf{Y}_l^n[k] = \text{diag}\left\{\mathbf{F}\begin{bmatrix} y_l[kL-L] & \cdots & y_l[kL+L-1] \end{bmatrix}^T\right\}$$

Store input data $\mathbf{Y}_l^n[k]$, $\mathbf{y}_0[k]$ in noise buffers $\tilde{\mathbf{B}}_y^n$, $\tilde{\mathbf{B}}_{y_0}^n$

Create $\mathbf{Y}_l[k]$ from data in speech+noise buffer $\tilde{\mathbf{B}}_y$

If speech detected:

$$\mathbf{Y}_l[k] = \text{diag}\left\{\mathbf{F}\begin{bmatrix} y_l[kL-L] & \cdots & y_l[kL+L-1] \end{bmatrix}^T\right\}$$

Store input data $\mathbf{Y}_l[k]$ in speech + noise buffer $\tilde{\mathbf{B}}_y$

Create $\mathbf{Y}_l^n[k]$, $\mathbf{d}[k]$ from data in noise buffers $\tilde{\mathbf{B}}_y^n$, $\tilde{\mathbf{B}}_{y_0}^n$

$$\mathbf{d}[k] = \frac{1}{\sqrt{1-\frac{1}{\mu}}}\begin{bmatrix} y_{\tilde{B}_{y_0}^n}[kL-\Delta] & \cdots & y_{\tilde{B}_{y_0}^n}[kL-\Delta+L-1] \end{bmatrix}^T$$

Update formula:

$$\bar{\mathbf{W}}_l[k] = \bar{\mathbf{W}}_l[k-1] + \mathbf{F}\mathbf{g}\mathbf{F}^{-1}\boldsymbol{\rho}[k]\mathbf{U}_l^H[k]\mathbf{E}[k]$$

with

$$\mathbf{U}_l[k] = \sqrt{1-\frac{1}{\mu}}\mathbf{Y}_l^n[k] + \sqrt{\frac{1}{\mu}}\mathbf{Y}_l[k]$$

$$\mathbf{E}[k] = \mathbf{F}\mathbf{k}^T\left(\mathbf{d}[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{U}_j[k]\bar{\mathbf{W}}_j[k-1]\right)$$

and

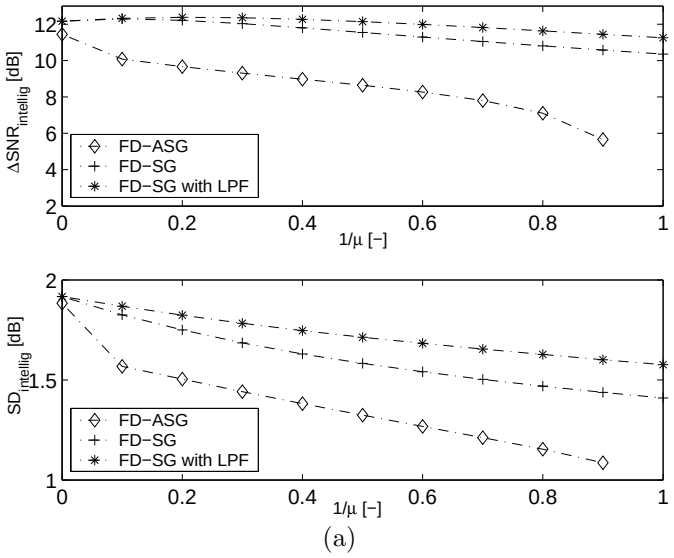
$$\boldsymbol{\rho}[k] = \frac{2\bar{\rho}}{L}\text{diag}\{P_0^{-1}[k], \cdots, P_{2L-1}^{-1}[k]\}$$

$$P_m[k] = \gamma P_m[k-1] + (1-\gamma)\sum_{l=M-N}^{M-1}|U_{l,m}[k]|^2$$

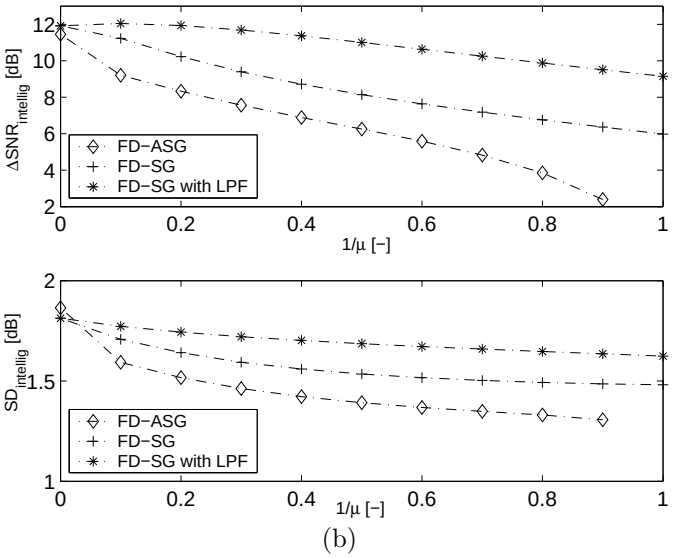
Output $\mathbf{z}[k] = \begin{bmatrix} z[kL] & \cdots & z[kL+L-1] \end{bmatrix}^T$:

$\mathbf{y}_0[k] = \begin{bmatrix} y_0[kL-\Delta] & \cdots & y_0[kL-\Delta+L-1] \end{bmatrix}^T$

- If noise-only detected: $\mathbf{z}[k] = \mathbf{y}_0[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j^n[k]\bar{\mathbf{W}}_j[k-1]$
- If speech + noise detected: $\mathbf{z}[k] = \mathbf{y}_0[k] - \mathbf{k}\mathbf{F}^{-1}\sum_{j=M-N}^{M-1}\mathbf{Y}_j[k]\bar{\mathbf{W}}_j[k-1]$



(a)



(b)

Figure 5.A.7: Performance of the FD-ASG compared to the performance of the FD-SG and the FD-SG with LPF for different values of $1/\mu$. (a) Stationary speech-weighted noise at 90°; (b) Multi-talker babble noise at 90°.

Chapter 6

Adaptive feedback cancellation techniques

ACOUSTIC feedback limits the maximum amplification that can be used in a hearing aid without making it unstable. In many cases this maximum amplification is too small to compensate for the hearing loss. Therefore, feedback suppression or cancellation algorithms are called for.

In this chapter, we give an overview of the existing techniques for feedback suppression and, in particular, (adaptive) feedback cancellation. We show that because of the presence of a closed signal loop, the standard adaptive filtering techniques for open-loop systems (e.g., echo cancellation) fail to provide a reliable feedback path estimate if the desired signal is spectrally colored. A more promising approach for adaptive feedback cancellation is to adopt the identification procedures for closed-loop systems [59, 60]. In this chapter, we provide the basic concepts of closed-loop system identification and discuss some recent applications to feedback cancellation, which form the basis for the adaptive feedback cancellation techniques that will be developed in Chapter 7.

Section 6.1 gives a mathematical formulation of the acoustic feedback problem in hearing aids and briefly describes the two possible approaches to reduce its negative effects, i.e., feedforward suppression and feedback cancellation. In addition, the performance measures for feedback cancellation are defined.

Section 6.2 discusses the standard continuous adaptation feedback cancellation (CAF) algorithm that is widely studied for application in hearing aids. We show that this standard feedback cancellation technique suffers from a model error or bias when the desired signal is spectrally colored.

In Section 6.3 and Section 6.4, existing solutions to reduce the bias of the standard CAF are discussed. A classical approach is to incorporate additional signal processing operations in the standard CAF that try to limit the bias to a certain extent (cf. Section 6.3). A more promising solution is to regard the feedback path as a part of a closed-loop system and to apply closed-loop system identification theory [59, 60] (cf. Section 6.4). This closed-loop identification approach has recently been adopted for feedback cancellation in [87, 88, 89, 217] and is further investigated in Chapter 7, with the aim of obtaining an unbiased adaptive feedback canceller.

6.1 Problem statement

In Section 6.1.1, we give a mathematical formulation of the acoustic feedback problem in hearing aids and define the notation. Section 6.1.2 and Section 6.1.3 describe the two approaches that can be followed to reduce acoustic feedback, i.e., feedforward suppression and feedback cancellation. In Section 6.1.4, we define the performance measures for feedback cancellation techniques.

6.1.1 Acoustic feedback in hearing aids

Notation

In the sequel, the following notation is used. The symbol q^{-1} denotes the discrete-time delay operator, i.e.,

$$q^{-1}u[k] = u[k-1]. \tag{6.1}$$

A discrete-time filter with coefficient vector $\mathbf{f}$

$$\mathbf{f} = \begin{bmatrix} f_0 & f_1 & \cdots & f_{L_F-1} \end{bmatrix}^T \tag{6.2}$$

and filter length L_F is represented as a polynomial $F(q)$ in q, i.e.,

$$F(q) = f_0 + f_1 q^{-1} + f_2 q^{-2} + \cdots + f_{L_F-1} q^{-L_F+1}. \tag{6.3}$$

The filtering of the signal $u[k]$ with $F(q)$ is denoted as

$$F(q)u[k] = \mathbf{f}^T \mathbf{u}[k], \tag{6.4}$$

with $\mathbf{u}[k] = \begin{bmatrix} u[k] & u[k-1] & \cdots & u[k-L_F+1] \end{bmatrix}^T$. The filter $F(q,k)$ refers to a time-varying filter with coefficient vector $\mathbf{f}[k]$.

Acoustic feedback

Figure 6.1 illustrates the problem of acoustic feedback for a hearing aid with a single microphone. The so-called *forward path* $G(q)$ represents the regular

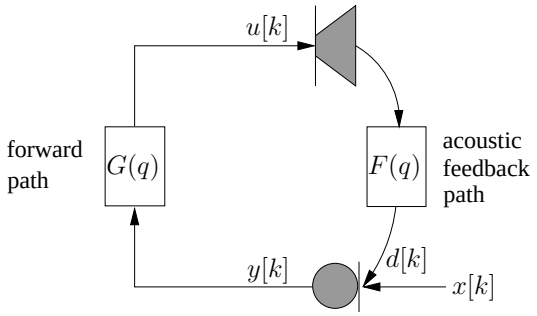


Figure 6.1: Acoustic feedback.

signal processing path of the hearing aid. We assume that $G(q)$ has a delay of at least one sample. In general, the forward path $G(q)$ consists of a frequency specific gain and compression. In the most recent hearing aids, it may also include a noise reduction algorithm. The *feedback path* between the loudspeaker and the microphone is denoted by $F(q)$. It includes the frequency characteristic of the loudspeaker and the microphone, the acoustic feedback path and the characteristics of the digital-to-analog (DAC) and analog-to-digital (ADC) converters. The desired signal is denoted by $x[k]$ and the feedback signal is denoted by $d[k]$. The loudspeaker and microphone signal are $u[k]$ and $y[k]$, respectively.

Ideally, the loudspeaker signal $u[k]$ equals the desired signal $x[k]$ processed by the forward path $G(q)$, i.e.,

$$u[k] = G(q)x[k]. \tag{6.5}$$

However, due to acoustical coupling $F(q)$ between the loudspeaker and the microphone, the loudspeaker signal $u[k]$ is fed-back into the microphone as

$$\begin{aligned} y[k] &= d[k] + x[k] \\ &= F(q)u[k] + x[k], \end{aligned} \tag{6.6}$$

resulting in a closed-loop system $C(q)$ from the desired signal $x[k]$ to the loudspeaker signal $u[k]$

$$C(q) = \frac{G(q)}{1 - G(q)F(q)}. \tag{6.7}$$

The closed-loop system $C(q)$ is stable iff (if and only if) the open-loop gain $|G(\omega)F(\omega)| < 1$ at all angular frequencies $\omega \in [0, \pi]$ with positive feedback, i.e., $\angle\left(G(\omega)F(\omega)\right) = n2\pi$ with $n \in \mathbb{Z}$. If the gain $|G(\omega)|$ is increased beyond this point, the system starts to oscillate. At low gain margins $-20\log_{10}|G(\omega)F(\omega)|$, acoustic feedback degrades the sound quality by introducing ringing or howling. To avoid significant distortion, a gain margin of at least 6 dB is advisable: the

effect of the acoustic feedback on the closed-loop frequency response $C(\omega)$ is then limited to [90]

$$\frac{2}{3}G(\omega) \leq C(\omega) \leq 2G(\omega). \tag{6.8}$$

Remark 6.1 Often, the open-loop response $G(q)F(q)$ of a hearing aid contains a group delay, which, e.g., originates from the ADC, the DAC and the signal processing algorithm in the forward path $G(q)$. As a result, the phase $\angle(G(\omega)F(\omega))$ frequently passes through a multiple of 2π and hence, multiple critical frequencies exist. Generally, the worst case is therefore assumed for the phase and only the gain margin is considered as a criterion for stability. □

The minimum attenuation of a typical feedback path lies between 20 and 50 dB, depending on the vent size and the type of hearing aid (ITC, ITE, BTE). In the latest hearing aids that use open fittings (e.g., ReSoundAir^{TM}), this minimum attenuation is even smaller. Placing a telephone handset close to the ear can temporarily reduce the feedback path attenuation by another 10 to 20 dB [91, 205, 246]. Hence, the maximum amplification that can be used in a hearing aid without compromising sound quality is limited to at most 44 dB, but often even less. Given that people with a moderate hearing loss suffer from an increase in hearing threshold of 40 to 70 dB compared to those with normal hearing, this maximum amplification is too small to compensate for the hearing loss in a patient.

6.1.2 Feedforward suppression

In the literature, several techniques have been developed to reduce the negative effects of acoustic feedback. They can be classified into feedforward suppression (cf. Section 6.1.2) and feedback cancellation techniques (cf. Section 6.1.3) [87].

Feedforward suppression techniques modify the forward path $G(q)$ of the hearing aid in such a way that it is stable in conjunction with the feedback path $F(q)$.

Notch filtering

The most common feedforward suppression technique is notch filtering: frequencies where instability is about to occur are attenuated by one or more narrow band notch filters [99, 166, 202]. The critical frequencies are determined beforehand or detected adaptively [99, 122, 269] when oscillation occurs. This technique can be useful to increase the maximum gain at other frequencies than the critical frequencies. The closed-loop gain $|C(q)|$ will, however, still be limited and the frequency response of the hearing aid can be severely compromised.

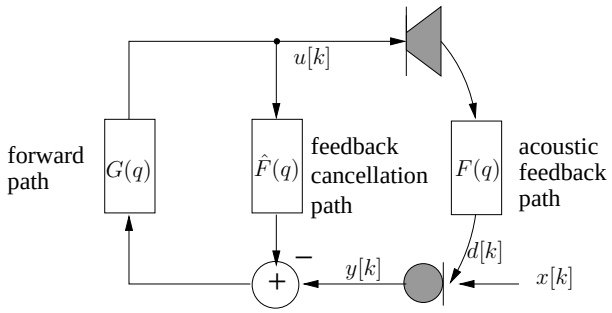


Figure 6.2: Feedback cancellation.

Time-varying elements in the forward path

In [21, 184, 216], instability is avoided by inserting time-varying elements in the forward path $G(q)$ such as delay modulation [21, 184], phase modulation [184] and frequency shifting [184, 216]. In this way, the build-up of feedback resonances is disrupted. Among the different time-varying elements, especially frequency shifting has been found successful in increasing the maximum stable gain (i.e., an increase of about 12 dB) [184, 216]. The improvement with delay and phase modulation, on the other hand, was found to be limited to 2 dB and 4 dB, respectively [21, 184]. The main disadvantages of time-varying elements are their relatively high cost and, in particular, the degradation of the sound quality [184]. While a small amount of modulation or a small frequency shift (e.g., a frequency shift up to 5 Hz [216]) may still be acceptable for speech signals, its effect on music is objectionable.

Equalizing of the phase of the open-loop response

In [84, 268], instability is avoided by modifying the phase of the open-loop response $G(q)F(q)$ such that it never reaches a multiple of 2π. However, because the transfer function of the feedback path $F(q)$ often contains some right-half plane zeros, complete cancellation of the phase of the open-loop response is impossible. In addition, the phase equalization technique is fixed and, hence, cannot respond to time-variations in the phase of the acoustic feedback path.

6.1.3 Feedback cancellation

A more promising solution for acoustic feedback is the use of a *feedback cancellation* technique, as illustrated in **Figure 6.2**. The feedback canceller $\hat{F}(q)$ estimates the feedback signal $d[k] = F(q)u[k]$ and subtracts it from the microphone signal $y[k]$. As a result, the closed-loop transfer function $C(q)$ from the

desired signal $x[k]$ to the loudspeaker input $u[k]$ is modified into

$$C(q) = \frac{G(q)}{1 - G(q)\left(F(q) - \hat{F}(q)\right)}. \tag{6.9}$$

Ideally, $\hat{F}(q) = F(q)$ so that the closed-loop system $C(q)$ equals the desired hearing aid response $G(q)$.

In [204, 275], a fixed feedback canceller $\hat{F}(q)$ is used. However, the acoustic path between the loudspeaker and the microphone is highly user-dependent. Furthermore, it can vary significantly depending on the acoustics outside the ear, e.g., because of jaw movements, hugging, a telephone set at the ear, room reverberation [91, 128, 129, 205], so that adaptive feedback cancellers $\hat{F}(q)$ are called for. Existing techniques for adaptive feedback cancellation are discussed in Section 6.2 and Section 6.3.

6.1.4 Performance measures for feedback cancellation

To assess the performance of a feedback cancellation algorithm, different measures can be used. Among them, the misalignment between the estimated and true feedback path, the maximum or added stable gain and the deviation of the closed-loop response $C(q)$ (cf. (6.9)) from the desired response $G(q)$.

Misalignment

The *misalignment* $\zeta\left(\hat{F}(q), F(q)\right)$ between the feedback path estimate $\hat{F}(q)$ and the true feedback path $F(q)$ is computed in the frequency domain as

$$\zeta\left(\hat{F}(q), F(q)\right) = \sqrt{\frac{\int_0^\pi \left|F(\omega) - \hat{F}(\omega)\right|^2 d\omega}{\int_0^\pi |F(\omega)|^2 \, d\omega}}. \tag{6.10}$$

The misalignment reflects the accuracy of the feedback path estimate.

Maximum stable gain (MSG) and added stable gain (ASG)

To get an idea of the benefit of a feedback cancellation algorithm for the hearing aid users, the *maximum stable gain* is often used. In [128], the maximum stable gain MSG is defined as the maximum gain without instability assuming a flat frequency response $|G(\omega)|$ and the worst case for the phase:

$$\mathrm{MSG} = 20 \log_{10} \left(\min_{\omega} \frac{1}{|F(\omega) - \hat{F}(\omega)|} \right). \tag{6.11}$$

The MSG is determined by the frequency where the mismatch between the feedback path estimate and the true feedback path is the largest. The system will however only be unstable when the phase at that frequency equals $n2\pi$, $n \in \mathbb{Z}$ [166]. The *added stable gain* ASG is defined as the additional gain that is possible by using the feedback canceller $\hat{F}(q)$, i.e.,

$$\text{ASG=MSG-}20\log_{10}\left(\min_{\omega}\frac{1}{|F(\omega)|}\right). \tag{6.12}$$

Remark 6.2 In [76], the maximum stable gain is defined as the largest gain for which the hearing aid does not oscillate and the desired signal is not substantially degraded by algorithm artefacts. The maximum stable gain is then determined by gradually and slowly increasing the gain while the adaptive procedure is running. Interestingly, an adaptive feedback canceller performs better if the gain $|G(q)|$ is increased during the adaptation: for larger $|G(q)|$, the feedback signal becomes more dominant w.r.t. the desired signal $x[k]$ (cf. Section 6.2.3). Hence, the MSG of [76] will be larger than the one defined in (6.11). □

Deviation of $C(q)$ from the desired transfer function $G(q)$

A third possible measure to assess the performance of the feedback cancellation algorithm, is the misalignment $\zeta(C(q), G(q))$ between the closed-loop response $C(q)$ in (6.9) and the desired response $G(q)$ of the hearing aid, measured as

$$\zeta(C(q),G(q)) = \sqrt{\int_0^{\pi}\left|1-\frac{1}{1-G(\omega)\left(F(\omega)-\hat{F}(\omega)\right)}\right|^2 d\omega}. \tag{6.13}$$

6.2 Standard adaptive feedback cancellation

This section describes the standard adaptive feedback cancellation technique that is widely studied for hearing aid applications [76, 129, 166]. The basic concepts of the adaptive feedback canceller are described in Section 6.2.1 and Section 6.2.2. In Section 6.2.3, we show that because of the presence of a closed signal loop, the standard feedback cancellation technique suffers from a large model error or bias, if the desired signal is spectrally colored.

6.2.1 Continuous versus non-continuous adaptation

Currently available adaptive feedback cancellers can be divided into two classes: algorithms with a continuous adaptation and algorithms with a *non-continuous*

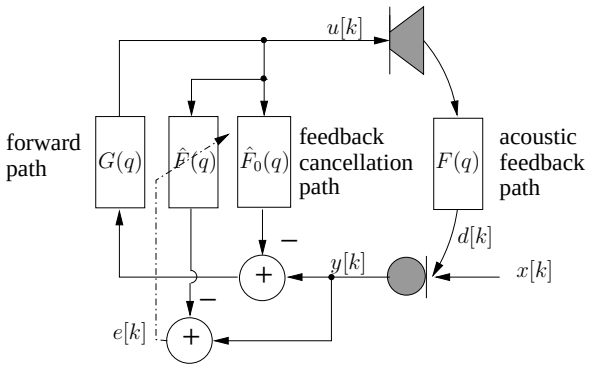


Figure 6.3: Continuous adaptation feedback canceller.

adaptation [76, 166]. The latter adapt the filter only when instability is detected or when the input signal level is low. During the filter adaptation, a training sequence $r[k]$ such as white noise is injected at the input of the loudspeaker and, in addition, the regular signal processing $G(q)$ of the hearing aid is sometimes interrupted [122, 126, 127, 196, 253]. This may be disturbing for the hearing aid user. Also, due to the reactive, rather than proactive, adaptation, these systems may be objectionable.

A *continuous adaptation* feedback canceller (CAF) continuously adapts the coefficients $\hat{\mathbf{f}}$ of the filter $\hat{F}(q)$, based on standard adaptive filtering (Wiener filtering) procedures (cf. Section 6.2.2). This is depicted in **Figure 6.3**. The filter $\hat{F}_0(q)$ in the feedback cancellation path is an initial estimate of $F(q)$, chosen such that the closed-loop system $C(q)$

$$C(q) = \frac{G(q)}{1 - G(q)\left(F(q) - \hat{F}_0(q)\right)} \tag{6.14}$$

is stable. It is typically replaced during adaptation by the estimate $\hat{F}(q)$.

The optimization criterion of *the standard CAF* is to minimize the energy $\mathcal{E}\{e^2[k]\}$ of the error signal $e[k]$

$$\boxed{e[k] = y[k] - \hat{F}(q)u[k]} \tag{6.15}$$

Ideally, $\hat{F}(q)$ converges to $F(q)$, so that, upon convergence, $e[k]$ equals the desired signal $x[k]$.

6.2.2 Adaptive filtering procedures

Standard adaptive filtering procedures (e.g., LMS, RLS) can be used for adapting the filter coefficients $\hat{\mathbf{f}}$ of the standard CAF $\hat{F}(q)$ of Figure 6.3.

Modified LMS algorithm

The most popular adaptive filtering technique for feedback cancellation is the LMS algorithm (cf. Chapter 2.3.1) [71, 76, 85]:

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \rho[k]\mathbf{u}[k]\epsilon[k] \tag{6.16}$$

where $\rho[k]$ is a time-varying step size, where

$$\mathbf{u}[k] = \begin{bmatrix} u[k] & u[k-1] & \cdots & u[k-L_{\hat{F}}+1] \end{bmatrix}^T, \tag{6.17}$$

$$\hat{\mathbf{f}}[k] = \begin{bmatrix} f_0[k] & f_1[k] & \cdots & f_{L_{\hat{F}}-1}[k] \end{bmatrix}^T, \tag{6.18}$$

and where $\epsilon[k]$ is the a priori residual, defined as

$$\epsilon[k] = y[k] - \hat{\mathbf{f}}^T[k-1]\mathbf{u}[k]. \tag{6.19}$$

Ideally the coefficient vector $\hat{\mathbf{f}}[k]$ converges to $\mathbf{f}[k]$, so that, upon convergence, the error signal $e[k]$ equals the desired signal $x[k]$, which is different from 0. Hence, to limit the excess MSE, the step size ρ should be kept small, at the expense of slow convergence. If the desired sound signal $x[k]$ exhibits large fluctuations in the short-time power (e.g., if $x[k]$ is a speech signal), the excess MSE of the LMS algorithm is further reduced by using a normalized time-varying step size $\rho[k]$ based on the sum-method, which is also referred to as the modified LMS algorithm [71, 76]:

$$\rho[k] = \frac{\bar{\rho}}{L_{\hat{F}}\left(p_u[k] + p_\epsilon[k]\right) + c}, \tag{6.20}$$

where $\bar{\rho} \in (0, 2]$ is a dimensionless step size constant, c is a positive constant to prevent singularities and p_u and p_ϵ are running power estimates of $u[k]$ and $\epsilon[k]$, respectively,

$$p_u[k] = \gamma p_u[k-1] + (1-\gamma)u^2[k], \tag{6.21}$$

$$p_e[k] = \gamma p_e + (1-\gamma)\epsilon^2[k], \tag{6.22}$$

with the forgetting factor γ for the power estimation set to $\gamma = 1 - \frac{1}{L_{\hat{F}}}$. For a highly time-varying desired signal $x[k]$ such as a speech signal, a burst in the error signal $\epsilon[k]$ mostly originates from fluctuations in the short-time power of $x[k]$. The sum-method (6.20) reduces the step size $\rho[k]$, when the desired signal $x[k]$ is strong and hence, mitigates the negative effect of a strong desired signal segment on the excess MSE.

Recursive Least Squares (RLS) algorithm

Faster convergence can be obtained when updating the adaptive filter $\hat{\mathbf{f}}[k]$ by means of the recursive least squares (RLS) algorithm, especially when the autocorrelation matrix $\mathcal{E}\{\mathbf{u}[k]\mathbf{u}^T[k]\}$ of the input data is ill-conditioned. At each time instant k, the RLS algorithm minimizes the least squares (LS) criterion

$$\sum_{\bar{k}=0}^{k} \lambda^{k-\bar{k}} e^2[\bar{k}] = \left\| \mathbf{y}[k] - \mathbf{U}[k]\hat{\mathbf{f}}[k] \right\|_2^2, \tag{6.23}$$

for $\hat{\mathbf{f}}[k]$, where $\lambda \in (0, 1]$ imposes an exponential data weighting and where

$$\mathbf{y}[k] = \begin{bmatrix} \lambda^0 y[k] \\ \lambda^{\frac{1}{2}} y[k-1] \\ \vdots \\ \lambda^{\frac{k}{2}} y[0] \end{bmatrix}; \quad \mathbf{U}[k] = \begin{bmatrix} \lambda^0 \mathbf{u}^T[k] \\ \lambda^{\frac{1}{2}} \mathbf{u}^T[k-1] \\ \vdots \\ \lambda^{\frac{k}{2}} \mathbf{u}^T[0] \end{bmatrix}. \tag{6.24}$$

It can be easily proven that minimization of (6.23) results in

$$\hat{\mathbf{f}}[k] = \mathbf{R}^{-1}[k]\mathbf{U}^T[k]\mathbf{y}[k], \tag{6.25}$$

where the matrix $\mathbf{R}[k]$ is defined as $\mathbf{U}^T[k]\mathbf{U}[k]$. Because of the matrix inversion in (6.25), computing $\hat{\mathbf{f}}[k]$ at each time instant k by means of (6.25) would require $\mathcal{O}\left(L_{\hat{F}}^3\right)$ computations.

The complexity can be reduced to $\mathcal{O}\left(L_{\hat{F}}^2\right)$ by recursively updating $\mathbf{R}^{-1}[k]$ from $\mathbf{R}^{-1}[k-1]$ using the matrix inversion lemma. This leads to the so-called RLS algorithm

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \mathbf{R}^{-1}[k]\mathbf{u}[k]\epsilon[k], \tag{6.26}$$

$$\mathbf{R}^{-1}[k] = \frac{1}{\lambda}\mathbf{R}^{-1}[k-1] - \frac{1}{\lambda^2}\frac{\mathbf{R}^{-1}[k-1]\mathbf{u}[k]\mathbf{u}^T[k]\mathbf{R}^{-1}[k-1]}{1 + \frac{1}{\lambda}\mathbf{u}^T[k]\mathbf{R}^{-1}[k-1]\mathbf{u}[k]}. \tag{6.27}$$

The RLS algorithm (6.26)-(6.27) suffers from poor numerical properties. For practical computation, numerically stable implementations based on a QR decomposition should be used [85, 173].

Some notes on the convergence of the adaptive feedback canceller

If the filter $\hat{F}_0(q)$ in the feedback cancellation path (see Figure 6.3) is *kept fixed during adaptation* with $C(q)$ (cf. (6.14)) stable, the input $u[k] = C(q)x[k]$ is bounded, for a bounded desired signal $x[k]$. Hence, the convergence properties of standard LMS and RLS apply. The LMS algorithm converges in the mean square, provided that its step size ρ is properly chosen (cf. Section 2.3.1) and

convergence of RLS is guaranteed if the auto-correlation matrix $\mathcal{E}\{\mathbf{u}[k]\mathbf{u}^T[k]\}$ of the input data is non-singular.

In practice, the *feedback canceller* $\hat{F}_0(q)$ *is replaced during adaptation* with new estimates $\hat{F}(q)$ of the feedback path $F(q)$. Typically, the filter $\hat{F}_0(q)$ is updated at the adaptation rate of the filter $\hat{F}(q)$ (i.e., after each time instant k), but it may also be updated at a lower rate, e.g., after each T time instants. The input data $u[k]$ used for identification is then determined by the adaptive filters $\hat{\mathbf{f}}[\bar{k}]$ at time instants $\bar{k} \leq k$ and hence, time-varying.

Example 6.1 For reasons of efficiency, $u[k]$ is generally derived from the a priori residual $\epsilon[k]$ as

$$\begin{aligned} u[k] &= G(q)\epsilon[k] \\ &= G(q)\left(y[k] - \hat{F}(q,k-1)u[k]\right). \end{aligned} \tag{6.28}$$

Alternatively, $u[k]$ could be computed from the a posteriori residual $\bar{\epsilon}[k]$ as

$$u[k] = G(q)\bar{\epsilon}[k], \tag{6.29}$$

with

$$\bar{\epsilon}[k] = y[k] - \hat{\mathbf{f}}[k]\mathbf{u}[k] = y[k] - \hat{F}(q,k)u[k]. \tag{6.30}$$

Since $\bar{\epsilon}_p[k] \leq \epsilon_p[k]$, (6.29) is expected to result in better performance than (6.28), but requires more computations. △

To guarantee convergence of the RLS or LMS algorithm, a boundness condition must now be satisfied, namely, the input data $u[k]$ should be bounded and hence, the filter $\hat{\mathbf{f}}[\bar{k}]$ (with $\bar{k} = n \cdot T$ ($n \in \mathbb{N}$) and $\bar{k} \leq k$) should not result in an unstable closed-loop system $C(q)$ for a significant length of time [152]. From simulations, it turns out that when $\hat{F}_0(q)$ is continuously being updated with $\hat{F}(q)$, the standard CAF has a self-stabilizing feature, whereby unstable poles of the closed-loop system $C(q)$ have a tendency to migrate back into the stable region. This tendency can be understood intuitively by considering the error signal $e[k]$. Using (6.6) and (6.28), $e[k]$ can be written as

$$e[k] = x[k] + \left(F(q) - \hat{F}(q)\right)\frac{G(q)x[k]}{1 - G(q)\left(F(q) - \hat{F}(q,k-1)\right)}. \tag{6.31}$$

If the feedback path estimate $\hat{F}(q,k-1)$ at time instant $k-1$ results in an unstable closed-loop system $C(q,k-1)$, the error $\left(F(q) - \hat{F}(q)\right)$ in the feedback path estimate gets a large weight at the oscillation frequencies, so that the next estimate $\hat{F}(q,k)$ will get close to $F(q)$ and thus stabilizes the closed-loop system again. In Appendix 6.A, the self-stabilizing feature has been proven for the RLS based CAF with $\lambda = 1$ and $u[k]$ computed as (6.29), assuming the sufficient order case, i.e., $L_{\hat{F}} \geq L_F$.

6.2.3 Bias of the standard continuous adaptation feedback canceller

The standard CAF assumes that the desired signal $x[k]$ is zero-mean and statistically independent of the input vector $\mathbf{u}[k]$, i.e.,

$$\mathcal{E}\{\mathbf{u}[k]x[k]\} = 0. \tag{6.32}$$

Because of the presence of the closed signal loop $G(q)$, the desired signal $x[k]$ and the input $u[k]$ to the adaptive filter $\hat{F}(q)$ are correlated when $x[k]$ is spectrally colored (e.g., a speech or music signal), generally causing a biased estimate $\hat{F}(q)$ of the feedback path $F(q)$ [221].

To illustrate this, we consider the RLS based CAF with $\lambda = 1$. At each time instant k, the adaptive filter $\hat{\mathbf{f}}[k]$ minimizes the LS criterion (6.23), i.e.,

$$\hat{\mathbf{f}}[k] = \left(\mathbf{U}^T[k]\mathbf{U}[k]\right)^{-1}\mathbf{U}^T[k]\mathbf{y}[k]. \tag{6.33}$$

Assume the sufficient-order case, i.e., $L_{\hat{F}} = L_F$. Then,

$$\mathbf{y}[k] = \underbrace{\mathbf{U}[k]\mathbf{f}^T}_{\mathbf{d}[k]} + \mathbf{x}[k], \tag{6.34}$$

with $\mathbf{f}$ the coefficient vector of the true feedback path $F(q)$ and with

$$\mathbf{x}[k] = \begin{bmatrix} x[1] & x[2] & \cdots & x[k] \end{bmatrix}^T. \tag{6.35}$$

Substituting (6.34) in (6.33), gives

$$\begin{aligned} \hat{\mathbf{f}}[k] &= \left(\mathbf{U}^T[k]\mathbf{U}[k]\right)^{-1}\mathbf{U}^T[k]\mathbf{d}[k] + \left(\mathbf{U}^T[k]\mathbf{U}[k]\right)^{-1}\mathbf{U}^T[k]\mathbf{x}[k] & (6.36)\\ &= \mathbf{f} + \underbrace{\left(\mathbf{U}^T[k]\mathbf{U}[k]\right)^{-1}\mathbf{U}^T[k]\mathbf{x}[k]}_{\text{bias}}. & (6.37) \end{aligned}$$

If $\mathcal{E}\{\mathbf{U}^T[k]\mathbf{x}[k]\} = 0$, the LS estimate $\hat{\mathbf{f}}[k]$ is *unbiased*, i.e.,

$$\mathcal{E}\{\hat{\mathbf{f}}\} = \mathbf{f}. \tag{6.38}$$

If however, $\mathcal{E}\{\mathbf{U}^T[k]\mathbf{x}[k]\} \neq 0$, the feedback path estimate $\hat{\mathbf{f}}[k]$ is biased. From (6.36), it follows that the larger the energy of the feedback signal $d[k]$ w.r.t. $x[k]$, the smaller the model error or bias will be. From Figure 6.3, we find

$$d[k] = \frac{F(q)G(q)}{1 - G(q)\left(F(q) - \hat{F}_0(q)\right)}x[k]. \tag{6.39}$$

Hence, the larger the open-loop gain $|G(q)F(q)|$, the smaller the bias in the feedback path estimate. In addition, (6.39) indicates that the bias will be smallest at the frequencies that are closest to instability.

Because of the presence of the closed signal loop $G(q)$, the input signal $u[k]$ to the adaptive filter relates to $x[k]$ as

$$u[k] = C(q)x[k], \tag{6.40}$$

where the closed-loop system $C(q)$ equals (cf. (6.14))

$$C(q) = \frac{G(q)}{1 - G(q)\left(F(q) - \hat{F}_0(q)\right)}. \tag{6.41}$$

Assume that $G(q)$ contains a delay $d_f \in \mathbb{N}$ with $d_f \geq 1$ and assume $G(q)$, $F(q)$, and $\hat{F}_0(q)$ to be causal and time-invariant. Then, $C(q)$ can be specified as

$$C(q) = c_{d_f} q^{-d_f} + c_{d_f+1} q^{-d_f-1} + \cdots + c_{L_C-1} q^{-L_C+1} \tag{6.42}$$

and hence

$$u[k] \quad = \quad \left(c_{d_f} + c_{d_f+1} q^{-1} + \cdots + c_{L_C-1} q^{-L_C+1}\right) x[k - d_f]. \tag{6.43}$$

Hence, $\mathcal{E}\{\mathbf{u}[k]x[k]\} = 0$ if and only if

$$\boxed{\mathcal{E}\{x[k-\Delta]x[k]\} = 0} \tag{6.44}$$

for $d_f \leq \Delta \leq L_C + L_{\hat{F}} - 2$.

White noise signal $x[k]$

If $x[k]$ is *white noise* with a noise variance σ_x^2

$$\mathcal{E}\{x[k-\Delta]x[k]\} = \sigma_x^2 \delta[\Delta], \tag{6.45}$$

where $\delta[k]$ is the Dirac-impulse and hence, the LS estimate $\hat{\mathbf{f}}$ is unbiased. If samples of $x[k]$ are independently and identically distributed, the variance of the estimation error can be computed as

$$\xi\left\{(\hat{\mathbf{f}}[k] - \mathbf{f}) \cdot (\hat{\mathbf{f}}[k] - \mathbf{f})^T\right\} = \sigma_x^2 (\mathbf{U}^T[k]\mathbf{U}[k])^{-1}. \tag{6.46}$$

Spectrally colored signal $x[k]$

However, most sound signals $x[k]$ in every-day life are *spectrally colored,* meaning that the signal values $x[k]$ are correlated in time (e.g., speech, music, car noise, alarm signals, etc). Many of the audio signals may be closely approximated as low-order, autoregressive (AR) random processes

$$x[k] = H(q)w[k] = \frac{1}{1 + q^{-1}P(q)} w[k], \tag{6.47}$$

with $w[k]$ a white noise signal. Hence, the desired signal model $H(q)$ is often IIR, so that for a small delay d_f in the forward path $G(q)$, the length L_H of $H(q)$ exceeds d_f and hence, $\mathcal{E}\{\mathbf{u}[k]x[k]\} \neq 0$. The CAF will then cancel the desired signal $x[k]$ instead of the feedback signal $d[k] = F(q)u[k]$ (see (6.36)), leading to signal distortion.

Example 6.2 If $x[k]$ is a speech signal, a segment of $x[k]$ can be modelled by an all-pole model (6.47) with then $w[k]$ a white noise signal (in case of unvoiced sounds) or a periodic impulse train with period equal to the pitch period P of speech (in case of voiced sounds) [31]. For a sampling frequency $f_s = 16$ kHz, the order L_P of the all-pole model of a speech segment equals about 20. △

6.3 Solutions to reduce the bias

In the literature, different methods have been proposed to reduce the bias in case of spectrally colored desired signals $x[k]$, such as decorrelating the desired signal $x[k]$ and the input $u[k]$ to the adaptive filter, inserting a probe signal $r[k]$ and controlling the adaptation of the feedback canceller $\hat{F}(q)$ [51, 67, 112, 118, 125, 129, 132, 147, 182, 183, 184, 221, 226].

6.3.1 Reducing the correlation between $x[k]$ and $u[k]$

A first solution to reduce the bias is to include some decorrelating signal operation in cascade with the forward path $G(q)$ or, in the case of a delay operation, the cancellation path $\hat{F}(q)$ (i.e., at the input of the adaptive filter $\hat{F}(q)$), as depicted in **Figure 6.4**. Ideally, the decorrelation unit removes any correlation between the desired signal $x[k]$ and the input to the adaptive filter $\hat{F}(q)$, so that the bias term in (6.37) goes to zero for $k \to \infty$. Often, the decorrelation unit however degrades the sound quality, making full decorrelation impossible.

Delays in the forward path or the cancellation path

A common technique to reduce the correlation is the use of delays in the forward path $G(q)$ or in the cancellation path [221].

For many signals $x[k]$, the autocorrelation sequence $\mathcal{E}\{x[k]x[k-\Delta]\}$ decreases with increasing lag Δ. As a result, inserting a delay $d_f > 1$ in $G(q)$ generally reduces the correlation between $x[k]$ and $u[k]$ (cf. (6.43)-(6.44)). Alternatively, the correlation between $x[k]$ and the input to the adaptive filter $\hat{F}(q)$ can be reduced by a delay d_c in the cancellation path. However, this implies that the first d_c coefficients of the true feedback path $F(q)$ are approximated with zeros. An unbiased feedback path estimate $\hat{F}(q)$ may then only be obtained if

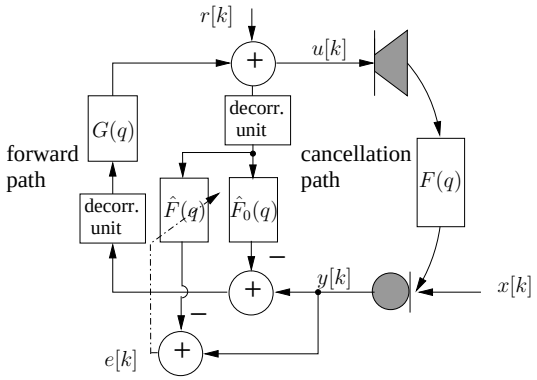


Figure 6.4: Continuous adaptation feedback cancellation with decorrelating signal operations.

the true feedback path $F(q)$ also contains a delay d_c [221]. A delay in the true feedback path $F(q)$ may arise from the propagation time of the sound from the loudspeaker to the microphone and the processing delay of the DAC and the ADC.

Many of the audio signals can be closely approximated by a low-order, all-pole model $H(q)$ (cf. Section 6.2.3) and hence, have a relatively strong autocorrelation. A large delay d_f may then be required to efficiently decorrelate $x[k]$ and $u[k]$. However, the delay d_f in the forward path $G(q)$ of the hearing aid should be kept small in order not to degrade intelligibility [247] and sound quality, e.g., by comb filter effects (i.e., sequence of notches and peaks in the frequency response) resulting from destructive and constructive interference between the direct sound to the eardrum and the amplified and delayed sound [88, 145]. In addition, adding delays will not help if $x[k]$ is tonal (e.g., a sinusoid) as is the case for voiced speech, music, alarm and whistling signals.

Inserting a probe signal $r[k]$ or non-linearities

Alternatively, the correlation between $x[k]$ and $u[k]$ can be reduced by inserting a probe signal $r[k]$ at the input $u[k]$ of the loudspeaker (cf. Figure 6.4) or by adding non-linearities to the forward path $G(q)$, such as frequency shifting, phase or delay modulation [51, 112, 118, 184]. The probe signal is generally a noise signal. Such methods may, however, again degrade the sound quality. In the former case, e.g., the probe signal should be inaudible so that it does not degrade the signal-to-noise ratio. Its level will thus have to be lower than the level of the ordinary output $u[k]$ with a result that significant residual bias still remains.

Remark 6.3 Inserting a probe signal $r[k]$ at the input of the loudspeaker can best be used in hearing aids designed for the severe hearing impaired where the increased hearing threshold and hence reduced dynamic range can be exploited to increase the level of the probe signal. □

6.3.2 Identification based on a probe signal $r[k]$

In [222], a probe signal $r[k]$ is inserted at the input of the loudspeaker and the adaptive filter $\hat{F}(q)$ is adapted based on the probe signal $r[k]$ instead of the loudspeaker signal $u[k]$. Because of the presence of the closed signal loop $G(q)$,

$$y[k] = \frac{F(q)}{1 - G(q)\left(F(q) - \hat{F}_0(q)\right)} r[k] + \frac{1 + G(q)\hat{F}_0(q)}{1 - G(q)\left(F(q) - \hat{F}_0(q)\right)} x[k]. \quad (6.48)$$

As a result, the CAF $\hat{F}(q)$ models the closed-loop transfer function from the probe signal $r[k]$ to the microphone signal $y[k]$ instead of the open-loop transfer function $F(q)$. In addition, because of the low level of the probe signal $r[k]$ with respect to the desired output $G(q)x[k]$ (which is required to preserve sound quality), only a minor part of the microphone signal $y[k]$ originates from $r[k]$, resulting in a large excess error or slow adaptation of the adaptive filter $\hat{F}(q)$.

6.3.3 Controlling the adaptation of the feedback canceller

In [24, 125, 129, 132, 147, 182, 183, 206], the bias of the adaptive feedback canceller $\hat{F}(q)$ is reduced by controlling the adaptation of the feedback canceller.

Reducing the adaptation speed of $\hat{F}(q)$

In [132, 182, 183], the adaptation speed of the feedback canceller is kept small to reduce distortion of the desired signal. When feedback or a sudden change in feedback path is detected, the adaptation speed is temporarily increased. A slow adaptation speed reduces the ability to track changes in the spectrum of the desired signal $x[k]$, and hence, limits the negative effect of a temporary signal with a strong autocorrelation, on the feedback path estimate [88]. Convergence of this technique may however be too slow to track variations in the feedback path so that the maximum amplification may still be limited.

Constrained adaptation

In [125, 126, 127, 129, 130], the bias is reduced by constrained adaptation: the filter coefficients $\hat{\mathbf{f}}$ of the adaptive feedback canceller $\hat{F}(q)$ are not allowed to deviate too far from reference filter coefficients $\hat{\mathbf{f}}_{ref}$, which are measured during start-up by means of a probe signal.

Two approaches to constrained adaptation have been proposed [125, 129]. In a first method, the hard constraint

$$\left\|\hat{\mathbf{f}}-\hat{\mathbf{f}}_{ref}\right\|_1 = \sum_{i=0}^{L_{\hat{F}}-1}\left|\hat{f}_i-\hat{f}_{ref,i}\right| < \zeta \sum_{i=0}^{L_{\hat{F}}-1}\left|\hat{f}_{ref,i}\right| \tag{6.49}$$

is imposed on the one-norm $\left\|\hat{\mathbf{f}}-\hat{\mathbf{f}}_{ref}\right\|_1$ by scaling the filter coefficients $\hat{\mathbf{f}}$ after each update if the constraint (6.49) is not fulfilled (cf. the SPA (2.95) of Section 2.3.3):

$$\text{If } \left\|\hat{\mathbf{f}}-\hat{\mathbf{f}}_{ref}\right\|_1 > \zeta\left\|\hat{\mathbf{f}}_{ref}\right\|_1 : \quad \hat{\mathbf{f}} = \hat{\mathbf{f}}_{ref} + \left(\hat{\mathbf{f}}-\hat{\mathbf{f}}_{ref}\right)\frac{\zeta\left\|\hat{\mathbf{f}}_{ref}\right\|_1}{\left\|\hat{\mathbf{f}}-\hat{\mathbf{f}}_{ref}\right\|_1}. \tag{6.50}$$

The scalar ζ trades off between desired signal cancellation and the ability to model variations on $\hat{\mathbf{f}}_{ref}$. In case of LMS-updating, a cheaper method is to add a term to the cost function $J(\hat{\mathbf{f}})$ of the feedback canceller that penalizes excessive deviation of the filter $\hat{\mathbf{f}}$ from the reference filter $\hat{\mathbf{f}}_{ref}$, i.e.,

$$J(\hat{\mathbf{f}}) = \mathcal{E}\{\left(y[k]-\hat{\mathbf{f}}^T\mathbf{u}[k]\right)^2\} + \eta\mathcal{E}\{\left(\hat{\mathbf{f}}-\hat{\mathbf{f}}_{ref}\right)^T\left(\hat{\mathbf{f}}-\hat{\mathbf{f}}_{ref}\right)\}, \tag{6.51}$$

where η is a trade-off factor. This results in the update equation

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \rho\mathbf{u}[k]\epsilon[k] - \rho\eta\left(\hat{\mathbf{f}}[k-1]-\hat{\mathbf{f}}_{ref}\right), \tag{6.52}$$

where $\epsilon[k] = y[k]-\hat{\mathbf{f}}^T[k-1]\mathbf{u}[k]$. In contrast to (6.49)-(6.50), this technique imposes a soft constraint on $\hat{\mathbf{f}}[k]$ (cf. the leaky LMS algorithm (2.97) of Section 2.3.3).

For tonal signals it is observed that even with a delay added to the forward path $G(q)$, the adaptive filter $\hat{F}(q)$ effectively cancels the input signal instead of the feedback signal. In addition, the variations on the reference feedback path that can be modelled with this technique are rather limited.

Band-limited feedback cancellation with a filtered-X algorithm

In [24, 67, 183, 206, 226], feedback cancellation is restricted to the frequency band that encompasses the unstable frequencies by using a filtered-X technique. This is illustrated in **Figure 6.5**. The filters $B_1(q)$ and $B_2(q)$ are chosen such that all possible oscillation frequencies are preserved while other desired signal components are removed as much as possible. The filter $B_1(q)$ in the feedback cancellation path limits the feedback cancellation signal to the frequency band of interest and hence, prevents the desired signal components outside the critical band from being distorted by the adaptive filter $\hat{F}(q)$. To focus the modeling effort of $\hat{F}(q)$ on the regions that contain the critical frequencies, the

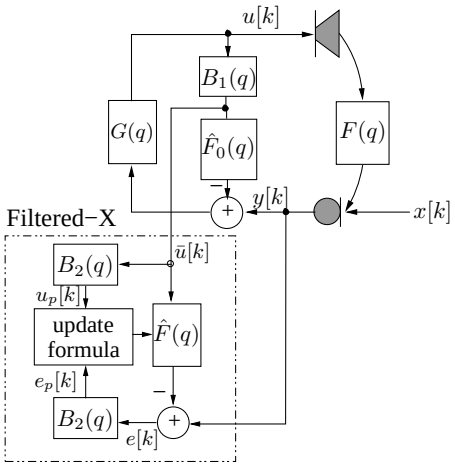


Figure 6.5: Band-limited feedback cancellation with a filtered-X algorithm.

input signal $u[k]$ and the error signal $e[k]$ are band-limited by the filter $B_2(q)$ before being used to adapt the adaptive filter $\hat{F}(q)$, resulting in a filtered-X algorithm.

The *filtered-X algorithm* [15, 17] is given by [24]

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \alpha \mathbf{R}^{-1}[k]\mathbf{u}_p[k]\epsilon_p[k] \tag{6.53}$$

where α is a scalar with $0 < \alpha \leq 1$ (cf. Remark 6.5) and $\mathbf{u}_p[k]$, $\epsilon_p[k]$ are defined as

$$\begin{aligned}
\mathbf{u}_p[k] &= \begin{bmatrix} u_p[k] & u_p[k-1] & \cdots & u_p[k-L_{\hat{F}}+1] \end{bmatrix}^T, && (6.54)\\
u_p[k] &= B_2(q)\bar{u}[k], && (6.55)\\
\epsilon_p[k] &= B_2(q)\epsilon[k], && (6.56)\\
\epsilon[k] &= y[k] - \hat{F}(q,k-1)\bar{u}[k], && (6.57)
\end{aligned}$$

with $\bar{u}[k] = B_1(q)u[k]$. For RLS updating

$$\mathbf{R}[k] = \lambda \mathbf{R}[k-1] + \mathbf{u}_p[k]\mathbf{u}_p^T[k], \tag{6.58}$$

where λ is the forgetting factor, while for LMS updating

$$\alpha \mathbf{R}^{-1}[k] = \rho[k]\mathbf{I}_{L_{\hat{F}}}. \tag{6.59}$$

As in the standard LMS based CAF, a time-varying step size $\rho[k]$ based on the sum-method is preferable for highly time-varying signals $x[k]$ (cf. Section 6.2.2):

$$\begin{aligned}
\rho[k] &= \frac{\bar{\rho}}{c + L_{\hat{F}} p[k]} && (6.60)\\
p[k] &= \gamma p[k-1] + (1-\gamma)\left(u_p^2[k] + \epsilon_p^2[k]\right). && (6.61)
\end{aligned}$$

Remark 6.4 Note that the input $\bar{u}[k]$ and error $e[k]$ are both filtered by the same filter $B_2(q)$. By doing so, the phase requirement for stable adaptation [24, 223] of the filtered-X LMS algorithm is always satisfied so that the algorithm does not suffer from the stability problems encountered in, e.g., active noise control because of plant model errors. □

Remark 6.5 If there is a group delay associated with the filter $B_2(q)$, changes to the adaptive filter $\hat{\mathbf{f}}[k]$ do not have an immediate impact on the error $\epsilon_p[k]$. If the group delay is large, the adaptive filter should change slowly in time to guarantee stability (i.e., stationary system hypothesis of filtered-X) and hence, a small step size may be required [17, 24, 154, 155]. For LMS, this amounts to using a small step size ρ. For RLS, the initialization of $\mathbf{R}[k]$ should not be too small and/or a scalar $\alpha < 1$ may have to be added before $\mathbf{R}^{-1}[k]$ in (6.53), at the expense of reduced convergence [17]. The reduction in convergence rate is counterbalanced by the removal of the unwanted components in the adaptive filter input and error signals, which accelerates convergence [24]. □

Typically, the feedback path of a hearing aid provides less attenuation at high frequencies [90], where most hearing aid users have the largest hearing loss. As a result, the risk for acoustic feedback is often highest in the high frequency range, while most of the desired signal energy is typically concentrated at low frequencies, e.g., for speech signals. Hence, $B_1(q)$ and $B_2(q)$ are usually high pass or band pass. However, to specify the bandwidth of the filters $B_1(q)$ and $B_2(q)$, identification of the oscillation frequencies for different hearing aids and different hearing aid users is required, which makes this technique rather cumbersome for practical use. In addition, the critical frequency region expands with increasing gain in the forward path $G(q)$ as well as, e.g., with the presence of a telephone handset or hand palm close to the ear. The bandwidth of the bandpass filters would then need to be increased accordingly and as a result, the improvement which comes from the band-limited adaptive feedback cancellation would be compromised.

Filtered-X feedback cancellation with adaptive decorrelation filters

In [147], a filtered-X LMS based feedback canceller with adaptive decorrelation filters has been proposed: the input signal $u[k]$ and the error signal $e[k]$ are pre-whitened through adaptive linear prediction (LP) filters $A(q)$ before updating the feedback canceller $\hat{F}(q)$ (cf. Figure 6.5 with $B_2(q) = A(q)$ and $B_1(q) = 1$). The adaptation of the LP filter $A(q) = 1 + q^{-1}\bar{A}(q)$ is performed on the error signal $e[k]$:

$$\bar{A}(q) = \arg\min_{\bar{A}(q)} \sum_{\bar{k}=0}^{k} \lambda_{\tilde{\mathbf{a}}}^{k-\bar{k}} (e[\bar{k}] + \bar{A}(q)e[\bar{k}-1])^2, \tag{6.62}$$

where $\lambda_{\bar{\mathbf{a}}}$ is the forgetting factor. Linear prediction removes the correlation between consecutive samples of a signal [31] and hence, reduces the autocorrelation function. To achieve a fast decorrelation, the LP filter $\bar{A}(q)$ is updated using an efficient lattice algorithm instead of LMS [161] (cf. Chapter 7).

Originally, the idea of adaptive pre-whitening was proposed in [167] to improve the convergence of an LMS based echo canceller: adaptive pre-whitening reduces the autocorrelation of colored input signals and hence, the eigenvalue spread of the input correlation matrix. In case of feedback cancellation, pre-whitening the input and the error signal also helps to reduce the bias of the feedback canceller (cf. Chapter 7). The filtered-X based feedback canceller in [147] has been derived from [167] by performing the adaptation of the linear prediction filter on the feedback compensated signal $e[k]$ instead of on the input signal $u[k]$. By doing so, the spectral maxima in the forward path $G(q)$ are preserved in the filtered input signal $u_p[k]$. As a result, the corresponding frequencies get a large weighting during the updating of the feedback canceller, which is desired because these frequencies are usually the ones that are the most susceptible to acoustic feedback.

In Chapter 7, we show that the here described filtered-X feedback cancellation technique with adaptive decorrelation can be motivated based on a closed-loop system identification point of view. In addition, we demonstrate that because of the presence of the closed signal loop $G(q)$, some conditions on the closed-loop system have to be satisfied to guarantee a good performance.

6.4 Closed-loop identification of the feedback path

In all of the above mentioned CAF techniques, the presence of the closed signal loop $G(q)$ has initially been ignored and the standard adaptive filtering techniques for open-loop systems (i.e., LMS, RLS) have been applied. In a second stage, signal processing operations are included to reduce the negative effect of the presence of the closed signal loop $G(q)$ on the performance. An alternative, more elegant approach is to immediately regard the feedback path as a part of a closed-loop system and to apply closed-loop system identification theory [59, 60]. This approach has recently been adopted for feedback cancellation in [87, 88, 89, 217] and is further investigated in Chapter 7, with the aim of providing solutions for the bias in the feedback path estimate.

Section 6.4.1 describes the closed-loop system set-up. Section 6.4.2 and Section 6.4.3 describe the three basic closed-loop system identification procedures in the context of feedback cancellation, i.e., the indirect method, the joint input-output method and the direct method. Among these different proce-

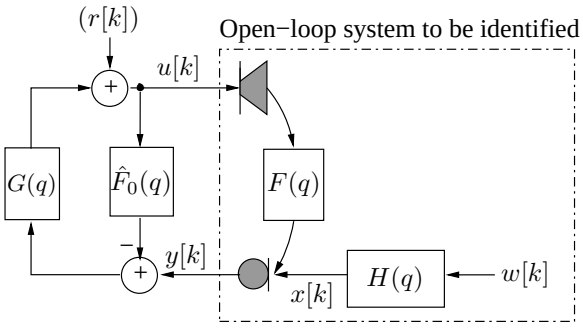


Figure 6.6: Closed-loop system set-up.

dures, especially the direct method seems attractive for feedback cancellation. In [87, 88, 89], a solution for the bias of the standard CAF has been proposed based on the direct method of closed-loop identification: it is shown that the bias can be avoided by means of a filtered-X algorithm, provided that the desired signal $x[k]$ is non-tonal and its spectrum is known. This technique is discussed in Section 6.4.4 and forms the basis for the adaptive feedback cancellation techniques that will be developed in Chapter 7.

6.4.1 Closed-loop system set-up

Figure 6.6 depicts the set-up of the closed-loop system. The *open-loop system to be identified* is given by

$$y[k] = F(q)u[k] + x[k]. \tag{6.63}$$

For the time being, we assume that the desired signal $x[k]$ can be modelled as

$$x[k] = H(q)w[k], \tag{6.64}$$

with $w[k]$ a zero mean, white noise sequence and $H(q)$ monic and inversely stable.

The output signal $y[k]$ is fed-back to the input $u[k]$ according to

$$u[k] = G(q)\left(y[k] - \hat{F}_0(q)u[k]\right) + r[k] \tag{6.65}$$

We (again) assume that the forward path $G(q)$ contains a delay d_f of at least one sample. The filter $\hat{F}_0(q)$ in the feedback cancellation path is an initial estimate of the feedback path $F(q)$ chosen such that the closed-loop system (6.14) is stable. The signal $r[k]$ is an optional probe signal that is independent of $x[k]$, and can be seen as an extra excitation signal of the feedback path $F(q)$. The probe signal is generally a noise signal.

By combining (6.63) and (6.65), we get the closed-loop relations [59, 60]

$$y[k] = (1 + F(q)C(q))\, x[k] + F(q)S(q)r[k] \quad (6.66)$$

$$u[k] = C(q)x[k] + S(q)r[k] \quad (6.67)$$

where $S(q)$ is the so-called sensitivity function

$$S(q) = \frac{1}{1 - G(q)\left(F(q) - \hat{F}_0(q)\right)} \quad (6.68)$$

and $C(q)$ denotes the closed-loop transfer function from the desired signal $x[k]$ to the loudspeaker input $u[k]$

$$C(q) = S(q)G(q). \quad (6.69)$$

For future use, we refer to the closed-loop transfer function $F(q)S(q)$ from the excitation $r[k]$ to the output $y[k]$ as $F_c(q)$:

$$F^c(q) = F(q)S(q). \quad (6.70)$$

Depending on the assumptions that are made about the forward path $G(q)$, closed-loop identification methods can be classified as *indirect, joint input-output* or *direct* methods [59, 60]. Section 6.4.2 describes the indirect and the joint input-output method, while Section 6.4.3 discusses closed-loop identification with the direct method.

6.4.2 Indirect and joint input-output method

Indirect method

The indirect method [59, 60] requires the use of a probe signal $r[k]$ as well as knowledge of the forward path $G(q)$. This makes it not that suitable for hearing aids because the forward path $G(q)$ of a hearing aid often contains user-dependent and adaptive signal processing features.

In the indirect method, the closed-loop transfer function $F^c(q)$ (see (6.70)) from the probe signal $r[k]$ to the microphone signal $y[k]$ is estimated. The feedback path estimate is then extracted from this closed-loop transfer function estimate $\hat{F}^c(q)$, using the knowledge of the forward path $G(q)$ and the feedback cancellation path $\hat{F}_0(q)$, as:

$$\hat{F}(q) = \hat{F}^c(q)\frac{1 + G(q)\hat{F}_0(q)}{1 + G(q)\hat{F}^c(q)}. \quad (6.71)$$

Since the identification of the closed-loop system $F^c(q)$ is degraded by the presence of the desired signal $x[k]$ (cf. (6.66)), a large noise level may be required to obtain an accurate feedback path estimate $\hat{F}(q)$ (cf. Section 6.3.2).

Joint input-output method

In [217], the joint input-output method is applied to obtain an unbiased feedback path estimate $\hat{F}(q)$. In this approach, the forward path $G(q)$ is assumed to have a certain structure (e.g., a linear filter) and the loudspeaker signal $u[k]$ and the microphone signal $y[k]$ are jointly viewed as the output of a system driven by the probe signal $r[k]$ (cf. (6.66)-(6.67)) [59, 60]:

$$\begin{bmatrix} y[k] \\ u[k] \end{bmatrix} = \begin{bmatrix} F^c(q) \\ S(q) \end{bmatrix} r[k] + \begin{bmatrix} 1 + F(q)C(q) \\ C(q) \end{bmatrix} x[k]. \tag{6.72}$$

The closed-loop transfer function $F^c(q)$ from $r[k]$ to $y[k]$ and the sensitivity function $S(q)$ are determined by means of standard adaptive filtering techniques. The feedback path estimate $\hat{F}(q)$ is then derived from these two model estimates $\hat{F}^c(q)$ and $\hat{S}(q)$ as:

$$\hat{F}(q) = \hat{F}^c(q)\hat{S}^{-1}(q). \tag{6.73}$$

Again, the identification is impeded by the presence of the desired signal $x[k]$. The low level of the probe signal $r[k]$ w.r.t. the desired output $G(q)x[k]$, which is required to preserve sound quality, causes slow adaptation or a large variance of the feedback path estimate [217]. In addition, large filter lengths are required to accurately identify the closed-loop transfer functions $\hat{F}^c(q)$ and $\hat{S}(q)$, as they can easily contain narrowband peaks when the closed-loop system $F^c(q)$ is close to instability.

Remark 6.6 There also exist joint input-output methods that do not require the use of a *measurable* probe signal $r[k]$. In these methods, $u[k]$ and $y[k]$ are jointly modelled as a multi-variable time-series, satisfying

$$\begin{bmatrix} y[k] \\ u[k] \end{bmatrix} = \underbrace{\begin{bmatrix} F^c(q) & (1 + F(q)C(q))\,H(q) \\ S(q) & C(q)H(q) \end{bmatrix}}_{\mathbf{H}_0(q)} \begin{bmatrix} r[k] \\ w[k] \end{bmatrix} \tag{6.74}$$

with $\begin{bmatrix} r[k] & w[k] \end{bmatrix}^T$ a white noise process and where $r[k]$ is a disturbance signal that is independent of $w[k]$ (e.g., originating from imperfect knowledge of $G(q)$ or internal noise of the loudspeaker). The feedback path estimate is then derived from the estimate of $\mathbf{H}_0(q)$. □

6.4.3 Closed-loop identification with the direct method

In the direct method, the open-loop system $\{F(q),\, H(q)\}$

$$y[k] = F(q)u[k] + H(q)w[k], \tag{6.75}$$

is directly identified from the loudspeaker signal $u[k]$ and the microphone signal $y[k]$, thereby ignoring the presence of the forward path $G(q)$. This implies that

the method can be applied for any arbitrary (unknown) forward path $G(q)$. In contrast to the indirect method and the joint input-output identification procedure proposed in [217], the use of a probe signal $r[k]$ is not required, making it an appealing approach for feedback cancellation.

The direct method is only applicable with some specific open-loop identification methods, such as the prediction error method (PEM) [59, 60]. Recall from Section 6.2.3 that the standard adaptive filtering techniques such as LMS and RLS, fail for closed-loop data because of signal correlation. The PEM produces an estimate $\hat{F}(q)$ and $\hat{H}(q)$ of the feedback path $F(q)$ and the desired signal model $H(q)$, respectively, by minimizing the difference between the measured output $y[k]$ and the predicted output $\hat{y}[k]$, called the prediction error $e_p[k]$. The optimal one step ahead predictor of $y[k]$ with the model $\{\hat{F}(q), \hat{H}(q)\}$ is given as [59, 60]

$$\hat{y}[k] = \hat{H}(q)^{-1}\hat{F}(q)u[k] + \left(1 - \hat{H}(q)^{-1}\right) y[k]. \tag{6.76}$$

Using (6.76), the prediction error $e_p[k] = y[k] - \hat{y}[k]$ can be written as

$$e_p[k] = \hat{H}(q)^{-1}\left(y[k] - \hat{F}(q)u[k]\right). \tag{6.77}$$

A standard optimization criterion for $\hat{F}(q)$ and $\hat{H}(q)$ is to minimize the time-average of squared prediction errors, i.e.,

$$\frac{1}{N}\sum_{k=0}^{N-1} e_p^2[k]. \tag{6.78}$$

Remark 6.7 In addition to an estimate of the inverse desired signal model $H^{-1}(q)$, the filter $\hat{H}^{-1}(q)$ can be interpreted as a pre-filter $A(q)$ that gives a frequency specific weight to the error signal $e[k] = y[k] - \hat{F}(q)u[k]$ of the standard CAF [59, 88]. □

In [87, 88, 89], the feedback path $F(q)$ is identified with the PEM method, using a *fixed* estimate of the desired signal model $H(q)$: it is shown that an unbiased feedback path estimate can be obtained if the desired signal model $H(q)$ is inversely stable and known. This technique is discussed in Section 6.4.4. In Chapter 7, adaptive feedback cancellation techniques with direct closed-loop identification of both the feedback path $F(q)$ and the desired signal model $H(q)$ will be developed.

6.4.4 Feedback cancellation with the direct method and a fixed estimate of the desired signal model

In [87, 88, 89], the desired signal model $H(q)$ is approximated by a *fixed* low-pass filter $\hat{H}(q) = A^{-1}(q)$, which represents the long-term average spectrum of

speech. The coefficient vector $\hat{\mathbf{f}}[k]$ of the feedback path $F(q)$ is then recursively estimated based on minimization at time k of the LS criterion

$$\sum_{\bar{k}=0}^{k} \lambda^{k-\bar{k}} e_p^2[\bar{k}], \tag{6.79}$$

where λ imposes an exponential data weighting and

$$e_p[k] = A(q)\left(y[k] - \hat{F}(q)u[k]\right). \tag{6.80}$$

Remark 6.8 Note that the standard CAF of Section 6.2 corresponds to a special case of the direct method, namely, with fixed desired signal model estimate $A(q) = 1$. □

Recursive PEM: Filtered-X algorithm

For a fixed filter $A(q)$, the recursive PEM (6.79)-(6.80) results in the following filtered-X update equation (cf. Section 6.3.3) [88]:

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \alpha \mathbf{R}^{-1}[k]\mathbf{u}_p[k]\epsilon_p[k], \tag{6.81}$$

with

$$\begin{aligned}
\mathbf{u}_p[k] &= \left[\; u_p[k] \quad u_p[k-1] \quad \cdots \quad u_p[k-L_{\hat{F}}+1] \;\right]^T, && (6.82)\\
u_p[k] &= A(q)u[k] && (6.83)\\
\epsilon_p[k] &= A(q)\epsilon[k], && (6.84)\\
\epsilon[k] &= y[k] - \hat{F}(q,k-1)u[k]. && (6.85)
\end{aligned}$$

The choice for $\mathbf{R}[k]$ in (6.81) depends on the algorithm (e.g., LMS, RLS) that is being used:

- If
$$\alpha \mathbf{R}^{-1}[k] = \rho[k]\mathbf{I}_{L_{\hat{F}}}, \tag{6.86}$$
(6.81) results in a *filtered-X LMS* algorithm with step size $\rho[k]$. Again, for highly time-varying input signals $x[k]$ such as a speech signals, better performance is obtained with the modified LMS algorithm (cf. Section 6.3.3):
$$\begin{aligned}
\rho[k] &= \frac{\bar{\rho}}{L_{\hat{F}}p[k] + c}, && (6.87)\\
p[k] &= \gamma p[k-1] + (1-\gamma)\left(u_p^2[k] + \epsilon_p^2[k]\right). && (6.88)
\end{aligned}$$
If there is a group delay associated with the filter $A(q)$, a small step size $\bar{\rho}$ may be required to guarantee stability (cf. Section 6.3.3).

- If
$$\mathbf{R}[k] = \lambda \mathbf{R}[k-1] + \mathbf{u}_p[k]\mathbf{u}_p^T[k], \tag{6.89}$$
(6.81) corresponds to *filtered-X RLS*. Again, if there is a large group delay associated with $A(q)$, a scalar $\alpha < 1$ may have to be added in (6.81) to guarantee stability [17] (cf. Remark 6.5 in Section 6.3.3).

Instrumental variable (IV) method with pre-filtering of the data

The recursive PEM or filtered-X algorithm (6.81) filters the input signal $u[k]$ and the error signal $\epsilon[k]$ before using them to update the adaptive filter $\hat{\mathbf{f}}$. By re-arranging the recursive PEM criterion (6.79)-(6.80), an alternative update equation for $\hat{\mathbf{f}}$ can be derived that corresponds to an (optimal) instrumental variable (IV) method with pre-filtering of the data. In the literature, this technique is also referred to as refined instrumental variable (RIV) method [152, Chapter 3].

Since $A(q)$ and $\hat{F}(q)$ in (6.79) are fixed, the LS criterion (6.79)-(6.80) can be written as:

$$\begin{aligned}\sum_{\bar{k}=0}^{k} \lambda^{k-\bar{k}} e_p^2[\bar{k}] &= \sum_{\bar{k}=0}^{k} \lambda^{k-\bar{k}} \left(A(q)y[k] - \hat{F}(q)A(q)u[k] \right)^2 \\ &= \left\| \mathbf{y}_p[k] - \mathbf{U}_p[k]\hat{\mathbf{f}} \right\|_2^2, \end{aligned} \tag{6.90}$$

where

$$\mathbf{y}_p[k] = \begin{bmatrix} \lambda^0 y_p[k] \\ \lambda^{\frac{1}{2}} y_p[k-1] \\ \vdots \\ \lambda^{\frac{k}{2}} y_p[0] \end{bmatrix}; \quad \mathbf{U}_p[k] = \begin{bmatrix} \lambda^0 \mathbf{u}_p^T[k] \\ \lambda^{\frac{1}{2}} \mathbf{u}_p^T[k-1] \\ \vdots \\ \lambda^{\frac{k}{2}} \mathbf{u}_p^T[0] \end{bmatrix}, \tag{6.91}$$

$$y_p[k] = A(q)y[k]. \tag{6.92}$$

Hence, the coefficient vector $\hat{\mathbf{f}}[k]$ can be adapted by applying the standard adaptive filtering techniques of Section 6.2.2 to the pre-filtered data $y_p[k]$ and $u_p[k]$, as depicted in Figure 6.7. For RLS, this results in an RIV update equation [152, Chapter 3]:

$$\begin{aligned}\hat{\mathbf{f}}[k] &= \hat{\mathbf{f}}[k-1] + \mathbf{R}^{-1}[k]\boldsymbol{\zeta}[k] \left(y_p[k] - \mathbf{u}_p^T[k]\hat{\mathbf{f}}[k-1] \right) \\ \mathbf{R}[k] &= \lambda \mathbf{R}[k-1] + \boldsymbol{\zeta}[k]\mathbf{u}_p^T[k] \end{aligned} \tag{6.93}$$

where the data is pre-filtered with the filter $A(q) = \hat{H}^{-1}(q)$ and the instrumental variable vector $\boldsymbol{\zeta}[k]$ is set to $\mathbf{u}_p[k]$. If $A(q) = H^{-1}(q)$ and $w[k]$ is white, this algorithm corresponds to the *optimal* RIV method, which means that it has the best accuracy among all choices for the instruments $\boldsymbol{\zeta}[k]$ and the pre-filter $A(q)$. For all other choices of $\boldsymbol{\zeta}[k]$ and $A(q)$, the accuracy of the RIV method will be worse than the PEM with optimal noise model [59].

Remark 6.9 In contrast to the filtered-X algorithm, any change to the adaptive filter $\hat{\mathbf{f}}$ now has an immediate effect on the prediction-error $e_p[k]$. Hence, if $A(q)$ has a group delay, larger step sizes may be used as compared to the filtered-X algorithm. To compute the feedback compensated signal $e[k]$, an additional filtering of $u[k]$ with $\hat{F}(q)$ is however required. □

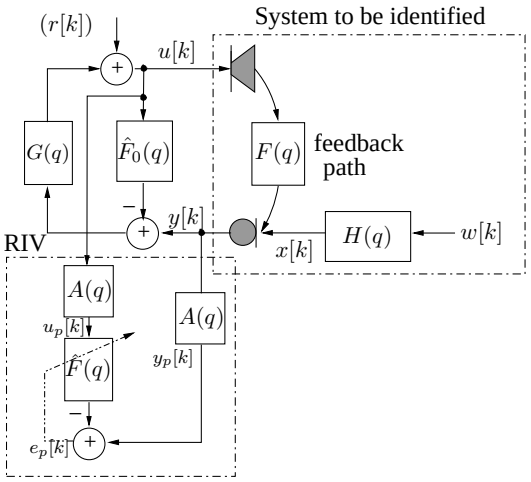


Figure 6.7: Instrumental variable method with pre-filtering of the data.

Unbiased feedback path estimate

If $A(q) = H^{-1}(q)$ and $w[k]$ is white, minimization of the LS criterion (6.79) or (6.90) results in an *unbiased* feedback path estimate $\hat{F}(q)$ [87, 88].

To illustrate this, consider the LS criterion (6.90). The coefficient vector $\hat{\mathbf{f}}[k]$ of $\hat{F}(q)$ that minimizes the LS is given by

$$\hat{\mathbf{f}}[k] = \left(\mathbf{U}_p^T[k]\mathbf{U}_p[k]\right)^{-1}\mathbf{U}_p^T[k]\mathbf{y}_p[k]. \tag{6.94}$$

Assume the sufficient order case, i.e., $L_{\hat{F}} = L_F$, and assume that $F(q)$ is time-invariant. Then, for $\lambda = 1$, (6.94) equals

$$\hat{\mathbf{f}}[k] = \mathbf{f} + \underbrace{\left(\mathbf{U}_p^T[k]\mathbf{U}_p[k]\right)^{-1}\mathbf{U}_p^T[k]\mathbf{x}_p[k]}_{\text{bias}} \tag{6.95}$$

with $x_p[k] = A(q)x[k]$ and

$$\mathbf{x}_p[k] \;=\; \left[\; x_p[k] \quad x_p[k-1] \quad \cdots \quad x_p[0] \;\right]^T, \tag{6.96}$$

If $A(q) = H^{-1}(q)$, $x_p[k]$ equals $w[k]$ and hence, if $w[k]$ is white and $G(q)$ contains a delay $d_f \geq 1$, $u_p[k]$ and $x_p[k]$ are uncorrelated, so that the bias term in (6.94) goes to zero for $k \to \infty$.

In practice, however, $H(q)$ is unknown and time-varying so that it is desirable to estimate $H(q)$ adaptively. In Chapter 7, we develop adaptive feedback cancellation techniques that are based on direct closed-loop identification of

both the feedback path and the desired signal model, with the aim of providing an unbiased feedback path estimate for unknown and time-varying desired signals $x[k]$.

6.5 Conclusion

In this chapter, we have given an overview of the existing feedforward suppression and, in particular, feedback cancellation techniques. The most promising approach to reduce acoustic feedback is the use of a continuous adaptation feedback (CAF) canceller. However, the adaptive modelling problem encountered here appears to be highly non-trivial, because of the presence of a closed signal loop, introducing specific signal correlation. As a result, the standard continuous adaptation feedback cancellers (CAF) fail to provide a reliable feedback path estimate, when the desired signal is spectrally colored.

Existing techniques to reduce the bias of the standard CAF have been discussed. Classical approaches include introducing signal decorrelating operations in the forward path, inserting a probe signal into the loudspeaker and controlling the adaptation of the feedback canceller. In all of these techniques, the closed signal loop is initially ignored and the standard adaptive filtering techniques for open-loop systems are applied. In a second stage, signal processing operations are included to compensate for the negative effect of the presence of the closed signal loop on the performance.

An attractive alternative is to regard the feedback path as a part of a closed-loop system and to apply closed-loop system identification theory. The three basic closed-loop system identification procedures, i.e., the indirect method, the joint input-output method and the direct method, have been discussed in the context of the feedback cancellation problem. Among the three methods, especially the direct method seems appealing for feedback cancellation, since it does neither rely on any assumptions about the forward path $G(q)$ (e.g., linear) nor on the use of an excitation signal $r[k]$. In [87, 88, 89], the feedback path $F(q)$ has been identified using the direct method and a *fixed* estimate of the desired signal model $H(q)$: it is shown that the bias of the feedback path estimate can be avoided by means of a filtered-X algorithm, provided that the desired signal model $H(q)$ is known. In practice, the desired signal model $H(q)$ is unknown and highly time-varying so that it is desirable to estimate this model adaptively. In Chapter 7, we develop adaptive feedback cancellation techniques that are based on direct closed-loop identification of the feedback path as well as the desired signal model, with the aim of providing an unbiased feedback path estimate for unknown and time-varying desired signals $x[k]$.

6.A Appendix: Continuous-adaptation feedback canceller

In this appendix, we demonstrate that the RLS based CAF with $\lambda = 1$ and $L_{\hat{F}} \geq L_F$ (i.e., sufficient order case) converges to a solution $\hat{F}(q)$ for which the closed-loop system is stable.

Lemma 6.A.1 *Consider the RLS based CAF (see (6.26)-(6.27)) with forgetting factor $\lambda = 1$ and assume the sufficient order case, i.e., $L_{\hat{F}} \geq L_F$.*

If the desired signal $x[k]$ is mean square bounded, i.e.,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x^2[k] < C \tag{6.A.1}$$

and $\mathcal{E}\{\mathbf{u}[k]\mathbf{u}^T[k]\}$ is non-singular. Then, the a posteriori error $\bar{\epsilon}[k]$, defined as $\bar{\epsilon}[k] = y[k] - \hat{\mathbf{f}}[k]\mathbf{u}[k]$is bounded, i.e.,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \bar{\epsilon}^2[k] < C. \tag{6.A.2}$$

Proof : Using $y[k] = x[k] + \mathbf{f}^T\mathbf{u}[k]$ and $L_{\hat{F}} \geq L_F$, (6.26)-(6.27) can be written as

$$\begin{aligned} \epsilon[k] &= x[k] - \boldsymbol{\theta}^T[k-1]\mathbf{u}[k] && (6.A.3)\\ \mathbf{R}[k] &= \lambda\mathbf{R}[k-1] + \mathbf{u}[k]\mathbf{u}^T[k] && (6.A.4)\\ \boldsymbol{\theta}[k] &= \boldsymbol{\theta}[k-1] + \mathbf{R}^{-1}[k]\mathbf{u}[k]\epsilon[k] && (6.A.5) \end{aligned}$$

with

$$\boldsymbol{\theta}[k] = \hat{\mathbf{f}}[k] - \mathbf{f}[k] \quad , \tag{6.A.6}$$

From (6.A.5),

$$\begin{aligned} \boldsymbol{\theta}^T[k]\mathbf{R}[k]\boldsymbol{\theta}[k] &= \boldsymbol{\theta}^T[k-1]\lambda\mathbf{R}[k-1]\boldsymbol{\theta}[k-1] + \left(\boldsymbol{\theta}^T[k-1]\mathbf{u}[k]\right)^2 \\ &\quad +2\epsilon[k]\mathbf{u}^T[k]\boldsymbol{\theta}[k-1] + \epsilon^2[k]\mathbf{u}^T[k]\mathbf{R}^{-1}[k]\mathbf{u}[k]. \end{aligned} \tag{6.A.7}$$

Using $\lambda = 1$ and (6.A.3), we find

$$\begin{aligned} \boldsymbol{\theta}^T[k]\mathbf{R}[k]\boldsymbol{\theta}[k] &= \boldsymbol{\theta}^T[k-1]\mathbf{R}[k-1]\boldsymbol{\theta}[k-1] + x^2[k] \\ &\quad -\epsilon^2[k]\left(1 - \mathbf{u}^T[k]\mathbf{R}^{-1}[k]\mathbf{u}[k]\right). \end{aligned} \tag{6.A.8}$$

Summing from $k = 1$ to $k = N$ gives

$$\boldsymbol{\theta}^T[N]\mathbf{R}[N]\boldsymbol{\theta}[N] = \boldsymbol{\theta}^T[0]\mathbf{R}[0]\boldsymbol{\theta}[0] + \sum_{k=1}^{N} x^2[k] - \sum_{k=1}^{N} \epsilon^2[k]\left(1 - \mathbf{u}^T[k]\mathbf{R}^{-1}[k]\mathbf{u}[k]\right). \tag{6.A.9}$$

Using (6.A.5), we find that the a posteriori residual $\bar{\epsilon}[k]$ equals

$$\bar{\epsilon}[k] \quad = \quad x[k] - \boldsymbol{\theta}^T[k]\mathbf{u}[k] = \epsilon[k]\left(1 - \mathbf{u}^T[k]\mathbf{R}^{-1}[k]\mathbf{u}[k]\right). \tag{6.A.10}$$

From (6.A.10), (6.A.4) and the matrix inversion lemma [69], i.e.,

$$1 - \mathbf{u}^T[k]\mathbf{R}^{-1}[k]\mathbf{u}[k] = 1 - \mathbf{u}^T[k]\left(\mathbf{R}[k-1] + \mathbf{u}[k]\mathbf{u}^T[k]\right)^{-1}\mathbf{u}[k], \tag{6.A.11}$$

$$= \left(1 + \mathbf{u}^T[k]\mathbf{R}^{-1}[k-1]\mathbf{u}[k]\right)^{-1}, \tag{6.A.12}$$

with $\mathbf{R}[k]$ non-singular, we obtain

$$\bar{\epsilon}[k] = \frac{\epsilon[k]}{1 + \mathbf{u}^T[k]\mathbf{R}^{-1}[k-1]\mathbf{u}[k]}. \tag{6.A.13}$$

Using (6.A.10) and (6.A.13) in (6.A.9), we get:

$$\sum_{k=1}^{N} \bar{\epsilon}^2[k]\left(1 + \mathbf{u}^T[k]\mathbf{R}^{-1}[k-1]\mathbf{u}[k]\right) + \boldsymbol{\theta}^T[N]\mathbf{R}[N]\boldsymbol{\theta}[N]$$

$$= \boldsymbol{\theta}^T[0]\mathbf{R}[0]\boldsymbol{\theta}[0] + \sum_{k=1}^{N} x^2[k]. \tag{6.A.14}$$

Hence, if $\mathbf{R}[k]$ is non-singular and $x[k]$ is mean square bounded, $\bar{\epsilon}[k]$ is mean square bounded [152, Chapter 4]. From (6.A.13) and (6.A.14), it follows that also the a priori residual $\epsilon[k]$ is bounded. □

If the input $u[k]$ to the adaptive filter is computed as $u[k] = G(q)\bar{\epsilon}[k] + r[k]$, it follows from lemma 6.A.1 that $u[k]$ is mean-square bounded if both the desired signal $x[k]$ and the probe signal $r[k]$ are bounded. Consequently, the step size of the RLS based CAF (6.26)-(6.27) with $\lambda = 1$ will tend to zero and the filter $\hat{F}(q)$ will change more slowly and slowly in time. Hence, if $\hat{F}(q)$ would result in an unstable closed-loop system $C(q)$, the generation of $\bar{\epsilon}[k]$ would be unstable, thus contradicting the lemma. Hence, the filter $\hat{F}(q)$ will converge to a solution for which the closed-loop system $C(q)$ is stable.

Chapter 7

Adaptive feedback cancellation with linear prediction of the desired signal

IN this chapter, we propose adaptive feedback cancellation techniques that are based on direct closed-loop identification of the feedback path as well as the (autoregressive) modelling of the desired signal, with the aim of providing an unbiased feedback path estimate for unknown and time-varying desired signals. In general, both models are not simultaneously identifiable in the closed-loop system at hand [89]. We show that, under certain conditions, e.g., if a delay is inserted in the forward path, identification of both models is indeed possible. Simulation results demonstrate that the proposed techniques outperform the standard CAF algorithm and the filtered-X algorithm of [88, 89] that uses a fixed estimate of the average desired signal spectrum, if the conditions for identifiability are satisfied.

In Section 7.2, we derive conditions under which the feedback path and the desired signal model can be both identified with the direct method of closed-loop identification. In [89], it has been shown that both models are not simultaneously identifiable in closed-loop unless a probe signal $r[k]$ is added to the loudspeaker signal or non-linearities are added to the forward path $G(q)$. We demonstrate that also adding a delay to the forward path $G(q)$ or the cancellation path $\hat{F}(q)$ can render the system identifiable. The required delay is found to be much smaller than the delay that is needed in the standard CAF to decorrelate $x[k]$ and $u[k]$ (cf. Section 6.3.1).

In Section 7.3, two classes of adaptive procedures for identifying the desired signal model $H(q)$ and the feedback path $F(q)$ are derived, namely, a two-channel identification method (2-ch AFC) as well as a prediction error method (PEM-AFC). Since most audio signals $x[k]$ such as a speech signal can be approximated by a short all-pole model (cf. Section 6.2.3), we assume an AR model $H(q)$ for $x[k]$ in the identification procedures. In contrast to the 2-ch AFC, the PEM-AFC allows the use of different adaptation speeds for the feedback path and for the desired signal model and hence, is found to be preferable for highly non-stationary desired signals. We also show that the filtered-X based feedback canceller with adaptive decorrelation filters of [147], which has been discussed in Section 6.3.3, can be interpreted as a PEM-AFC. Hence, to guarantee an unbiased feedback path estimate, identifiability conditions should be satisfied. In [147], no conditions for identifiability were discussed.

Simulations in Section 7.4 demonstrate that when the conditions for identifiability are fulfilled, the proposed recursive techniques outperform the standard CAF algorithm, and that for highly time-varying desired signals such as speech, the PEM-AFC offers the best performance and has a clear benefit over the filtered-X algorithm of [88, 89] that uses a *fixed* estimate of the average desired signal spectrum.

7.1 Introduction

In Section 6.4, the three basic closed-loop system identification procedures, i.e., the indirect method, the joint input-output method and the direct method, have been discussed in the context of feedback cancellation: among the three methods, especially the direct method (cf. Section 6.4.3) seems appealing for feedback cancellation. In [89], an adaptive feedback cancellation technique based on the direct method of closed-loop identification and a fixed estimate of the desired signal model has been proposed: it has been shown that the bias of the feedback path estimate can be avoided by means of the PEM or a filtered-X algorithm provided that the desired signal $x[k]$ can be modelled as $H(q)w[k]$ with $w[k]$ white noise, and the desired signal model $H(q)$ is inversely stable and known.

In practice, the desired signal model $H(q)$ is unknown and highly time-varying. In addition, the quality of the filtered-X algorithm strongly depends on the accuracy of the signal model estimate [89] so that it is desirable to identify both the feedback path $F(q)$ and the desired signal model $H(q)$ with the direct PEM method of closed-loop identification. In general though, the feedback path $F(q)$ and the desired signal model $H(q)$ are not both identifiable in closed-loop, unless a probe signal $r[k]$ is added to the loudspeaker signal $u[k]$ or non-linearities are added to $G(q)$ [89]. This implies that besides the desired

solution $\hat{F}(q) = F(q)$ and $\hat{H}(q) = H(q)$, alternative spurious solutions for $\hat{F}(q)$ and $\hat{H}(q)$ may exist. Hence, convergence of the adaptive filters $\hat{F}(q)$ and $\hat{H}(q)$ towards the desired solution is not guaranteed. The unidentifiability is caused by the forward path $G(q)$, which induces a linear relationship between the loudspeaker signal $u[k]$ and the microphone signal $y[k]$.

In this chapter, we revise the conditions for identifiability of $F(q)$ and $H(q)$ with the direct PEM of closed-loop identification (cf. Section 6.4.3). We demonstrate that besides inserting a probe signal or adding non-linearities to the forward path $G(q)$, also adding a delay to the forward path $G(q)$ or the cancellation path can render the system identifiable. Two classes of adaptive feedback cancellation techniques that are based on direct closed-loop identification of the feedback path as well as the desired signal model are developed, namely, a 2-ch AFC as well as a PEM-AFC. In the derivation of the identifiability conditions, we assume that the desired signal $x[k]$ is generated by passing white noise $w[k]$ through an all-pole filter $H(q)$. If $x[k]$ is a speech signal, however, the excitation of voiced speech segments corresponds to an impulse train, so that the performance of the algorithms is expected to be affected. Simulations show that also for real speech signals the PEM-AFC has a clear benefit over the standard CAF and the filtered-X algorithm [88, 89] that uses a *fixed* estimate of the average speech spectrum.

7.2 Conditions for identifiability

In this section, we derive conditions under which both the inverse desired signal model $H^{-1}(q)$ and the feedback path $F(q)$ can be identified by minimizing the PEM cost function (6.78) of the direct closed-loop identification method described in Section 6.4.3.

Like in Section 6.4, we assume that the forward path $G(q)$ contains a delay $d_f \geq 1$ and we first assume that the desired signal $x[k] = H(q)w[k]$, where $w[k]$ is white noise and $H(q)$ is monic and inversely stable. Since most audio signals $x[k]$ can be approximated by a short AR model (cf. Section 6.2.3), we assume that

$$H^{-1}(q) = 1 + q^{-1}P(q) \tag{7.1}$$

with $P(q)$ an FIR filter.

To derive the conditions for identifiability, we convert the non-linear PEM cost function (6.78) into a linear cost function by replacing $\hat{H}^{-1}(q)$ by $A(q)$ and the product $-\hat{H}^{-1}(q)\hat{F}(q)$ by $B(q)$, where the adaptive FIR filters $A(q)$ and $B(q)$ have coefficient vectors $\mathbf{a}$ and $\mathbf{b}$ and filter lengths L_A and L_B, respectively. This results in the two-channel identification method depicted in **Figure 7.1**.

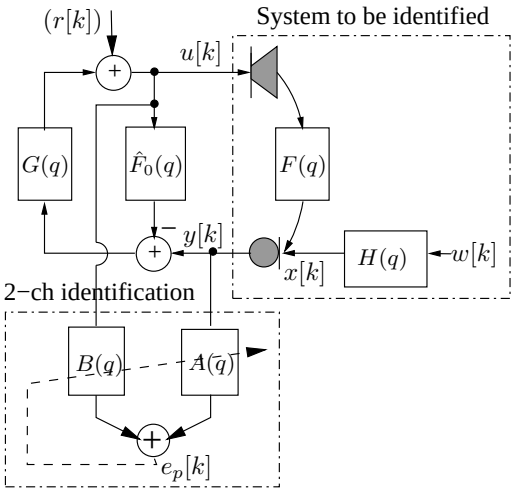


Figure 7.1: Two-channel identification method.

The two-channel adaptive filter $\{A(q),\, B(q)\}$ minimizes

$$\frac{1}{N}\sum_{k=0}^{N-1} e_p^2[k], \tag{7.2}$$

with

$$e_p[k] = B(q)u[k] + A(q)y[k]. \tag{7.3}$$

To avoid the trivial solution $A(q) = B(q) = 0$, the first tap a_0 of $A(q)$ is set to one, i.e., $\boxed{A(q) = 1 + q^{-1}\bar{A}(q)}$. We would like the filter $B(q)$ to identify the product $-H^{-1}(q)F(q)$ and the filter $A(q)$ to identify $H^{-1}(q)$ so that $e_p[k]$ equals $w[k]$. The feedback path estimate $\hat{F}(q)$ then corresponds to the IIR-filter $-A^{-1}(q)B(q)$. Hence, the filter $A^{-1}(q)$ should be constrained to be stable.

Remark 7.1 Again, the filter $\hat{F}_0(q)$ in the feedback cancellation path (cf. Figure 7.1) is an initial estimate of $F(q)$ such that the closed-loop system is stable (cf. Section 6.4.1). The filter $\hat{F}_0(q)$ is typically replaced during adaptation by the estimate $\hat{F}(q) = -A^{-1}(q)B(q)$ of $F(q)$. The data $u[k]$ and $y[k]$ used for identification, then depend on previous values of $A(q)$ and $B(q)$. To guarantee convergence of the adaptive filters $A(q)$ and $B(q)$, a boundness condition must be satisfied (cf. Section 6.2.2 and Section 7.3). □

Recall from Section 6.4.1 that the inputs $u[k]$ and $y[k]$ of the two-channel adaptive filter are given as

$$u[k] = C(q)x[k] + S(q)r[k], \tag{7.4}$$

$$y[k] = (1 + F(q)C(q))\, x[k] + F(q)S(q)r[k], \tag{7.5}$$

where $r[k]$ is the optional probe signal (noise) injected at the input of the loudspeaker, $S(q)$ is the sensitivity function and $C(q)$ the closed-loop transfer function from $x[k]$ to $u[k]$ defined in (6.68)-(6.69). The probe signal $r[k]$ is generally a noise signal. Using (7.4)-(7.5) and (6.68)-(6.69), the output $e_p[k]$ of the two-channel adaptive filter equals

$$\boxed{\begin{aligned} e_p[k] &= e_x[k] + e_r[k] && (7.6)\\ e_x[k] &= \left(A(q) + \frac{(B(q)+A(q)F(q))G(q)}{1-G(q)(F(q)-\hat{F}_0(q))}\right)x[k] && (7.7)\\ e_r[k] &= \frac{B(q)+A(q)F(q)}{1-G(q)(F(q)-\hat{F}_0(q))}r[k] && (7.8)\end{aligned}}$$

Below, we study conditions under which minimization of (7.2), has the unique solution

$$A(q) = H^{-1}(q), \quad B(q) = -H^{-1}(q)F(q), \tag{7.9}$$

so that $-A^{-1}(q)B(q)$ is an unbiased estimate of the feedback path $F(q)$. The conditions for identifiability also apply to the non-linear PEM criterion (6.77)-(6.78) in Section 6.4.3, provided that $H(q)$ is stable (so that $H^{-1}(q)$ does not have any zeros on the unit circle). To analyze (7.6)-(7.8), we distinguish between two cases: $r[k] = 0$ (no noise injection) and $r[k] \neq 0$ (noise injection). In the derivation of the identifiability conditions, we assume that $G(q) = q^{-d_f}\bar{G}(q)$ with $d_f \geq 1$ and that $\bar{G}(q)$, $H^{-1}(q)$, $F(q)$ and $\hat{F}_0(q)$ are causal. In addition, we assume the sufficient order case for $A(q)$ and $B(q)$, i.e., $\boxed{L_A \geq L_{H^{-1}}}$ and $\boxed{L_B \geq L_{H^{-1}} + L_F - 1}$, so that $A(q)$ and $B(q)$ can model the desired solution (7.9). Section 7.2.3 considers the case of undermodelling.

7.2.1 Case I: no noise injection

If $r[k] = 0$, minimization of (7.2) reduces to minimization of

$$\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k] \tag{7.10}$$

with $e_x[k]$ defined in (7.7).

Lemma 7.1 *Assume that $x[k] = H(q)w[k]$ with $w[k]$ white noise and $H^{-1}(q) = 1 + q^{-1}P(q)$ an FIR filter. Then, minimization of (7.10) with $A(q) = 1 + q^{-1}\bar{A}(q)$, results in the unique desired solution (7.9) if at least one of the following conditions is satisfied:*

- The forward path $G(q)$ has a delay $\boxed{d_f \geq L_A}$;

- The cancellation path $B(q)$ has a delay d_c with $\boxed{d_f + d_c \geq L_A}$, i.e., $B(q) = q^{-d_c}\bar{B}(q)$ and the feedback path $F(q)$ has a delay of at least d_c samples, i.e., $\boxed{F(q) = q^{-d_c}\bar{F}(q)}$.
- $\bar{G}(q)$ or $\hat{F}_0(q)$ are time-varying or $\bar{G}(q)$ is nonlinear.

Proof : Since $G(q) = q^{-d_f}\bar{G}(q)$ with $d_f \geq 1$ and $\bar{G}(q)$, $F(q)$, $\hat{F}_0(q)$, $A(q)$ and $B(q)$ are causal, $e_x[k]$ can be written as

$$e_x[k] = \big(A(q) + q^{-d_f}\Gamma(q)\big)\, x[k], \tag{7.11}$$

where

$$\Gamma(q) = \frac{\bar{G}(q)(B(q) + A(q)F(q))}{1 - q^{-d_f}\bar{G}(q)(F(q) - \hat{F}_0(q))} \tag{7.12}$$

is a causal IIR filter, which may be specified as $\Gamma(q) = \gamma_0 + q^{-1}\gamma_1 + \ldots$. Since $d_f \geq 1$, the first tap of $A(q) + q^{-d_f}\Gamma(q)$ in (7.11) equals the first tap $a_0 = 1$ of $A(q)$ and hence, minimization of $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k]$ corresponds to a linear prediction of $x[k]$.

If $x[k] = H(q)w[k]$ with $w[k]$ white noise, the optimal solution corresponds to $[A(q) + q^{-d_f}\Gamma(q)]x[k] = H^{-1}(q)x[k]$ or

$$\boxed{A(q) + q^{-d_f}\frac{\bar{G}(q)(B(q) + A(q)F(q))}{1 - q^{-d_f}\bar{G}(q)(F(q) - \hat{F}_0(q))} = H^{-1}(q).} \tag{7.13}$$

In general, $A(q)$ and $B(q)$ are not uniquely determined by equating coefficients of equal powers of q on both sides in (7.13). Below, we derive conditions under which the solution to (7.13) is unique and equals the desired solution (7.9).

Delay d_f in the forward path $G(q)$

If the delay $d_f \geq L_A \geq L_{H^{-1}}$, (7.13) has the unique solution

$$A(q) = H^{-1}(q) \quad \text{and} \quad \Gamma(q) = 0. \tag{7.14}$$

Since $a_0 = 1$, the equation $\Gamma(q) = 0$ is satisfied if and only[1] if $B(q) = -A(q)F(q)$, resulting in the desired solution (7.9).

Delay d_c in the cancellation path

If the cancellation path $B(q)$ has a delay d_c, i.e., $B(q) = q^{-d_c}\bar{B}(q)$ and the feedback path $F(q)$ has a delay of at least d_c samples, i.e., $F(q) = q^{-d_c}\bar{F}(q)$,

[1]Note that this assumes that $G(\omega) \neq 0$ for all $\omega \in [0, \pi]$. At angular frequencies ω where $G(\omega) = 0$, the feedback path cannot be identified. However, at those frequencies, no instability can occur and hence no feedback cancellation is required.

(7.13) can be expressed as

$$A(q) + q^{-d_f-d_c}\bar{\Gamma}(q) = A(q) + q^{-d_f-d_c}\frac{\bar{G}(q)(\bar{B}(q)+A(q)\bar{F}(q))}{1-q^{-d_f}\bar{G}(q)(F(q)-\hat{F}_0(q))}$$
$$= H^{-1}(q). \tag{7.15}$$

If $d_c + d_f \geq L_A \geq L_{H^{-1}}$, the solution of (7.15) is unique and equals

$$A(q) = H^{-1}(q); \qquad \bar{B}(q) = -A(q)\bar{F}(q). \tag{7.16}$$

Time-varying $\hat{F}_0(q)$, $\bar{G}(q)$ or nonlinear $\bar{G}(q)$

In general, (7.13) implies that there are several solutions for $A(q)$ and $B(q)$. If $\bar{G}(q)$ or $\hat{F}_0(q)$ are time-varying, equation (7.13) and hence, the positions of the spurious solutions will change with time. Several, different equations (7.13) then have to be satisfied, one for each time segment, so that it is likely that, with sufficient averaging, the algorithm will converge to the desired solution. Hence, if $\hat{F}_0(q)$ is at each time instant replaced by the most recent estimate of $-A^{-1}(q)B(q)$, the adaptive algorithm may converge to the desired solution, even without adding a delay in the forward path.
A nonlinear $\bar{G}(q)$ reduces the correlation between $x[k]$ and $S(q)\bar{G}(q)x[k]$ so that it decouples the minimization of $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k]$ into a minimization of $\frac{1}{N}\sum_{k=0}^{N-1}\left(A(q)x[k]\right)^2$ and $\frac{1}{N}\sum_{k=0}^{N-1}\left[\left(B(q)+A(q)F(q)\right)x[k]\right]^2$. If $w[k]$ is white noise, both terms are minimized if and only if $A(q)$ and $B(q)$ equal the desired solution (7.9). □

Lemma 7.1 shows that adding a delay d_f to the forward path $G(q)$ and/or a delay d_c to the cancellation path with $d_t = d_f + d_c \geq L_A$ renders the system $\{F(q),\, H^{-1}(q)\}$ identifiable. Adding a delay d_c to the cancellation path only results in an unbiased feedback path estimate if the first d_c taps of $F(q)$ are zero. A delay in the feedback path may arise from the propagation time of the sound from the loudspeaker to the microphone and the processing delay of the DAC and ADC converters. Since most audio signals such as speech, can be modelled by an all-pole model $H(q)$, the filter $H^{-1}(q)$ is usually much shorter than $H(q)$. Hence, the minimum required delay $d_t = d_f + d_c = L_{H^{-1}}$ is smaller than the delay that is needed in the standard CAF to decorrelate $x[k]$ and $u[k]$ (cf. 6.3.1).

Remark 7.2 The above derived identifiability conditions assume that the excitation $w[k]$ of the AR-model $H(q)$ is white noise. This assumption does not apply for periodic signals $x[k]$ such as a pure sinusoid or a voiced speech segment, where the frequency spectrum consists of discrete pulses. For periodic signals, the filter $A(q) + q^{-d_f}\Gamma(q)$ (cf. (7.13)) tries to cancel the discrete frequency pulses. As a result, the feedback path $F(q)$ cannot be identified if $r[k] = 0$.

To illustrate this, consider, e.g., a voiced speech signal $x[k]$. The excitation $w[k]$ of voiced speech resembles a periodic impulse train with pitch period P, so that a voiced speech signal $x[k]$ can be reasonably approximated by a periodic signal with period P and hence, by a sum of harmonics [179]:

$$x[k] = \sum_{i=1}^{L_h} c_i \cos(2\pi i \frac{f_0}{f_s} k + \phi_i), \tag{7.17}$$

where c_i and ϕ_i denote the amplitude and phase of the i-th harmonic, $f_0 = \frac{1}{P}$ is the pitch frequency and L_h is the number of harmonics given by $L_h = \left\lfloor \frac{f_s}{2f_0} \right\rfloor$ with $f_s = 16$ kHz the sampling frequency. For this periodic signal $x[k]$, the error $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k]$ equals 0 iff

$$A(q) + q^{-d_f}\Gamma(q) = Q(q)\Pi_{i=1}^{L_h}(1 - e^{j2\pi i \frac{f_0}{f_s}} q^{-1})(1 - e^{-j2\pi i \frac{f_0}{f_s}} q^{-1}), \tag{7.18}$$

with $Q(q)$ any random polynomial. Hence, the filter $A(q) + q^{-d_f}\Gamma(q)$ will not model $H^{-1}(q)$ but will try to cancel the impulse train by putting notches at the normalized frequencies $\frac{if_0}{f_s}$ with $i \in \mathbb{N}$ and $\frac{if_0}{f_s} \leq \frac{1}{2}$ [179]. Consequently, the feedback path $F(q)$ cannot be identified. □

7.2.2 Case II: noise injection

Lemma 7.2 *Assume that $x[k] = H(q)w[k]$ where $w[k]$ is white noise. Then, minimization of (7.2) results in the unique desired solution (7.9) if $r[k] \neq 0$ and $r[k]$ and $x[k]$ are uncorrelated.*

Proof: If $r[k] \neq 0$ and if $r[k]$ and $x[k]$ are uncorrelated, the cost function $\frac{1}{N}\sum_{k=0}^{N-1} e_p^2[k]$ equals

$$\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k] + \frac{1}{N}\sum_{k=0}^{N-1} e_r^2[k], \tag{7.19}$$

with $e_x[k]$ and $e_r[k]$ defined in (7.7)-(7.8). If there exist an $A(q)$ and $B(q)$ that minimize both $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k]$ and $\frac{1}{N}\sum_{k=0}^{N-1} e_r^2[k]$ (which is the case if $L_B \geq L_{H^{-1}} + L_F - 1$ and $L_A \geq L_{H^{-1}}$), then these $A(q)$ and $B(q)$ are the solution to the minimization of $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k] + \frac{1}{N}\sum_{k=0}^{N-1} e_r^2[k]$.
The term $\frac{1}{N}\sum_{k=0}^{N-1} e_r^2[k]$ is minimized iff

$$\boxed{B(q) + A(q)F(q) = 0,} \tag{7.20}$$

which results in a true estimate $-A^{-1}(q)B(q)$ of the feedback path provided that $A^{-1}(q)$ is stable. If $w[k]$ is white noise and $G(q)$ contains a delay $d_f \geq 1$,

the term $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k]$ is minimized if (7.13) is satisfied (cf. the proof of Lemma 7.1). Both (7.13) and (7.20) are fulfilled iff

$$\begin{aligned} A(q) &= H^{-1}(q) \\ B(q) &= H^{-1}(q)F(q). \end{aligned} \tag{7.21}$$

Hence, the solution is unique and equals the desired solution. □

Note that if the level of the probe signal $r[k]$ is too low compared to the level of $x[k]$, the error energy $\frac{1}{N}\sum_{k=0}^{N-1} e_r^2[k]$ is negligible compared to $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k]$. If the identifiability conditions derived in Section 7.2.1 are not satisfied, several solutions $A(q)$ and $B(q)$ exist if $r[k] = 0$ so that, for $r[k] \to 0$, the identification problem may become ill-conditioned. The level of $r[k]$ thus determines the condition of the identification problem and hence the convergence.

7.2.3 Biased feedback path estimate in case of undermodelling

Section 7.2.1 shows that adding a delay to the forward path $G(q)$ and/or the cancellation path can render the system identifiable without adding a probe signal $r[k]$ or non-linearities. If the total delay $d_t = d_f + d_c$ with d_f the delay in the forward path and d_c the common delay in the feedback $F(q)$ and cancellation path $B(q)$, exceeds L_A and if $L_A \geq L_{H^{-1}}$ and $L_B \geq L_{H^{-1}} + L_F - 1$, an unbiased estimate $-A^{-1}(q)B(q)$ is obtained. In practice, the filter length L_{H-1} of the inverse speech model $H^{-1}(q)$ and the filter length L_F of the feedback path $F(q)$ are unknown, so that undermodelling may occur, i.e., $L_A < L_{H^{-1}}$ or/and $L_B < L_A + L_F - 1$. In this case, the solution to (7.13) is still unique if $d_t \geq L_A$ but biased. In this section, we show that, like for the standard CAF [221] or the filtered-X algorithm that uses a fixed estimate of the desired signal model $H(q)$ [89], the amount of bias depends on the amplitude characteristic of the closed-loop system, the delay d_t and –for $r[k] \neq 0$ with $r[k]$ and $x[k]$ uncorrelated– the power spectral density of the probe signal $r[k]$.

- The larger the amplitude characteristic $|C(\omega)|$ of the closed-loop system $C(q) = S(q)G(q)$ at angular frequency ω, the smaller the bias of the feedback path estimate at this angular frequency ω. Note that the sensitivity function $S(\omega)$ (cf. (6.68)) is especially large at frequencies with a small gain margin. Hence, the bias of the feedback path estimate will be smallest at the critical frequencies, which is a desired property.

  Proof: Consider minimization of (7.10). The term $(B(q) + A(q)F(q))$ in $e_x[k]$ is weighted by the closed-loop system $C(q) = S(q)G(q)$ (cf. (7.7)). Hence, the larger the amplitude characteristic $|C(\omega)| = |S(\omega)G(\omega)|$, the smaller the bias in the feedback path estimate $\hat{F}(q) = -A^{-1}(q)B(q)$ at angular frequency ω. □

- Increasing the delay d_t for a given L_A will in general reduce the bias if $w[k]$ is white.

  Proof : If $w[k]$ is white noise, $q^{-d_t}w[k]$ is uncorrelated with $q^{-l}w[k]$ for $l = 0, \cdots, d_t - 1$, so that minimization of $\frac{1}{N}\sum_{k=0}^{N-1} e_x^2[k]$ corresponds to a minimization of

  $$\frac{1}{N}\sum_{k=0}^{N-1} e_a^2[k] + \frac{1}{N}\sum_{k=0}^{N-1} e_b^2[k], \tag{7.22}$$

  with

  $$e_a[k] = [A(q)H(q)]_{1:d_t}\, w[k], \tag{7.23}$$
  $$e_b[k] = [A(q)H(q)]_{d_t+1:end}\, w[k] + C(q)\,(B(q) + A(q)F(q))\,H(q)w[k], \tag{7.24}$$

  where $[A(q)H(q)]_{1:d_t}$ is the filter whose impulse response corresponds to the first d_t coefficients of $A(q)H(q)$ and has zeros everywhere else and where $[A(q)H(q)]_{d_t+1:end}$ is the filter whose impulse response has zeros in the first d_t positions, but otherwise is the same as the impulse response $A(q)H(q)$. The adaptive filter $B(q)$ attempts to minimize $\frac{1}{N}\sum_{k=0}^{N-1} e_b^2[k]$: if $[A(q)H(q)]_{d_t+1:end} \neq 0$, the feedback path estimate $\hat{F}(q) = -A^{-1}(q)B(q)$ is biased. Since in most cases the impulse response of $H(q)$ and hence, $A(q)H(q)$ decreases with increasing lag, increasing the delay d_t for a given, fixed L_A will generally reduce the term $[A(q)H(q)]_{d_t+1:end}\, w[k]$ and hence, the bias of the feedback path estimate $\hat{F}(q)$ [89]. □

- Inserting a probe signal $r[k]$ with $r[k]$ and $x[k]$ uncorrelated, reduces the bias of the feedback path estimate: the larger the power spectral density $|R(\omega)|^2$ with $R(\omega)$ the spectrum of $r[k]$, the smaller the bias at angular frequency ω.

  Proof : Increasing $|R(\omega)|^2$ puts more weight on the error term $\frac{1}{N}\sum_{k=0}^{N-1} e_r^2[k]$ in (7.19) at angular frequency ω and hence, reduces the bias of the feedback path estimate $\hat{F}(q) = -A^{-1}(q)B(q)$ at angular frequency ω. □

7.3 Recursive identification

Section 7.2 shows that under certain conditions the feedback path $F(q)$ and the inverse signal model $H^{-1}(q)$ are both identifiable even if no additional noise $r[k]$ is injected in the system and no non-linearities are inserted in $G(q)$. Inserting, e.g., a large enough delay d_f (i.e., $d_f \geq L_A$) in the forward path $G(q)$ renders

the system identifiable. Inserting a delay d_c in the cancellation path may only result in an unbiased solution if the first d_c taps of $F(q)$ are also zero.

In this section, two different recursive procedures for estimating the feedback path $F(q)$ and the desired signal model $H(q)$ are proposed based on the two-channel identification method of Section 7.2 and the PEM of Section 6.4.3, resulting in a two-channel adaptive feedback canceller (2-ch AFC) and an PEM based adaptive feedback canceller (PEM-AFC).

The *two-channel identification method* of Figure 7.1 has the advantage of having a linear cost function. However, to obtain the feedback path estimate $\hat{F}(q)$, the adaptive filter $B(q)$ has to be multiplied by $-A^{-1}(q)$. Hence, a stability constraint should be imposed on the adaptive filter $A^{-1}(q)$. The *prediction error method* estimates $F(q)$ and $H^{-1}(q)$ instead of $-H^{-1}(q)F(q)$ and $H^{-1}(q)$, but has a non-linear cost function. Filtering with $A^{-1}(q)$ is then not required and shorter adaptive filters can be used. In practice, the desired signal $x[k]$ is often highly time-varying, while the feedback path varies more slowly in time. In contrast to the 2-ch AFC, the PEM-AFC allows to use a different adaptation scheme or adaptation speed for estimating the feedback path $F(q)$ and the inverse speech model $H^{-1}(q)$ and hence, is found to be preferable for highly non-stationary desired signals. It will also be shown that the LMS based feedback canceller in [147] can be interpreted as an implementation of the PEM with a decoupled adaptation of $F(q)$ and $H^{-1}(q)$.

7.3.1 Two-channel adaptive feedback canceller

At each time instant k, the two-channel adaptive filter $\begin{bmatrix} \mathbf{a}^T & \mathbf{b}^T \end{bmatrix}^T$ in Figure 7.1 minimizes the cost function

$$J_k(\mathbf{a},\mathbf{b}) = \sum_{\bar{k}=0}^{k} \lambda^{k-\bar{k}} e_p^2[\bar{k}], \tag{7.25}$$

with $e_p[\bar{k}] = B(q)u[\bar{k}] + A(q)y[\bar{k}]$ (see (7.3)) and where $\lambda \in (0,1]$ is a forgetting factor to allow for tracking of time-varying scenarios. To avoid the trivial solution $\mathbf{a} = \mathbf{0}$, $\mathbf{b} = \mathbf{0}$, the first coefficient a_0 of $\mathbf{a}$ is set to one, resulting in

$$e_p[\bar{k}] = y[\bar{k}] + \mathbf{b}^T\mathbf{u}[\bar{k}] + \bar{\mathbf{a}}^T\bar{\mathbf{y}}[\bar{k}], \tag{7.26}$$

with

$$\bar{\mathbf{a}}^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_{L_A-1} \end{bmatrix}, \tag{7.27}$$

$$\mathbf{u}[\bar{k}] = \begin{bmatrix} u[\bar{k}] & u[\bar{k}-1] & \cdots & u[\bar{k}-L_B+1] \end{bmatrix}^T \tag{7.28}$$

$$\bar{\mathbf{y}}[\bar{k}] = \begin{bmatrix} y[\bar{k}-1] & y[\bar{k}-2] & \cdots & y[\bar{k}-L_A+1] \end{bmatrix}^T \tag{7.29}$$

The cost function (7.25) is *linear* in the filters $\bar{\mathbf{a}}$ and $\mathbf{b}$ so that standard recursive algorithms such as (modified) LMS [71] and RLS can be used:

$$\begin{bmatrix} \bar{\mathbf{a}}[k] \\ \mathbf{b}[k] \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{a}}[k-1] \\ \mathbf{b}[k-1] \end{bmatrix} - \mathbf{R^{-1}}[k] \begin{bmatrix} \bar{\mathbf{y}}[k] \\ \mathbf{u}[k] \end{bmatrix} \epsilon_p[k], \tag{7.30}$$

where $\epsilon_p[k]$ is the a priori residual, defined as

$$\begin{aligned} \epsilon_p[k] &= A(q,k-1)y[k] + B(q,k-1)u[k] \\ &= y[k] + \mathbf{b}^T[k-1]\mathbf{u}[k] + \bar{\mathbf{a}}^T[k-1]\bar{\mathbf{y}}[k], \end{aligned} \tag{7.31}$$

and where the choice for $\mathbf{R}[k]$ depends on the algorithm that is being used:

- For *LMS-updating,*

  $$\mathbf{R}^{-1}[k] = \rho[k]\mathbf{I}_{L_A+L_B-1}$$

  where $\rho[k]$ is the step size. Again, for highly non-stationary sound signals, better performance is achieved with a time-varying step size $\rho[k]$ based on the sum-method (cf. Section 6.2.2):

  $$\rho[k] = \frac{\bar{\rho}}{(L_A-1)p_{\bar{y}}(k) + L_B p_u[k] + (L_A-1+L_B)p_{\epsilon_p}[k] + c}, \tag{7.32}$$

  where c is a small positive constant and

  $$p_u[k] = \gamma p_u[k-1] + (1-\gamma)u^2[k], \tag{7.33}$$
  $$p_{\bar{y}}[k] = \gamma p_{\bar{y}}[k-1] + (1-\gamma)y^2[k-1], \tag{7.34}$$
  $$p_{\epsilon_p}[k] = \gamma p_{\epsilon_p}[k-1] + (1-\gamma)\epsilon_p^2[k], \tag{7.35}$$

  with the forgetting factor $\gamma \in [0,1)$ set to $\gamma = 1 - \frac{1}{L_A+L_B-1}$.

- For *RLS-updating*,

  $$\mathbf{R}[k] = \lambda\mathbf{R}[k-1] + \begin{bmatrix} \bar{\mathbf{y}}[k] \\ \mathbf{u}[k] \end{bmatrix} \begin{bmatrix} \bar{\mathbf{y}}^T[k] & \mathbf{u}^T[k] \end{bmatrix}. \tag{7.36}$$

In the simulations, RLS will be used to update the adaptive filters $\bar{\mathbf{a}}$ and $\mathbf{b}$. The implementation of the 2-ch AFC is summarized in the tabulated **Algorithm 7.1**. For reasons of efficiency, the matrix $\mathbf{R}^{-1}[k]$ in the RLS algorithm is recursively updated using the matrix inversion lemma (cf. Section 6.2.2). For practical computation, numerically stable RLS implementations based on a QR decomposition should be used [85, 173] (cf. Section 6.2.2).

Below, we consider two different scenarios for the 2-ch AFC, i.e., the case where the feedback canceller $\hat{F}_0(q)$ is kept fixed during adaptation and the case where $\hat{F}_0(q)$ is updated during adaptation.

Algorithm 7.1 Implementation of the 2-ch AFC.

Initialization:

$$\bar{\mathbf{a}}[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{R}^{(L_A-1)\times 1}; \quad \mathbf{b}[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{R}^{L_B \times 1}$$

For RLS: $\mathbf{R}^{-1}[0] = \frac{1}{c}\mathbf{I}_{L_A+L_B-1}$ with c a small positive constant (e.g., $c = 10^{-8}$)

For LMS: $p_u[0] = p_{\bar{y}}[0] = p_{\epsilon_p}[0] = c$ with c a small positive constant

For each time instant $k = 0, \ldots, \infty$**:**

$$u[k] = G(q)\left(y[k] - \hat{F}_0(q)u[k]\right) + r[k]$$

$$\mathbf{u}[k] = \begin{bmatrix} u[k] & u[k-1] & \cdots & u[k-L_B+1] \end{bmatrix}^T$$

$$\bar{\mathbf{y}}[k] = \begin{bmatrix} y[k-1] & y[k-2] & \cdots & y[k-L_A+1] \end{bmatrix}^T$$

$$\epsilon_p[k] = y[k] + \mathbf{b}^T[k-1]\mathbf{u}[k] + \bar{\mathbf{a}}^T[k-1]\bar{\mathbf{y}}[k]$$

Update equation:

- RLS

$$\begin{bmatrix} \bar{\mathbf{a}}[k] \\ \mathbf{b}[k] \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{a}}[k-1] \\ \mathbf{b}[k-1] \end{bmatrix} - \mathbf{R}^{-1}[k] \begin{bmatrix} \bar{\mathbf{y}}[k] \\ \mathbf{u}[k] \end{bmatrix} \epsilon_p[k]$$

$$\mathbf{R}^{-1}[k] = \frac{1}{\lambda}\mathbf{R}^{-1}[k-1] - \frac{1}{\lambda^2}\frac{\mathbf{R}^{-1}[k-1]\begin{bmatrix} \bar{\mathbf{y}}[k] \\ \mathbf{u}[k] \end{bmatrix}\begin{bmatrix} \bar{\mathbf{y}}^T[k] & \mathbf{u}^T[k] \end{bmatrix}\mathbf{R}^{-1}[k-1]}{1 + \frac{1}{\lambda}\begin{bmatrix} \bar{\mathbf{y}}^T[k] & \mathbf{u}^T[k] \end{bmatrix}\mathbf{R}^{-1}[k-1]\begin{bmatrix} \bar{\mathbf{y}}[k] \\ \mathbf{u}[k] \end{bmatrix}}$$

- LMS

$$\begin{bmatrix} \bar{\mathbf{a}}[k] \\ \mathbf{b}[k] \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{a}}[k-1] \\ \mathbf{b}[k-1] \end{bmatrix} - \rho[k] \begin{bmatrix} \bar{\mathbf{y}}[k] \\ \mathbf{u}[k] \end{bmatrix} \epsilon_p[k]$$

with

$$\rho[k] = \frac{\bar{\rho}}{(L_A-1)p_{\bar{y}}(k) + L_B p_u[k] + (L_A-1+L_B)p_{\epsilon_p}[k] + c}$$

$$\begin{aligned} p_u[k] &= \gamma p_u[k-1] + (1-\gamma)u^2[k] \\ p_{\bar{y}}[k] &= \gamma p_{\bar{y}}[k-1] + (1-\gamma)y^2[k-1] \\ p_{\epsilon_p}[k] &= \gamma p_{\epsilon_p}[k-1] + (1-\gamma)\epsilon_p^2[k] \end{aligned}$$

Feedback canceller $\hat{F}_0(q)$ is kept fixed during adaptation

If the filter $\hat{F}_0(q)$ in the feedback cancellation path is kept fixed during adaptation and the closed-loop system $C(q)$ is stable, the input data $u[k]$ and $y[k]$ are bounded so that convergence of the adaptive filters $\bar{\mathbf{a}}$ and $\mathbf{b}$ is guaranteed provided that the correlation matrix $\mathbf{R}[k]$ of the input data is non-singular (cf. Section 6.2.2).

The feedback path estimate $\hat{F}(q)$ corresponds to the IIR-filter $-A^{-1}(q)B(q)$. To guarantee a stable feedback path estimate, a stability constraint on $A^{-1}(q)$ is thus required. The adaptive filters are therefore only updated when the update results in a stable filter $A^{-1}(q)$. A simple (conservative) test for stability is to check that $\sum_{i=1}^{L_A-1}|a_i|$ is less than 1 [218]. All unstable updates will be detected by this approach, but the test may indicate that $A^{-1}(q)$ is unstable when it is not. For a stable model $H^{-1}(q) = 1+q^{-1}P(q)$ with $\sum_{i=0}^{L_P-1}|p_i| \geq 1$, convergence of $A(q)$ to the optimal solution $H^{-1}(q)$ may then never be reached. A better, but computationally more complex approach, is to test after each update that the zeros of $A(q)$ lie inside the unit circle by using the Shur-Cohn test [218].

Feedback canceller $\hat{F}_0(q)$ is updated during adaptation

In practice, the filter $\hat{F}_0(q)$ in the feedback cancellation path (see Figure 7.1) will be replaced during adaptation with new estimates $\hat{F} = -A^{-1}(q)B(q)$ of the feedback path $F(q)$. Typically, the feedback canceller $\hat{F}_0(q)$ is updated at the adaptation rate of the filters $A(q)$ and $B(q)$ (i.e., after each time instant k), but it may also be updated at a lower rate, e.g., after each T time instants. The input-output data $u[k]$, $y[k]$ used for the identification is then determined by the adaptive filters $\mathbf{a}[\bar{k}]$ and $\mathbf{b}[\bar{k}]$ with $\bar{k} \leq k-1$.

Example 7.1 If $\hat{F}_0(q)$ is updated at the adaptation rate of $A(q)$ and $B(q)$, $u[k]$ can be computed as:

$$u[k] = r[k] + G(q)A^{-1}(q,k-1)\epsilon_p[k] \tag{7.37}$$

with $\epsilon_p[k]$ the a priori residual defined in (7.31) or as

$$u[k] = r[k] + G(q)A^{-1}(q,k)\bar{\epsilon}_p[k] \tag{7.38}$$

with $\bar{\epsilon}_p[k]$ the a posteriori residual, defined as

$$\bar{\epsilon}_p[k] = A(q,k)y[k] + B(q,k)u[k]. \tag{7.39}$$

Since $\bar{\epsilon}_p[k] \leq \epsilon_p[k]$, (7.38) is expected to result in better performance than (7.37), but requires more computations. For reasons of efficiency, we use (7.38).

To guarantee convergence of the adaptive filters $\mathbf{a}$ and $\mathbf{b}$, a boundness condition must now be satisfied, namely, the input-output data $u[k]$, $y[k]$ should be bounded. Therefore, the filters $\mathbf{a}[\bar{k}]$ and $\mathbf{b}[\bar{k}]$ (with $\bar{k} = n \cdot T$ ($n \in \mathbb{N}$) and $\bar{k} \leq k$) should not result in an unstable closed-loop system $C(q)$ for a significant length of time (cf. Section 6.2.2). △

From simulations, we observed that when $\hat{F}_0(q)$ is continuously being updated with $\hat{F}(q)$, also the 2-ch AFC has a self-stabilizing feature, whereby unstable poles of the closed-loop system $C(q)$ have a tendency to migrate back into the stable region (cf. the self-stabilizing property of the standard CAF in Section 6.2.2). Since a stable $C(q)$ implies a stable feedback path estimate $\hat{F}(q, k-1)$, a stability constraint on $A(q)$ is then not required. The self-stabilizing feature can be understood intuitively by considering the prediction error $e_p[k]$, which, for $u[k]$ computed as in (7.37), equals (7.6)-(7.8) with

$$\hat{F}_0(q) = -A^{-1}(q, k-1)B(q, k-1). \tag{7.40}$$

If the feedback path estimate $\hat{F}(q, k-1) = -A^{-1}(q, k-1)B(q, k-1)$ at time instant $k-1$ results in a closed-loop system

$$S\,(q, k-1) = \frac{1}{1 - G(q)\left(F(q) - \hat{F}(q, k-1)\right)} \tag{7.41}$$

that is close to instability, the error $B(q) + A(q)F(q)$ in the feedback path estimate gets a large weight $S\,(q, k-1)$ in the prediction error $e_p[k]$, so that the next estimate $\hat{F}(q, k) = -A(q, k)^{-1}B(q, k)$ will get close to $F(q)$ and thus stabilizes the closed-loop system $C(q) = G(q)S(q)$ again. As a result of the self-stabilizing property, $u[k]$ and $y[k]$ remain bounded and $A(q, k)$ and $B(q, k)$ are found to converge to the desired solution if the identifiability conditions are satisfied. In addition, faster convergence may be obtained than when $\hat{F}_0(q)$ is kept fixed. E.g., if the correlation matrix $\mathbf{R}[k]$ is initialized with a very small identity matrix, the initial updates of the 2-ch AFC may be quite large. Because of the presence of the near-end signal $x[k]$, these initial updates are quite noisy so that the closed-loop system may become unstable. Simulations indicated that, if this occurs, $\hat{F}(q)$ sometimes immediately converges to the true feedback path $F(q)$.

7.3.2 PEM based adaptive feedback canceller

The 2-ch AFC of Section 7.3.1 estimates the product $-F(q)H^{-1}(q)$ and the inverse signal model $H^{-1}(q)$. Hence, a joint adaptation algorithm and/or adaptation speed is required for $F(q)$ and $H^{-1}(q)$. In practice, the desired signal $x[k]$ is, e.g., a speech signal and thus strongly time-varying: the inverse speech model $H^{-1}(q)$ is only stationary during approximately 20 ms. To obtain a

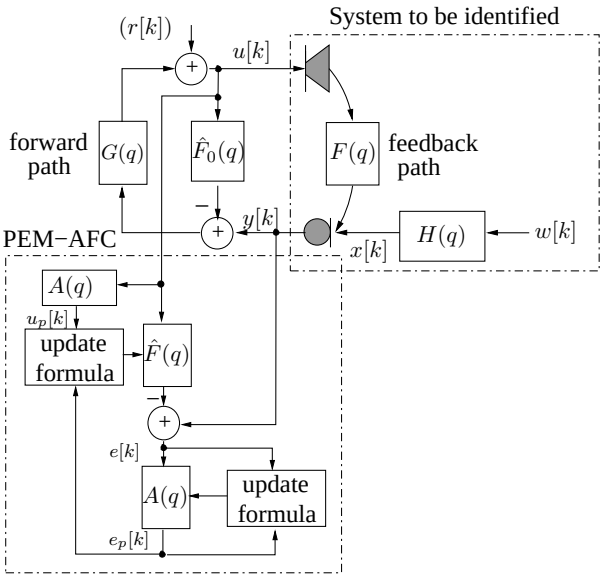


Figure 7.2: PEM-AFC.

good estimate of the instantaneous speech model, fast convergence is required for the filter $A(q)$. However, the identification of the feedback path is strongly degraded by the presence of the desired signal $x[k]$ because of its high-energy with respect to the feedback signal $F(q)u[k]$. As a consequence, a large amount of data samples N (or a slow convergence time) is required to produce a reliable feedback path estimate (i.e., N corresponds to a duration $\gg 20$ ms), especially for small $|G(q)F(q)|$. Therefore, it seems preferable to use a different adaptation scheme or adaptation speed for the estimation of the slowly time-varying feedback path $F(q)$ on the one hand and the estimation of the highly time-varying inverse speech model $H^{-1}(q)$ on the other hand. Instead of estimating the product $-F(q)H^{-1}(q)$ and $H^{-1}(q)$, the non-linear PEM directly estimates $F(q)$ and $H^{-1}(q)$. By introducing some approximations in the Hessian, the joint prediction error update equation for estimating $F(q)$ and $H^{-1}(q)$ decouples into two separate update equations for the two models, making a different adaptation possible.

PEM based adaptive feedback canceller (PEM-AFC)

Figure 7.2 illustrates the PEM-AFC: the filter $A(q)$ estimates $H^{-1}(q)$ and the filter $\hat{F}(q)$ estimates $F(q)$. At each time instant k, the coefficient vectors $\mathbf{a}$ and $\hat{\mathbf{f}}$ of $A(q)$ and $\hat{F}(q)$, respectively, defined as

$$\mathbf{a} = \begin{bmatrix} 1 & \bar{\mathbf{a}} \end{bmatrix}^T; \; \hat{\mathbf{f}} = \begin{bmatrix} \hat{f}_0 & \hat{f}_1 & \cdots & \hat{f}_{L_{\hat{F}}-1} \end{bmatrix}^T, \tag{7.42}$$

minimize the PEM cost function $J_k(\bar{\mathbf{a}}, \hat{\mathbf{f}})$

$$J_k(\bar{\mathbf{a}}, \hat{\mathbf{f}}) = \sum_{\bar{k}=0}^{k} \lambda^{k-\bar{k}} e_p^2[\bar{k}], \tag{7.43}$$

where $\lambda \in (0, 1]$ is the forgetting factor and where $e_p[\bar{k}]$ is the prediction error at time instant $\bar{k}$

$$e_p[\bar{k}] = A(q) \left(y[\bar{k}] - \hat{F}(q) u[\bar{k}] \right) \tag{7.44}$$

The PEM cost function $J_k(\bar{\mathbf{a}}, \hat{\mathbf{f}})$ is *non-linear* in the parameters $\bar{\mathbf{a}}$ and $\hat{\mathbf{f}}$. Hence, local minima exist and convergence to the global minimum is not guaranteed. The recursive update equation at time instant k equals [152]

$$\begin{bmatrix} \bar{\mathbf{a}}[k] \\ \hat{\mathbf{f}}[k] \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{a}}[k-1] \\ \hat{\mathbf{f}}[k-1] \end{bmatrix} - \left[\mathbf{J}_k^{''}(\bar{\mathbf{a}}[k-1], \hat{\mathbf{f}}[k-1]) \right]^{-1} \frac{1}{2} \mathbf{J}_k^{'}(\bar{\mathbf{a}}[k-1], \hat{\mathbf{f}}[k-1]), \tag{7.45}$$

where $\mathbf{J}_k^{''}(\bar{\mathbf{a}}[k-1], \hat{\mathbf{f}}[k-1])$ is the Hessian and $\mathbf{J}_k^{'}(\bar{\mathbf{a}}[k-1], \hat{\mathbf{f}}[k-1])$ the gradient of the cost function $J_k(\bar{\mathbf{a}}, \hat{\mathbf{f}})$ computed using $\bar{\mathbf{a}}[k-1]$ and $\hat{\mathbf{f}}[k-1]$. The gradient $\mathbf{J}_k^{'}(\bar{\mathbf{a}}[k-1], \hat{\mathbf{f}}[k-1])$ equals

$$\mathbf{J}_k^{'}(\bar{\mathbf{a}}[k-1], \hat{\mathbf{f}}[k-1]) = 2 \begin{bmatrix} \frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]} \\ \frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} \end{bmatrix} \epsilon_p[k], \tag{7.46}$$

where $\epsilon_p[k]$ is the a priori prediction error

$$\epsilon_p[k] \quad = \quad A(q, k-1) \left(y[k] - \hat{F}(q, k-1) u[k] \right). \tag{7.47}$$

In practice, the Hessian $\mathbf{J}_k^{''}$ in (7.45) is approximated by a positive semi-definite matrix $\mathbf{R}[k]$. Possible choices for $\mathbf{R}[k]$ are

- the stochastic gradient method:

$$\mathbf{R}^{-1}[k] = \rho[k] \mathbf{I}_{L_A + L_{\hat{F}} - 1} \tag{7.48}$$

  with ρ the step size. Again, a time-varying step size $\rho[k]$ based on the sum-method is preferable if there are large fluctuations in the desired signal power.

- the stochastic Gauss-Newton method:

$$\mathbf{R}[k] = \lambda \mathbf{R}[k-1] + \begin{bmatrix} \frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]} \\ \frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} \end{bmatrix} \begin{bmatrix} \left(\frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]} \right)^T & \left(\frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} \right)^T \end{bmatrix}. \tag{7.49}$$

In the sequel, we focus on the stochastic Gauss-Newton method.

As in [152, chapter 2], the a priori residual $\epsilon_p[k]$ and the derivatives $\frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]}$ and $\frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]}$ are, for reasons of efficiency[2], approximated as:

$$\epsilon_p[k] \approx A(q, k-1)\epsilon[k] \tag{7.50}$$

$$\frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]} = \left[\begin{array}{cccc} \epsilon[k-1] & \epsilon[k-2] & \cdots & \epsilon[k-L_A+1] \end{array} \right]^T, \tag{7.51}$$

$$\frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} = \left[\begin{array}{cccc} u_p[k] & u_p[k-1] & \cdots & u_p[k-L_{\hat{F}}+1] \end{array} \right]^T, \tag{7.52}$$

where $\epsilon[k]$ and $u_p[k]$ are computed as:

$$\epsilon[k] = y[k] - \hat{F}(q, k-1)u[k] \tag{7.53}$$

$$u_p[k] = A(q, k-1)u[k]. \tag{7.54}$$

Note that this approximation assumes that $\bar{\mathbf{a}}[k-1] = \bar{\mathbf{a}}[k-\bar{k}_1]$ for $1 < \bar{k}_1 < L_{\hat{F}}$ and that $\hat{\mathbf{f}}[k-1] = \hat{\mathbf{f}}[k-\bar{k}_2]$ for $1 < \bar{k}_2 < L_A$. Since L_A is rather small (e.g., 20 taps), it is reasonable to expect that $\hat{\mathbf{f}}[k-1] \approx \hat{\mathbf{f}}[k-\bar{k}_2]$ for $1 < \bar{k}_2 < L_A$. Furthermore, in hearing aids, the dominant part of the true feedback path $F(q)$ is quite short, so that a relatively short filter length $L_{\hat{F}}$ is typically used [128], i.e., $L_{\hat{F}}/f_s < 20$ ms (i.e., smaller than the short-time stationarity of speech). Hence, we expect that $\bar{\mathbf{a}}[k-1] \approx \bar{\mathbf{a}}[k-\bar{k}_1]$ for $1 < \bar{k}_1 < L_{\hat{F}}$, even for highly time-varying speech signals $x[k]$.

Remark 7.3 For a non-linear cost function, the forgetting factor λ has an influence on the convergence time: a large λ results in a slower convergence than a smaller λ [152, chapter 5]. Initially, the estimates $\hat{\mathbf{f}}[k-1]$ and $\bar{\mathbf{a}}[k-1]$ in the gradient estimate (7.46) may be quite bad. Therefore, it is better to discount the initial prediction errors by using a time-varying forgetting factor $\lambda[k]$ that grows exponentially with k to a final λ_{final}, i.e.,

$$\lambda[k] = \lambda_0 \lambda[k-1] + (1-\lambda_0)\lambda_{\text{final}}, \tag{7.55}$$

where the rate λ_0, the initial value $\lambda[0]$ and the final value λ_{final} are design variables [152, chapter 5]. □

Remark 7.4 Like the 2-ch AFC, $u[k]$ and $y[k]$ should be bounded to guarantee convergence. Hence, if $\hat{F}_0(q)$ is updated during adaptation, the estimate $\hat{F}(q, \bar{k})$ with $\bar{k} \leq k$ should not result in an unstable closed-loop system $C(q)$ for a significant length of time. □

[2]Otherwise, $L_{\hat{F}} + L_A - 2$ additional filtering operations would be required, i.e., $A(q, k-1)u[k-i]$, $i = 1, \cdots, L_{\hat{F}} - 1$ and $\hat{F}(q, k-1)u[k-i]$, $i = 1, \cdots, L_A - 1$.

Decoupling the adaptation of $\hat{F}(q)$ and $A(q)$

The stochastic Gauss-Newton update equation (7.45) corresponds to a joint update equation for $\hat{\mathbf{f}}[k]$ and $\bar{\mathbf{a}}[k]$. Forcing the off-diagonal blocks of $\mathbf{R}[k]$ to zero decouples the update equations for the feedback path $\hat{\mathbf{f}}[k]$ and the noise model $\bar{\mathbf{a}}[k]$:

$$\begin{cases} \bar{\mathbf{a}}[k] = \bar{\mathbf{a}}[k-1] - \mathbf{R}_{\bar{\mathbf{a}}}^{-1}[k] \frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]} \epsilon_p[k] \\ \hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] - \mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k] \frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} \epsilon_p[k] \end{cases} \tag{7.56}$$

with

$$\begin{cases} \mathbf{R}_{\bar{\mathbf{a}}}[k] = \lambda_{\bar{\mathbf{a}}}[k] \mathbf{R}_{\bar{\mathbf{a}}}[k-1] + \frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]} \left(\frac{\partial \epsilon_p[k]}{\partial \bar{\mathbf{a}}[k-1]} \right)^T \\ \mathbf{R}_{\hat{\mathbf{f}}}[k] = \lambda_{\hat{\mathbf{f}}}[k] \mathbf{R}_{\hat{\mathbf{f}}}[k-1] + \frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} \left(\frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} \right)^T . \end{cases} \tag{7.57}$$

Since the block-diagonal approximation

$$\begin{bmatrix} \mathbf{R}_{\bar{\mathbf{a}}}[k] & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\hat{\mathbf{f}}}[k] \end{bmatrix} \tag{7.58}$$

of $\mathbf{R}[k]$ is positive semi-definite, (7.56) converges to a minimum (local or global) of $\mathbf{J}_k$, provided that the update equations (7.56) are stable[3] [152]. Algorithm (7.56)-(7.57) is expected to have a convergence rate in-between the convergence rate of the stochastic gradient algorithm and the stochastic Gauss-Newton algorithm. Different forgetting factors $\lambda_{\bar{\mathbf{a}}}[k]$ and $\lambda_{\hat{\mathbf{f}}}[k]$ or different adaptation schemes (e.g., stochastic Gauss-Newton or stochastic gradient adaptation) can now be used for $\bar{\mathbf{a}}$ and $\hat{\mathbf{f}}$. Because speech is short-term stationary for about $10 - 20$ ms, $\lambda_{\bar{\mathbf{a}}} = 1 - \frac{1}{160}$ seems a good choice for a sampling frequency $f_s = 16$ kHz.

Note that the update equation for the feedback path $\hat{\mathbf{f}}[k]$ corresponds to the filtered-X algorithm of Section 6.4.4. Hence, if there is a large group delay associated with the filter $A(q)$, instability may occur when the filter coefficients $\hat{\mathbf{f}}$ change too quickly in time [17]. In this case, a gain scalar $\alpha < 1$ should be added before $\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k]$ in (7.56) (cf. Section 6.4.4) and the initialization for $\mathbf{R}_{\hat{\mathbf{f}}}[k]$ should not be too small, at the expense of reduced convergence.

Remark 7.5 The PEM-AFC does not require a filtering with $A^{-1}(q)$ so that there is no need for a stability constraint on $A^{-1}(q)$. However, since the desired signal $x[k]$ is bounded, the desired signal model $H(q)$ is assumed to be stable or marginally stable. Hence, the performance of the non-linear PEM-AFC can be improved by limiting the search space for $A(q)$ to the inversely stable or marginally stable filters. This stability constraint can be easily imposed by using, e.g., the Burg lattice algorithm [161] instead of RLS for updating the AR-coefficients $\bar{\mathbf{a}}$. The Burg lattice algorithm, described by **Algorithm 7.2**,

[3] In the simulations, no instability occurred. If needed, instability can be avoided by including a small step size $\rho < 1$ in the update equations (7.56).

Algorithm 7.2 Burg lattice algorithm for computing $\kappa_i[k]$, $i = 1, \ldots, L_A - 1$.

Definitions:

$f_i[k]$, $i = 0, \ldots, L_A - 1$ = forward residuals
$b_i[k]$, $i = 0, \ldots, L_A - 1$ = backward residuals

Initialization:

$f_0[k] = \epsilon[k] = y[k] - \hat{\mathbf{f}}[k-1]\mathbf{u}[k]$;
$b_0[k] = \epsilon[k]$;

For $i = 1, \ldots, L_A - 1$:

$d_i[k] = \lambda_{\bar{\mathbf{a}}} d_i[k-1] + (1 - \lambda_{\bar{\mathbf{a}}}) \left[f_{i-1}^2[k] + b_{i-1}^2[k-1] \right]$
$n_i[k] = \lambda_{\bar{\mathbf{a}}} n_i[k-1] + (1 - \lambda_{\bar{\mathbf{a}}})(-2) f_{i-1}[k] b_{i-1}[k-1]$
$\kappa_i[k] = \frac{n_i[k]}{d_i[k]}$
$f_i[k] = f_{i-1}[k] + \kappa_i[k] b_{i-1}[k-1]$
$b_i[k] = \kappa_i[k] f_{i-1}[k] + b_{i-1}[k-1]$

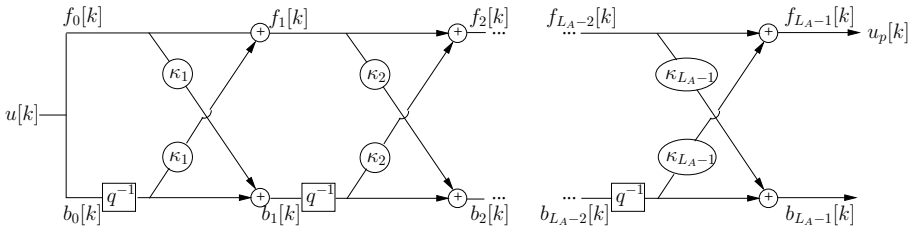


Figure 7.3: Filtering of $u[k]$ with the lattice filter $A(q)$ (with reflection coefficients κ_i, $i = 1, \ldots, L_A - 1$).

guarantees stability by ensuring that the reflection coefficients $\kappa_i \leq 1$, $i = 1, \cdots, L_A - 1$. The a priori prediction error $\epsilon_p[k]$ and the derivative $\frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]}$ (see (7.50)-(7.54)) in (7.56) are then obtained by filtering the feedback compensated signal $\epsilon[k]$ and the loudspeaker signal $u[k]$, respectively, through the lattice filter determined by the reflection coefficients $\kappa_i[k-1]$, $i = 1, \ldots, L_A - 1$ from the previous iteration. As an illustration, the filtering of $u[k]$ through the lattice filter $A(q)$ with reflection coefficients κ_i is depicted in **Figure 7.3**. In the simulations, we concentrate on this particular implementation of the PEM-AFC, referred to as decoupled (lattice) PEM-AFC. The implementation of the decoupled (lattice) PEM-AFC is summarized in the tabulated **Algorithm 7.3**. □

Remark 7.6 If the stochastic gradient method is used to update the feedback canceller $\hat{F}(q)$ (cf. Algorithm 7.3), the decoupled (lattice) PEM-AFC corresponds to the filtered-X based feedback canceller with adaptive decorrelation

Algorithm 7.3 Implementation of the decoupled (lattice) PEM-AFC .

Initialization:

Initialization $\hat{\mathbf{f}}$*:* $\hat{\mathbf{f}}[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{R}^{L_{\hat{F}} \times 1}$

For filtered-X RLS: $\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[0] = \frac{1}{c}\mathbf{I}_{L_{\hat{F}}}$ with c a small positive constant

For filtered-X LMS: $p[0] = c$ with c a small positive constant

Initialization $\bar{\mathbf{a}}$*:* reflection coefficients $\kappa_i[0] = 0$, $i = 1, \ldots, L_A - 1$

For each time instant $k = 0, \ldots, \infty$**:**

$$u[k] = G(q)\left(y[k] - \hat{F}_0(q)u[k]\right) + r[k]; \quad \epsilon[k] = y[k] - \hat{\mathbf{f}}^T[k-1]u[k]$$

Compute $u_p[k] = A(q, k-1)u[k]$:

Filter $u[k]$ through the lattice filter $\kappa_i[k-1]$, $i = 1, \ldots, L_A - 1$ (cf. Figure 7.3)

Compute $\epsilon_p[k] = A(q, k-1)\epsilon[k]$:

Filter $e[k]$ through the lattice filter $\kappa_i[k-1]$, $i = 1, \ldots, L_A - 1$ (cf. Figure 7.3)

Update equation for $\bar{\mathbf{a}}$*:*

Compute the reflection coefficients $\kappa_i[k]$, $i = 1, \ldots, L_A - 1$ with the Burg lattice algorithm (cf. Algorithm 7.2)

Update equation for $\hat{\mathbf{f}}$*:*

$$-\frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} = \mathbf{u}_p[k] = \begin{bmatrix} u_p[k] & u_p[k-1] & \cdots & u_p[k - L_{\hat{F}} + 1] \end{bmatrix}^T$$

- Filtered-X RLS (with $\alpha \in (0, 1]$) (stochastic Gauss-Newton method)

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \alpha \mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k]\mathbf{u}_p[k]\epsilon_p[k]$$

$$\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k] = \frac{1}{\lambda_{\hat{\mathbf{f}}}}\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1] - \frac{1}{\lambda_{\hat{\mathbf{f}}}^2}\frac{\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1]\mathbf{u}_p[k]\mathbf{u}_p^T[k]\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1]}{1 + \frac{1}{\lambda_{\hat{\mathbf{f}}}}\mathbf{u}_p^T[k]\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1]\mathbf{u}_p[k]}$$

- Filtered-X LMS (stochastic gradient method)

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \frac{\bar{\rho}}{L_{\hat{F}}p[k] + c}\mathbf{u}_p[k]\epsilon_p[k]$$

with

$$p[k] = \gamma p[k-1] + (1-\gamma)\left(u_p^2[k] + \epsilon_p^2[k]\right)$$

filters of [147] (cf. Section 6.3.3). The filtered-X LMS based feedback canceller of [147] can thus also be interpreted as an implementation of the (decoupled) PEM-AFC so that the conditions for identifiability derived in Section 7.2 also apply to [147]. In [147], however, no identifiability conditions were derived. □

Alternative decoupled adaptation with instrumental variable method.

Instead of adapting the feedback path estimate $\hat{\mathbf{f}}[k]$ with the filtered-X update equation, the RIV update equation (6.93) that has been described in Section 6.4.4, could be used. This amounts to computing the a priori prediction error $\epsilon_p[k]$ from the pre-filtered data $u_p[k]$ and $y_p[k]$ as:

$$\epsilon_p[k] = y_p[k] - \hat{F}(q, k-1)u_p[k], \tag{7.59}$$

instead of computing $\epsilon_p[k]$ by filtering the feedback-compensated signal $e[k]$ with $A(q, k-1)$. The implementation of this alternative decoupled adaptation scheme is described in the tabulated **Algorithm 7.4**. In contrast to the filtered-X update equation, any change to the adaptive filter $\hat{\mathbf{f}}$ now has an immediate effect on the prediction error $\epsilon_p[k]$, so that no stability problems may occur when $A(q)$ has a large group delay. To compute the feedback-compensated signal $e[k] = y[k] - \hat{F}(q)u[k]$, an additional filtering with $\hat{\mathbf{f}}$ is required. The algorithm (7.56)-(7.57) with (7.59) corresponds to the refined instrumental variable approximate maximum likelihood algorithm of [281].

Remark 7.7 The non-linear PEM cost function (7.43) is linear in $\mathbf{a}$ if $\hat{\mathbf{f}}$ is kept fixed and vice-versa, linear in $\hat{\mathbf{f}}$ if $\mathbf{a}$ is kept fixed. Hence, the cost function (7.43) corresponds to a separable non-linear least-squares problem. Alternative decoupled implementations of the PEM-AFC may be derived from the theory of separable non-linear least-squares problems [68, 181]. In separable non-linear least squares problems, one parameter set, e.g., $\mathbf{a}$, is estimated using a linear technique. The new estimate $\mathbf{a}$ is then used to update the other parameter set $\hat{\mathbf{f}}$ by means of a non-linear optimization technique. □

7.4 Simulation results

In this section, we compare the performance of the 2-ch AFC and the decoupled PEM-AFC with the standard CAF algorithm[4] [76, 221] (cf. Section 6.2) and the filtered-X algorithm with fixed $A(q)$ [88, 89] (cf. Section 6.4.4) for both stationary (cf. Section 7.4.2) and time-varying desired signals (cf. Section 7.4.3 and Section 7.4.4).

[4] I.e., the 2-ch AFC with fixed $A(q) = 1$.

Algorithm 7.4 Alternative decoupled adaptation with RIV method.

Initialization:

Initialization $\hat{\mathbf{f}}$*:* $\hat{\mathbf{f}}[0] = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{R}^{L_{\hat{F}} \times 1}$

For RLS: $\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[0] = \frac{1}{c}\mathbf{I}_{L_{\hat{F}}}$ with c a small positive constant

For LMS: $p[0] = c$ with c a small positive constant

Initialization $\bar{\mathbf{a}}$*:* reflection coefficients $\kappa_i[0] = 0$, $i = 1, \ldots, L_A - 1$

For each time instant $k = 0, \ldots, \infty$**:**

$$u[k] = G(q)\left(y[k] - \hat{F}_0(q)u[k]\right) + r[k]$$

Pre-filter the input-output data $u[k]$ *and* $y[k]$ *with* $A(q, k-1)$*:*

Compute $u_p[k] = A(q, k-1)u[k]$:

Filter $u[k]$ through the lattice filter $\kappa_i[k-1]$, $i = 1, \ldots, L_A - 1$ (cf. Figure 7.3)

Compute $y_p[k] = A(q, k-1)y[k]$:

Filter $y[k]$ through the lattice filter $\kappa_i[k-1]$, $i = 1, \ldots, L_A - 1$ (cf. Figure 7.3)

Update equation for $\bar{\mathbf{a}}$*:*

Compute the reflection coefficients $\kappa_i[k]$, $i = 1, \ldots, L_A - 1$ with the Burg lattice algorithm (cf. Algorithm 7.2)

Update equation for $\hat{\mathbf{f}}$*:*

$$-\frac{\partial \epsilon_p[k]}{\partial \hat{\mathbf{f}}[k-1]} = \mathbf{u}_p[k] = \begin{bmatrix} u_p[k] & u_p[k-1] & \cdots & u_p[k - L_{\hat{F}} + 1] \end{bmatrix}^T$$

$$\epsilon_p[k] = y_p[k] - \hat{\mathbf{f}}^T[k-1]\mathbf{u}_p[k]$$

- RLS

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k]\mathbf{u}_p[k]\epsilon_p[k]$$

$$\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k] = \frac{1}{\lambda_{\hat{\mathbf{f}}}}\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1] - \frac{1}{\lambda_{\hat{\mathbf{f}}}^2}\frac{\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1]\mathbf{u}_p[k]\mathbf{u}_p^T[k]\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1]}{1 + \frac{1}{\lambda_{\hat{\mathbf{f}}}}\mathbf{u}_p^T[k]\mathbf{R}_{\hat{\mathbf{f}}}^{-1}[k-1]\mathbf{u}_p[k]}$$

- LMS

$$\hat{\mathbf{f}}[k] = \hat{\mathbf{f}}[k-1] + \frac{\bar{\rho}}{L_{\hat{F}}p[k] + c}\mathbf{u}_p[k]\epsilon_p[k]$$

with

$$p[k] = \gamma p[k-1] + (1-\gamma)\left(u_p^2[k] + \epsilon_p^2[k]\right)$$

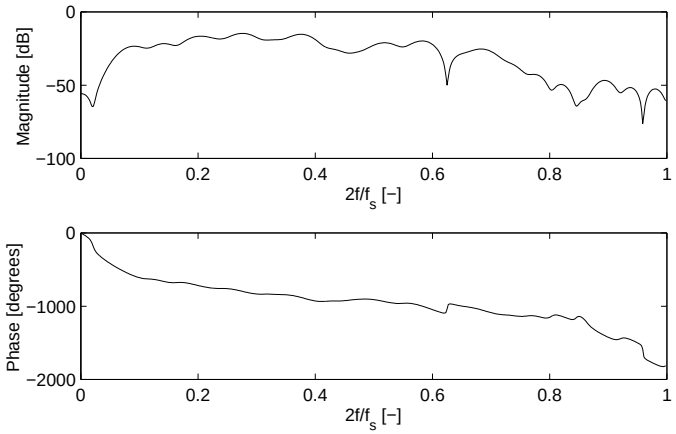


Figure 7.4: Frequency characteristic of the feedback path $F(q)$.

Among the different conditions for identifiability of $H(q)$ and $F(q)$, inserting a delay d_f in the forward path seems the preferred option. Inserting a delay d_c in the cancellation path only results in an unbiased feedback path estimate $\hat{F}(q)$ if the first d_c taps of $F(q)$ are also zero and adding non-linearities to $G(q)$ or inserting a probe signal $r[k]$ may degrade the sound quality. In the simulations we, therefore, concentrate on the influence of the delay d_f on the performance of the different algorithms. In the derivation of the identifiability conditions, we assumed that the desired signal $x[k]$ is generated by passing white noise $w[k]$ through an all-pole filter. If $x[k]$ is a speech signal, the excitation of voiced segments however corresponds to an impulse train, so that the performance of the feedback path cancellation techniques may be degraded. The importance of the white-noise assumption for the performance of the different techniques will be evaluated.

7.4.1 Set-up

In the simulations, the acoustic feedback path model $F(q)$ is an FIR filter with filter length $L_F = 50$, obtained from a measured feedback path. The frequency characteristic of $F(q)$ is depicted in **Figure 7.4**. The forward path $G(q)$ equals $q^{-d_f}|G|$ with $|G| = 5$. The sampling frequency $f_s = 16$ kHz. Without feedback cancellation, the MSG (cf. Section 6.1.4) equals 14.63 dB.

All feedback cancellers are implemented using RLS with $\lambda = 1$. In the PEM-AFC, the filter $A(q)$ is implemented by the Burg lattice algorithm with $\lambda_{\bar{\mathbf{a}}} = 1 - \frac{1}{160}$. In the simulations, the filter $\hat{F}_0(q)$ in feedback cancellation path is continuously updated during adaptation with the estimate $\hat{F}(q)$. Unless specified otherwise, the filter lengths L_A, L_B and $L_{\hat{F}}$ of the adaptive filters

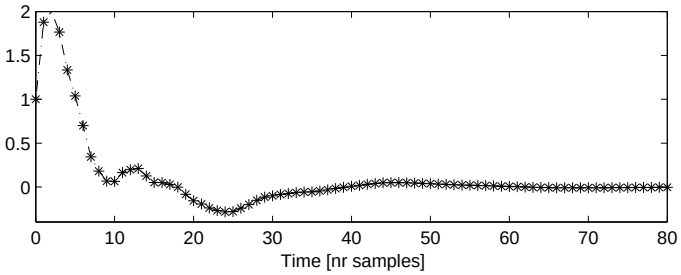


Figure 7.5: Impulse response of the all-pole filter $H(q)$.

$A(q)$ in the 2-ch AFC and PEM-AFC, $B(q)$ in the 2-ch AFC and $\hat{F}(q)$ in the CAF, the PEM-AFC and the filtered-X algorithm are set to $L_A = 21$, $L_B = L_A + L_F - 1$ and $L_{\hat{F}} = L_F = 50$.

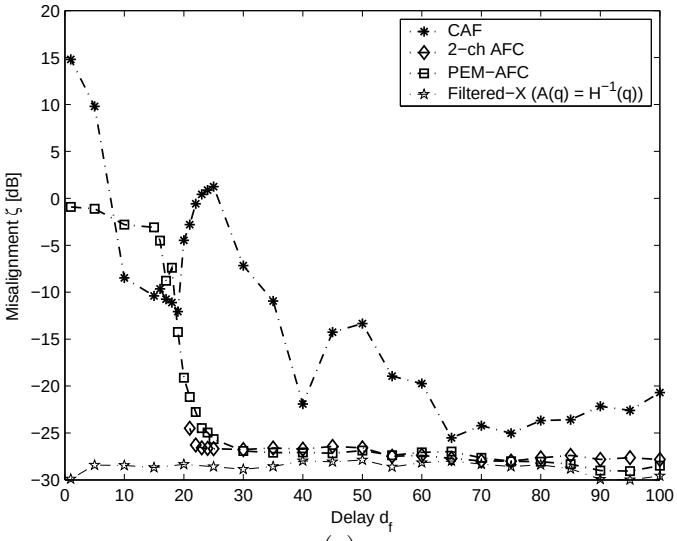
To assess the performance of the feedback cancellation algorithms we use the misalignment $\zeta(F(q), \hat{F}(q))$ between the true and estimated feedback path and the added stable gain ASG. For the definitions, we refer to Section 6.1.4.

7.4.2 Stationary desired signal $x[k]$

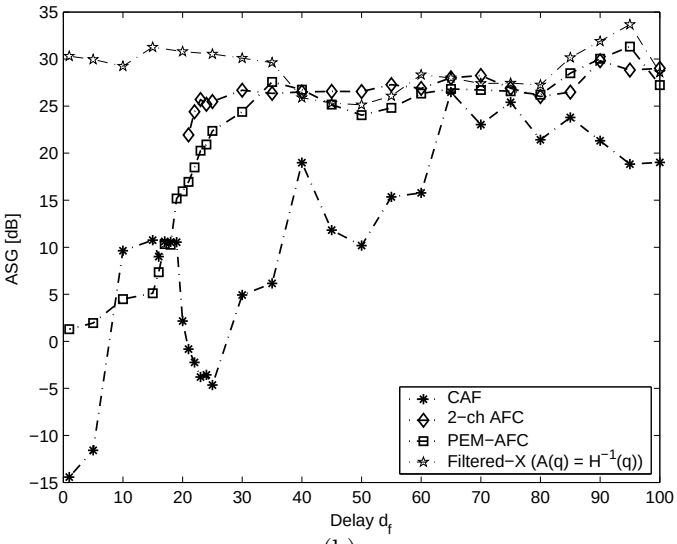
In this section, the performance of the 2-ch AFC and the decoupled PEM-AFC are evaluated for a stationary desired signal $x[k]$.

In a first experiment, the desired signal $x[k]$ is a *stationary speech-shaped noise signal* created by passing Gaussian noise through a 20-th order all-pole filter $H(q)$. The impulse response of $H(q)$ is depicted in **Figure 7.5**.

Figure 7.6 plots the misalignment $\zeta(F, \hat{F})$ and the ASG of the CAF, the 2-ch AFC and the (decoupled) PEM-AFC as a function of the delay d_f in the forward path. For comparison, the misalignment $\zeta(F, \hat{F})$ of the feedback path estimate obtained with filtered-X RLS using the correct speech model, i.e., $A(q) = H^{-1}(q)$, which we consider in some sense an optimal solution, is depicted too. The filter lengths of the adaptive filters are set to the true model orders, i.e. $L_A = L_{H^{-1}} = 21$, $L_B = L_{H^{-1}} + L_F - 1$ in the 2-ch AFC and $L_{\hat{F}} = L_F$ in the CAF, the PEM-AFC and the optimal filtered-X algorithm. Figure 7.6(a)-(b) illustrate the condition for identifiability: if the delay $d_f \geq L_A$, the 2-ch AFC and the PEM-AFC obtain an unbiased feedback path estimate. The 2-ch AFC and PEM-AFC perform nearly as well as the optimal filtered-X algorithm and clearly outperform the CAF. If $d_f < L_A$, the feedback path $F(q)$ and the model $H^{-1}(q)$ are not both identifiable: multiple spurious solutions for the feedback path $F(q)$ then exist. If $d_f < L_A$, the correlation matrix $\mathbf{R}[k]$ (cf. (7.36)) that is used in the update equation of the 2-ch AFC is non-singular and hence, instability occurs. The performance of the CAF improves with increasing delay



(a)



(b)

Figure 7.6: (a) Misalignment $\zeta(F, \hat{F})$ and (b) ASG of CAF, 2-ch AFC, PEM-AFC and filtered-X with $A(q) = H^{-1}(q)$ as a function of the delay d_f.

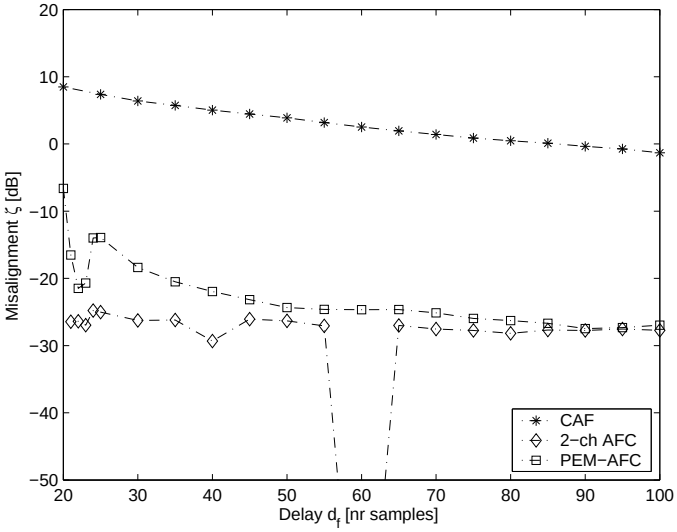


Figure 7.7: Misalignment $\zeta(F, \hat{F})$ of CAF, 2-ch AFC and PEM-AFC as a function of the delay d_f for a more tonal desired signal $x[k]$.

d_f. For $d_f > 60$, the correlation between $x[k]$ and $x[k-d_f]$ is quite small. Note that the impulse response of $H(q)$ (see Figure 7.5) has most of it energy at the first 60 samples.

For a more tonal signal $x[k]$, the performance of the CAF improves less with increasing delay d_f. This is depicted in **Figure 7.7**, which shows the misalignment $\zeta(F, \hat{F})$ of the CAF, the 2-ch AFC and the PEM-AFC for a desired signal $x[k]$ that is generated by passing white, Gaussian noise through the filter $H(q)$

$$H(q) = \frac{1}{(1 - 0.99e^{-j2\pi 0.2}q^{-1})(1 - 0.99e^{j2\pi 0.2}q^{-1})}. \tag{7.60}$$

The filter $\hat{F}_0(q)$ is continuously updated during adaptation. Because of the self-stabilizing feature of the 2-ch AFC, the true feedback path $F(q)$ may be immediately found when the noisy initial estimates $\hat{F}(q)$ result in an unstable closed-loop system (cf. Section 7.3.1). This explains the good performance for $d_f = 60$.

7.4.3 Synthetic time-varying desired signal $x[k]$

In practice, the desired signal $x[k]$ will often be a speech signal and hence, highly time-varying. In the first example, we generate a *synthetic speech signal* by passing white, Gaussian noise through a highly time-varying AR-filter, so that

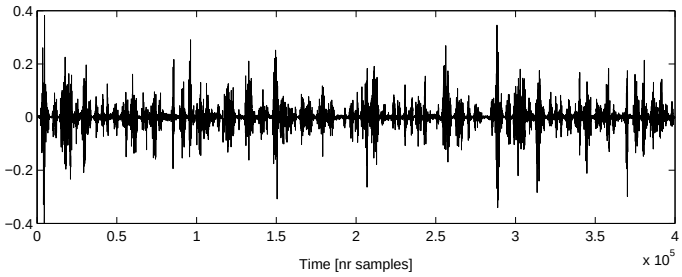


Figure 7.8: Synthetic speech signal $x[k]$.

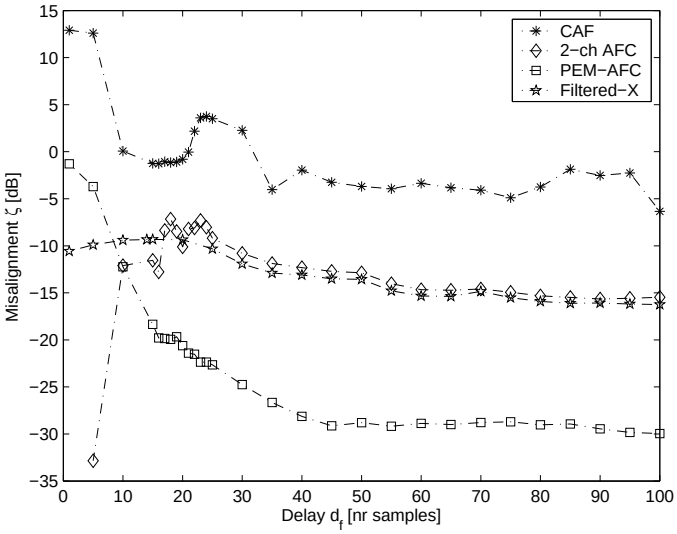
the white noise assumption for $w[k]$ is still satisfied. The AR-filter coefficients have been derived from the reflection coefficients of an adaptive lattice filter with $\lambda_{\tilde{\mathbf{a}}} = 1 - \frac{1}{160}$ operated on a real speech signal. The synthetic signal $x[k]$ is depicted in **Figure 7.8**.

Figure 7.9 depicts the misalignment $\zeta(F, \hat{F})$ and the ASG of the different feedback cancellation algorithms. For comparison, the filtered-X algorithm using the average speech model,i.e., an AR model of the long-term average spectrum of the desired signal, is depicted too. As expected, the PEM-AFC offers the best performance for a highly time-varying desired signal $x[k]$. In contrast to the 2-ch AFC and the filtered-X algorithm, the PEM-AFC succeeds in tracking the instantaneous AR model, resulting in a good feedback path estimate if $d_f \geq 21$. Even for $15 < d_f < 21$, the performance of the PEM-AFC for a highly time-varying desired signal is quite good. In the PEM-AFC implementation, the forgetting factor $\lambda_{\hat{\mathbf{f}}}$ of the feedback path estimate $\hat{F}(q)$ is much larger than the forgetting factor $\lambda_{\mathbf{a}}$ of $A(q)$. As a result, the filter $A(q)$ is able to model the fast changes in the speech model $H(q)$ for highly time-varying signals $x[k]$, even when $d_f < L_A$. Hence, for highly time-varying signals $x[k]$ and $\lambda_{\mathbf{a}} \ll \lambda_{\hat{\mathbf{f}}}$, the identifiability conditions are less crucial than for stationary signals $x[k]$ or $\lambda_{\mathbf{a}} \approx \lambda_{\hat{\mathbf{f}}}$. The 2-ch AFC and filtered-X algorithm provide a less accurate estimate than the PEM-AFC but outperform the CAF if $d_f \geq 21$.

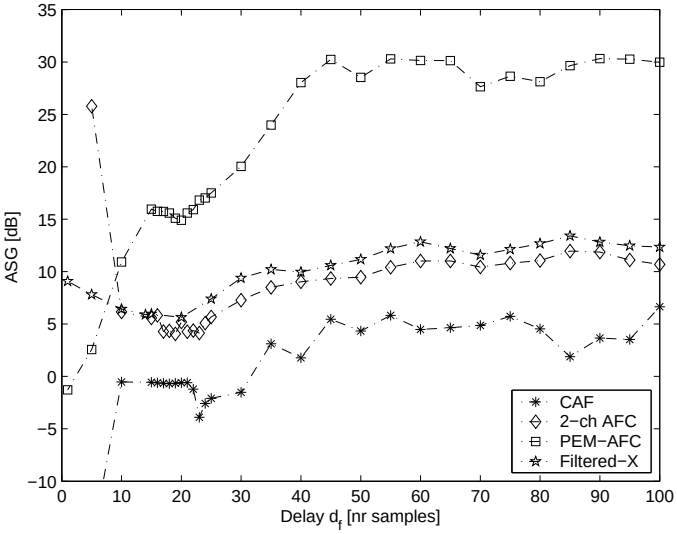
Figure 7.10 depicts the filter $A(q)$ obtained by the 2-ch AFC for $d_f = L_A = 21$ and the inverse of the average PSD of $x[k]$: for sufficiently large d_f, the filter $A(q)$ of the 2-ch AFC is found to converge to the average inverse speech model and hence, the 2-ch AFC approaches the performance of the filtered-X algorithm that uses a model of the long-term average speech spectrum (cf. Figure 7.9).

7.4.4 Real speech signal

In the second example, $x[k]$ is a *real speech signal*, consisting of sentences spoken by a male speaker. Note that the white noise assumption for $w[k]$ is not



(a)



(b)

Figure 7.9: (a) Misalignment $\zeta(F, \hat{F})$ and (b) ASG of CAF, 2-ch AFC, PEM-AFC and filtered-X as a function of the delay d_f for a highly time-varying synthetic speech signal $x[k]$.

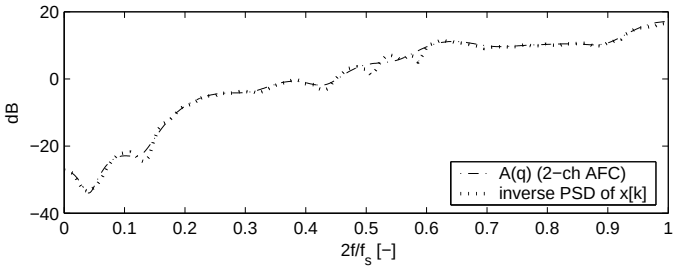


Figure 7.10: Comparison between the amplitude characteristic of the filter $A(q)$ of the 2-ch AFC and the inverse of the average PSD of $x[k]$.

completely satisfied anymore: the excitation $w[k]$ of voiced speech segments corresponds to an impulse train rather than to white noise, so that $w[k]$ and $w[k-d]$ may be correlated. As a consequence, the performance of the adaptive feedback cancellation techniques is expected to be affected: during voiced segments, the feedback canceller will not only model $H(q)$ and $F(q)$, but will also make an attempt to cancel the impulse train (see also Section 7.2.1). In the 2-ch AFC, the filter $A(q)$ is seen to converge to the average speech model. The filter $B(q)$, and hence, the feedback canceller $\hat{F}(q) = -A^{-1}(q)B(q)$ then tries to cancel the impulse train excitation. In the PEM-AFC, the fast filter $A(q)$ tries to track the instantaneous speech spectrum: theoretically the filter $A(q)$ ought to notch out all frequencies caused by the impulse train excitation during voiced segments, so that $e_p[k]$ goes to 0 (for any $\hat{F}(q)$). So, adaptation of $\hat{F}(q)$ is effectively performed only during non-voiced speech segments. However, due to the small filter length[5] of $A(q)$ and the limited tracking performance, $e_p[k] \neq 0$ during voiced speech, causing a bias to the feedback path estimate being introduced during voiced speech.

Figure 7.11 depicts the misalignment $\zeta(F, \hat{F})$ of the CAF, the 2-ch AFC and the PEM-AFC as a function of time (expressed in number of samples at f_s) for $d_f = 30$. The performance of the filtered-X method using the average speech model is depicted too. The CAF has the worst performance among all methods. Initially, the CAF suffers from a huge bias. Due to the time-variability of $x[k]$, the average correlation between $u[k]$ and $x[k]$ decreases for an increasing number of data samples N used in the RLS algorithm. Hence, a slow convergence rate (i.e., a large forgetting factor λ or, in the case of LMS, a small step size ρ) is beneficial to the CAF at the expense of reduced tracking performance (cf. Section 6.3.3). The 2-ch AFC performs better than the CAF.

[5]Note that to cancel the impulse train with period P effectively, the filter $A(q)$ should model $\Pi_{i=1}^{L_h}(1 - e^{j2\pi i \frac{f_0}{f_s}} q^{-1})(1 - e^{-j2\pi i \frac{f_0}{f_s}} q^{-1})$ with $L_h = \left\lfloor \frac{f_s}{2f_0} \right\rfloor$ (see Remark 7.2 in Section 7.2). The pitch frequency f_0 lies between 100 Hz and 300 Hz. For $f_s = 16$ kHz, the number of harmonics L_h lies between 20 and 80, hence, L_A should be between 40 and 160 taps to effectively cancel the impulse train.

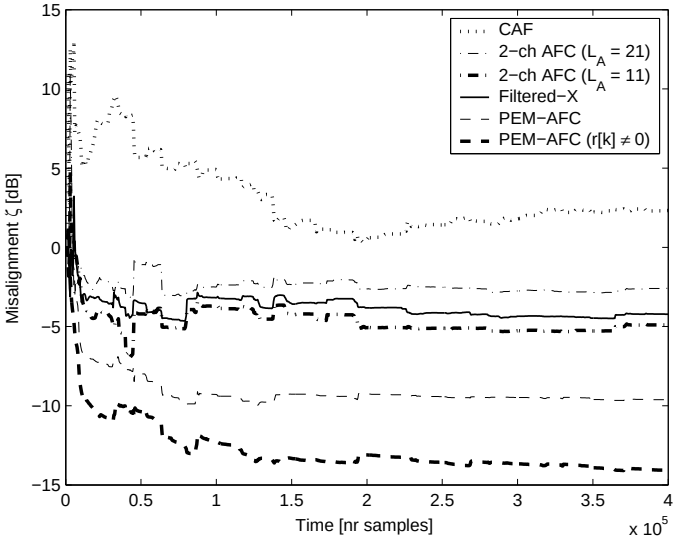


Figure 7.11: Misalignment as a function of time (number samples) for a real speech signal $x[k]$ ($d_f = 30$).

Compared to the filtered-X algorithm, the 2-ch AFC has a larger number of filter taps, i.e., $L_A + L_B$ taps, and hence, more degrees of freedom to attenuate the frequencies caused by the impulse train excitation. For a smaller $L_A + L_B$ (e.g., $L_A = 11$, $L_B = L_A + L_F - 1$), the 2-ch AFC performs better. The PEM-AFC offers the best performance. The performance of the PEM-AFC can be further improved by inserting uncorrelated noise $r[k]$ at the input of the loudspeaker. In Figure 7.11, the performance of the PEM-AFC with $r[k] \neq 0$ is depicted too. The noise $r[k]$ has the long-term average speech spectrum and a level of -20 dB with respect to $G(q)x[k]$. The performance curves of the CAF, the 2-ch AFC and the filtered-X algorithm with $r[k] \neq 0$ almost coincide with the curves for $r[k] = 0$ and are, therefore, omitted. The noise is too low to reduce the bias of the CAF, 2-ch AFC and filtered-X algorithm significantly, but clearly improves the performance of the PEM-AFC.

Figure 7.12 plots the amplitude characteristic of the closed-loop system $S(q) = 1/\left(1 - G(q)(F(q) - \hat{F}(q))\right)$ for the CAF, the 2-ch AFC and the PEM-AFC (with $r[k] = 0$). Ideally, the amplitude characteristic of the closed-loop system $S(q)$ should equal 0 dB at all frequencies. The 2-ch AFC outperforms the CAF, but exhibits fast fluctuations due to cancellation of the impulse train excitation of voiced speech segments. The PEM-AFC has the smallest amount of signal distortion.

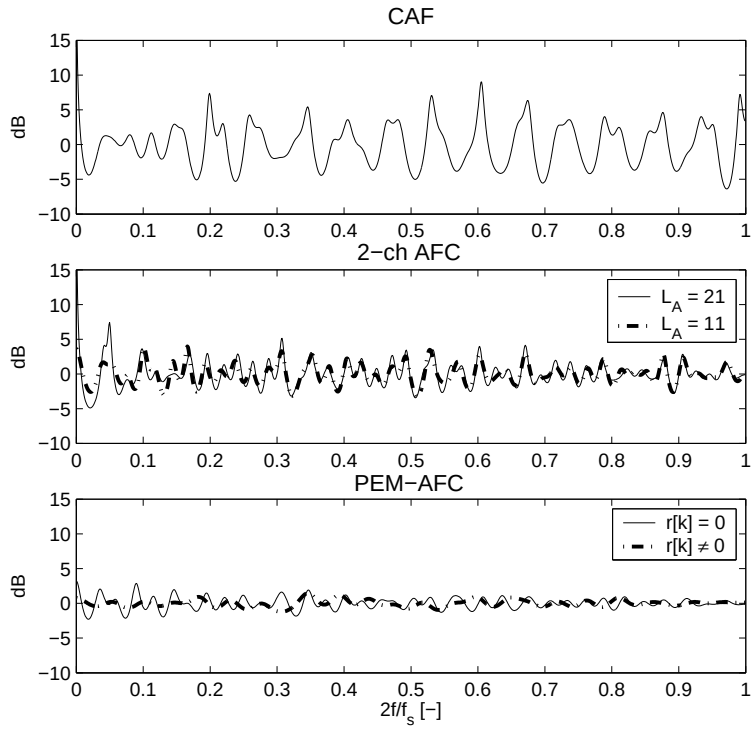


Figure 7.12: Amplitude characteristic of the closed-loop response $S(q) = 1/(1 - G(q)(F(q) - \hat{F}(q)))$ of the CAF, the 2-ch AFC and the PEM-AFC ($d_f = 30$).

Figure 7.13 plots the misalignment $\zeta(F, \hat{F})$ and the ASG of the CAF and PEM-AFC as a function of the delay d_f for $r[k] = 0$ and $r[k] \neq 0$. The noise has a level of -20 dB with respect to $G(q)x[k]$. In general, the performance measures improve for an increasing delay d_f. For all values of the delay $d_f \geq 21$, the PEM-AFC outperforms the CAF: on average, the PEM-AFC has a 5 dB larger ASG than the CAF. Injection of low-level noise further improves the performance of the PEM-AFC, while the gain in performance of the CAF is negligible.

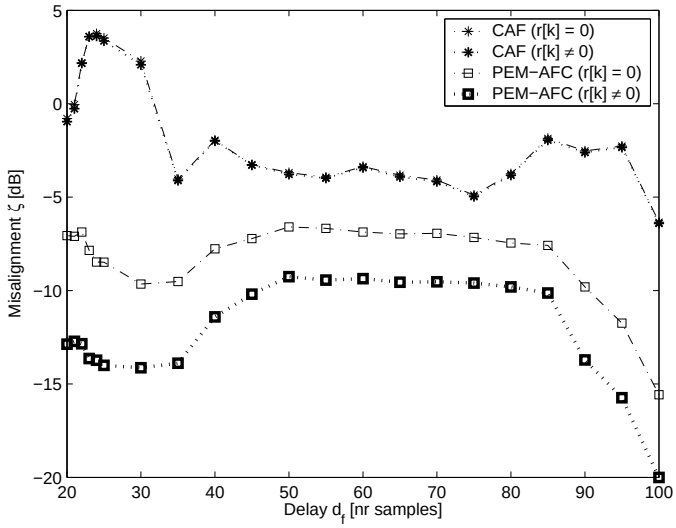
7.5 Conclusions

In this chapter, we have proposed adaptive feedback cancellation techniques that are based on closed-loop identification of the feedback path as well as the (autoregressive) model of the desired signal with the aim of reducing the bias in the feedback path estimate.

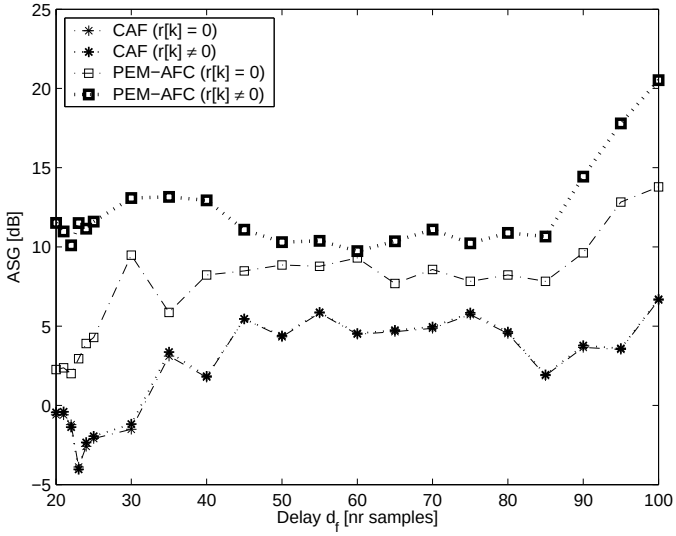
In general, both models are not simultaneously identifiable in the closed-loop system at hand. We demonstrated that not only inserting non-linearities or a probe signal, but also adding a delay in the forward path or in the cancellation path can render the system identifiable. The required delay is much smaller than the delay that is needed in the CAF to decorrelate the received sound signal and the loudspeaker signal.

Two classes of recursive procedures for identifying the desired signal model and the feedback path have been derived: a two-channel adaptive feedback canceller (2-ch AFC) as well as a decoupled PEM-AFC. In contrast to the 2-ch AFC, the decoupled PEM-AFC uses a different adaptation scheme for the feedback path and the desired signal model and hence, is found preferable for highly non-stationary sound signals.

Simulations demonstrated that both recursive techniques significantly outperform the existing CAF when the identifiability conditions are fulfilled. For a time-varying desired signal $x[k]$, the decoupled PEM-AFC offers the best performance and outperforms the filtered-X algorithm of [88, 89] that uses a *fixed* estimate of the average desired signal spectrum. In the derivation of the identifiability conditions, we assumed that the desired signal is generated by passing white noise through an all-pole filter. If $x[k]$ is a speech signal, the excitation of voiced speech segments corresponds to an impulse train, and hence, the performance of the algorithms is expected to be affected. Simulations showed that also for real speech signals the decoupled PEM-AFC has a clear benefit with respect to the standard CAF and the filtered-X algorithm of [89, 88] that a *fixed* estimate of the average desired signal spectrum. In addition, injection of low-level noise further improves the performance of the PEM-AFC, while the benefit of low-level noise injection to the standard CAF and the filtered-X algorithm of [88, 89] is negligible.



(a)



(b)

Figure 7.13: (a) Misalignment $\zeta(F, \hat{F})$ and (b) ASG of CAF and PEM-AFC ($L_A = 21$) as a function of the delay d_f for a real speech signal $x[k]$. The dashed-dotted lines correspond to $r[k] = 0$, the dotted lines correspond to $r[k] \neq 0$.

Chapter 8

Conclusions and further research

THIS chapter summarizes the conclusions of this thesis and provides some suggestions for further research.

8.1 Conclusions

Understanding speech in noise and acoustic feedback belong to the major problems of current hearing aid users. With the advent of low power DSPs, an urgent demand exists for efficient and well-working algorithms that offer a solution to these problems. In this thesis, we have developed several digital signal processing techniques for noise reduction and feedback cancellation. In order to be able to deal with time-varying acoustic environments and non-stationary signals, all the proposed algorithms are adaptive. In the algorithmic design, the specific design criteria for hearing aids, such as small-sized systems, low computational complexity and a low processing delay, have been taken into account.

8.1.1 Noise reduction

In the field of noise reduction, we have concentrated on multi-microphone techniques. Because of the miniaturization in the hearing aid industry, robustness of the noise reduction algorithm against imperfections such as microphone mismatch has become a crucial issue. Indeed, commercial hearing aids often suffer from a performance degradation in real-life use because of a non-robust algorithmic design.

Among the different adaptive multi-microphone noise reduction techniques in the literature, we can distinguish between two classes, namely, adaptive beamforming (of which the Generalized Sidelobe Canceller is an example) and multi-channel Wiener filtering. In particular, we have concentrated on a multi-channel Wiener filtering technique that, in contrast to the GSC and other MWF techniques, requires neither a priori knowledge about the desired signal model (such as the microphone configuration, the speaker position) nor a calibration procedure. This MWF technique, although attractive from a robustness point of view, has not been considered yet for implementation in hearing aids because of its high computational cost.

In Chapter 2, we have shown that an implementation of the MWF becomes feasible for future hearing aids, when it is implemented in subbands. Furthermore, a subband implementation better achieves the major goal of a noise reduction algorithm in a hearing aid, which is to improve speech intelligibility.

In Chapter 3, we have analyzed the robustness of the MWF technique and the GSC against different model errors. In the analysis, we have concentrated on small-sized arrays as used in hearing aids. We have provided a theoretical framework to predict the impact of different types of model errors on the intelligibility weighted performance of the algorithms. The theoretical results have been verified through experiments with a three-microphone BTE. We have shown that because of the high sensitivity of small-sized systems, the *number of useful microphones in a commercial BTE is limited to three*, provided that the noise reduction algorithm is sufficiently robust. Theoretical and experimental analysis has pointed out that *robustness of the GSC* is especially crucial in *complicated noise scenarios* and that *microphone mismatch* is particular harmful to the GSC, even if the ANC is adapted during noise-only. As a consequence, a noise sensitivity constraint (e.g., a quadratic inequality constraint on the ANC) is required. The *MWF* on the other hand, *is not degraded by errors in the signal model* and reduces more noise than a robust GSC with a noise sensitivity constraint (NSC-GSC). Because the MWF does not require a noise sensitivity constraint, the performance benefit of a third microphone in a BTE can be fully exploited.

However, since the MWF does not require a priori information about the desired signal model, it is expected to be more sensitive to *inaccurate estimation of the second order statistics* caused by, e.g., *non-stationarities in the noise* or *speech detection errors*. Experiments demonstrated that the MWF preserves its benefit in robustness and performance over the GSC and the NSC-GSC for spectrally short-time non-stationary noise, provided that the spatial and l*ong-term* spectral characteristics of the noise vary relatively slowly in time (on the order of seconds). In addition, it has been shown both theoretically and experimentally that the noise reduction performance of the MWF is, in contrast to the GSC and the NSC-GSC, not degraded by speech detection errors, but that the speech gets distorted when speech + noise samples are

wrongly detected as noise only. For a reasonable speech detection error rate of 20% or less, the distortion by the MWF is however limited, while compared to the NSC-GSC, more noise is reduced.

In Chapter 4, we have integrated the GSC and the MWF in a generalized noise reduction scheme, called the *spatially pre-processed speech distortion weighted MWF*, that consists of a spatial pre-processor and an adaptive speech distortion weighted MWF. The SP-SDW-MWF encompasses the GSC and the MWF as extreme cases and allows for in-between solutions that can be interpreted as a speech distortion regularized GSC. In contrast to the robust GSC with quadratic inequality constraint (QIC-GSC), the SDR-GSC shifts emphasis towards speech distortion when the amount of speech leakage grows. As a result, a better noise reduction performance is obtained for small model errors. In addition, the extra filtering of the speech reference in the SP-SDW-MWF further improves the performance. We have proven that in the absence of model errors and for infinitely long filter lengths, the SP-SDW-MWF corresponds to an SDR-GSC cascaded with a single-channel speech distortion weighted Wiener post-filter. In contrast to the GSC, the SDR-GSC and the QIC-GSC, its performance does not degrade due to microphone mismatch. Experimental results with a BTE have demonstrated that the SP-SDW-MWF indeed achieves a better noise reduction performance than the QIC-GSC for a given maximum allowable speech distortion level.

However, in contrast to the GSC and the QIC-GSC, no cheap stochastic gradient based implementation of the SP-SDW-MWF was available yet. In Chapter 5, we have developed a low-cost, *stochastic gradient implementation* of the SP-SDW-MWF that is *feasible for implementation in commercial hearing aids*. To speed up convergence and reduce computational complexity, the algorithm is implemented in the frequency-domain. Experimental results have shown that the proposed frequency-domain stochastic gradient algorithm preserves the benefit of the exact SP-SDW-MWF over the QIC-GSC, while its computational cost is comparable to the LMS based scaled projection algorithm for implementing the QIC-GSC. This makes the stochastic gradient SP-SDW-MWF an ideal candidate for the next generation of commercial hearing aids.

8.1.2 Feedback cancellation

Standard continuous adaptation feedback cancellers (CAF) suffer from a serious model error or bias, when the desired signal is spectrally colored. As a result, the desired signal gets cancelled and distorted. Classical solutions for reducing the model error trade off between desired signal cancellation on the one hand and tracking performance or sound quality on the other hand, so that typically a significant residual bias remains. An appealing alternative is to regard the feedback path as a part of a closed-loop system and to apply closed-loop system identification theory. However, solutions proposed so far in this context, either

require perfect knowledge of the (unknown and time-varying) desired signal spectrum or the use of a high-level probe signal.

In Chapter 7, we have proposed adaptive feedback cancellation techniques that are based on closed-loop identification of the feedback path as well as the (autoregressive) model of the desired signal with the aim of providing an unbiased feedback path estimate for unknown and time-varying desired signals. In general, the feedback path and the desired signal model are not simultaneously identifiable in the closed-loop system at hand. We demonstrated that not only inserting non-linearities or a probe signal, but also adding a small delay in the forward path or in the cancellation path can render the system identifiable.

Two classes of adaptive procedures for identifying the desired signal model and the feedback path have been derived, namely, a two-channel identification method (2-ch AFC) as well as a prediction error method (PEM-AFC). In contrast to the 2-ch AFC, the PEM-AFC allows the use of a different adaptation speed for the feedback path and for the desired signal model and hence, is found to be preferable for highly non-stationary desired signals. Simulations demonstrated that both recursive techniques significantly outperform the existing CAF when the identifiability conditions are fulfilled. For time-varying desired signals $x[k]$, the decoupled PEM-AFC offers the best performance and outperforms the filtered-X algorithm of [88, 89] that uses a *fixed* estimate of the average desired signal spectrum. In addition, injection of low-level noise further improves the performance of the PEM-AFC, while the benefit of low-level noise injection for the standard CAF and the filtered-X algorithm of [88, 89] is negligible.

8.2 Suggestions for further research

8.2.1 Noise reduction

Improvements to the SP-SDW-MWF

In Chapter 4, we have proposed a generalized noise reduction scheme, called the SP-SDW-MWF, that outperforms the robust QIC-GSC. In Chapter 5, a low-cost stochastic gradient implementation has been derived, making the SP-SDW-MWF feasible for implementation in commercial hearing aids. Based on the observations in Chapter 3 concerning the robustness of the MWF, we conclude that this SP-SDW-MWF could be further improved in the following ways.

Like the MWF, the SP-SDW-MWF requires an estimate of a clean speech correlation matrix, namely, for computing the regularization term. In Chapter 3, we observed that because of the non-continuous estimation of the noise and the speech + noise statistics, the estimation accuracy of the second order statis-

tics is degraded by non-stationarities in the noise. In addition, the accuracy is strongly dependent of the quality of the speech detection mechanism. In cocktail party noise and persistent speech activity, classical speech detection mechanisms (e.g., energy-based) may however fail to provide a reliable decision. A topic of further research would be to *improve the estimation accuracy of the regularization term*. In the field of single-channel speech enhancement, a minimum statistics procedure has been proposed that, in contrast to classical speech detection driven noise estimators, is also capable of updating the noise PSD during speech activity. The noise PSD is estimated by tracking the minima of the noise PSD [163]. Compared to classical speech detection algorithms, the minimum statistics procedure allows for a quasi-simultaneous, and hence, more accurate estimation of the speech + noise PSD and the noise PSD. To improve the accuracy of regularization term, it would be convenient to extend this minimum statistics approach to the multi-channel case, where also the cross-power spectral densities have to be estimated. A possible extension can be found in [2].

In scenarios with low SNR, producing a reliable estimate of the second order statistics of the speech + noise and the noise is difficult, so that the performance of the adaptive SDW-MWF stage becomes unreliable. In these scenarios, the use of a fixed beamformer output is advisable. Therefore, it would be interesting to *incorporate a control mechanism* in the SP-SDW-MWF that establishes a transition to the output of a fixed beamformer when the SNR drops below a certain threshold. The SP-SDW-MWF structure lends itself to the use of such a mechanism: indeed, by setting the adaptive filters to zero, the output of the fixed beamformer of the spatial pre-processor is obtained. A less stringent procedure is to still allow some, be it restricted, adaptation of the SDW-MWF, e.g., by imposing a quadratic inequality constraint on the adaptive filter or by putting more focus on speech distortion in the SDW-MWF criterion (i.e., by decreasing the trade-off factor μ), when the SNR decreases (cf. Section 4.2).

In Chapter 4 and Chapter 5, we especially concentrated on the adaptive SDW-MWF stage of the SP-SDW-MWF. The performance of the SP-SDW-MWF could be further improved by using a more optimized spatial pre-processor. In this thesis, a DS beamformer has been used as a fixed beamformer. For small-sized multi-microphone systems, this beamformer offers a high robustness against signal model errors. However, the amount of noise suppression is negligible. With the design procedures for robust broadband beamformers [37] (cf. Section 2.2.5), one could design a better fixed beamformer, which reduces more noise than the DS beamformer while still being sufficiently robust. In addition, also the quality of the noise references could be improved with the design procedures of [37]. Given statistical knowledge of the microphone characteristics, the speech leakage in the noise references could be minimized for all possible microphone characteristics. As a result, the average amount of speech leakage is reduced so that more attention could be paid to noise reduc-

tion in the adaptive SDW-MWF stage. Typically, the design procedures for fixed beamformers are applied under the assumption of free field. In order to obtain a spatial pre-processor that is still effective when used for hearing aids, it may be preferable to take into account the head shadow effect in the design (see e.g. [171]).

Validation by perceptual tests

In this thesis, the performance of the developed algorithms has been evaluated based on objective intelligibility weighted performance measures. To get an idea of the actual benefit of the algorithms for hearing aid users in everyday listening situations, the major objective results should be *validated by perceptual tests* with normal hearing and hearing impaired people. In particular, the benefit of the SP-SDW-MWF over the QIC-GSC (i.e., a better intelligibility weighted performance and a low sensitivity to model errors), the benefit of a subband implementation for speech intelligibility and the contribution of an additional third microphone in a BTE require perceptual validation. Furthermore, real-life testing allows to *detect potential additional deficiencies* of the developed algorithms. So may the *own voice* of the hearing aid user disturb the adaptation of the noise reduction algorithms.

Perceptually motivated noise reduction

In Chapter 2, we have shown that a subband MWF implementation better optimizes intelligibility than a fullband realization. However, the MMSE criterion on which the MWF is based is not so relevant from a psychoacoustic point of view. In the field of single-channel noise reduction, alternative spectral weighting rules have been proposed that are based on the concept of masking threshold [4, 79, 80, 104, 110, 256, 265]. Basically, these techniques only attenuate the noise when its level exceeds the masking threshold, resulting in less speech distortion and musical noise for single-channel procedures. In further research, one could investigate whether the concept of masking or some kind of hearing model could also be successfully incorporated in the optimization criterion of multi-channel noise reduction algorithms.

Binaural noise reduction

In this thesis, we have concentrated on monaural hearing aid systems that are located at one single ear. With the future advent of a wireless link between left and right ear, binaural noise reduction systems become commercially feasible. Hence, it would be interesting to evaluate the potential of the here developed algorithms for binaural systems [32, 81, 170, 186, 271, 274]. In addition to spatial noise reduction, binaural signal processing techniques try to provide the natural benefits of sound localization and speech intelligibility that origi-

nate from binaural listening. The MWF with unknown reference estimates the speech as it is received at the microphone, which seems advantageous for binaural hearing. However, there is typically a trade-off between noise reduction performance and the preservation of localization cues (such as inter-aural time delay and inter-aural level differences), so that additional constraints may be required (cf. [32, 186, 271]).

8.2.2 Feedback cancellation

In Chapter 7, adaptive feedback cancellation techniques have been developed that are based on direct closed-loop identification as well as the autoregressive modelling of the desired signal model. We have shown that under certain identifiability conditions, an unbiased feedback path can be obtained. Among the different conditions, we have especially concentrated on the use of a delay in the forward path, because this technique does not affect the sound quality. Inserting an external probe signal or adding non-linearities to the forward path can however be useful alternatives. In Chapter 7, we have shown that the feedback path can not be estimated if the desired signal is periodic, e.g., a sinusoid or a voiced speech segment. Initial experiments in Chapter 7 have indicated that the performance for a real speech signal is improved by inserting a low-level probe signal. In addition, the use of a probe signal allows to apply other closed-loop identification procedures such as the joint input-output method. In further research, the *potential benefits of noise injection and non-linearities* should therefore be investigated in further detail. An important issue when using these techniques is to preserve the sound quality. Additive noise is, e.g., not audible if its level is below the hearing threshold of the hearing aid user or if it is masked by the desired signal. Hence, it would be interesting to translate the concepts of loudness and masking threshold into low-cost algorithms for the generation of inaudible noise signals or non-linearities.

In Chapter 7, we have proposed two classes of adaptive procedures for identifying the feedback path and the desired signal model, namely, a two-channel identification based (2-ch AFC) and a prediction error method (PEM-AFC). The PEM-AFC allows the use of a different adaptation scheme for the feedback path and for the desired signal model and hence, is preferable for highly time-varying signals such as speech. However, in contrast to the 2-ch AFC, the optimization criterion is non-linear, so that local minima may occur. Further research could focus on *the design of alternative identification procedures.* In Chapter 7, e.g., the speech model is estimated by means of a lattice based linear prediction method. The identification of a highly time-varying signal such as speech could, e.g., be improved by the use of time-varying autoregressive (TVAR) modelling [262]. As already indicated in Chapter 6 and Chapter 7, the feedback path could be identified by means of an IV-method instead of an PEM-method. In Remark 7.7, it has been noted that the cost function of the PEM-AFC corresponds to a separable non-linear least-squares problem. Hence,

alternative decoupled implementations of the PEM-AFC may be derived from the theory of separable non-linear least-squares problems [68, 181].

8.2.3 Combined solution for acoustic feedback and noise reduction

Finally, it would be interesting to combine the developed feedback cancellation and noise reduction algorithms into one signal processing scheme (cf. the combination of noise reduction and echo cancellation). A straightforward approach is to perform the noise reduction and the feedback cancellation algorithm independently, either by first cancelling the acoustic feedback in all microphone signals and then performing the noise reduction, or vice-versa, by first performing multi-microphone noise reduction, followed by a single-channel feedback canceller. Although the latter seems attractive in terms of computational complexity, we believe that the former approach is preferable in terms of performance. So may the time-varying character of the noise reduction algorithm and hence, the forward path improve the identifiability of the feedback path estimate and the desired signal model (cf. Chapter 7). In addition, it would be interesting to investigate whether the use of information in additional microphone signals does not weaken the conditions for identifiability. Finally, a more elegant approach would be to integrate the noise reduction and the feedback cancellation into one multi-microphone noise reduction algorithm, where acoustic feedback and noise are jointly being reduced. This problem shows similarities with the combination of noise reduction and acoustic echo cancellation, for which an integrated solution has been proposed in [33, 210, 211].

Bibliography

[1] Acoustical Society of America. ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index, June 1997.

[2] R. Aichner, W. Herbordt, H. Buchner, and W. Kellermann. Least-squares error beamforming using minimum statistics and multichannel frequency-domain adaptive filtering. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 223–226, Kyoto, Japan, September 2003.

[3] S. Allegro, S. Büchler, and S. Launer. Automatic sound classification inspired by auditory scene analysis. In *Eurospeech*, Aalborg, Denmark, September 2001.

[4] K. H. Arehart, J. H. L. Hansen, and S. Gallant. Evaluation of an auditory masked threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners. *Speech Communication*, 40(4):575–592, June 2003.

[5] Auditec. Auditory Tests (Revised), Compact Disc, Auditec, St. Louis.

[6] T. Baer and B. C. J. Moore. Effects of spectral smearing on the intelligibility of sentences in noise. *Journal of the Acoustical Society of America*, 94(3):1229–1241, September 1993.

[7] T. Baer and B. C. J. Moore. Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times. *Journal of Rehabilitation Research and Development*, 30(1):49–72, 1993.

[8] J. Benesty and D. R. Morgan. Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, 2000.

[9] J. Bitzer and K. U. Simmer. *Superdirective Microphone Arrays*, chapter 2 in "Microphone Arrays: Signal Processing Techniques and Applications"

(Brandstein, M. S. and Ward, D. B., Eds.), pages 19–38. Springer-Verlag, May 2001.

[10] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer. An alternative implementation of the superdirective beamformer. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 7–10, New Paltz NY, USA, October 1999.

[11] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer. Multi-microphone noise reduction by post-filter and superdirective beamformer. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 100–103, Pocono Manor, USA, September 1999.

[12] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer. Multichannel noise reduction - Algorithms and theoretical limits-. In *Proc. European Signal Processing Conf. (EUSIPCO)*, volume 1, pages 105–108, Rhodes, Greece, September 1999.

[13] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer. Theoretical noise reduction limits of the Generalized Sidelobe Canceller (GSC) for speech enhancement. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 2965–2968, Phoenix, Arizona, USA, May 1999.

[14] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer. Multi-microphone noise reduction techniques as front-end devices for speech recognition. *Speech Communication*, 34:3–12, April 2001.

[15] E. Bjarnasson. Analysis of the filtered-X LMS algorithm. *IEEE Trans. Speech, Audio Processing*, 3(6):504–514, November 1995.

[16] S. F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech, and Signal Processing*, 27(2):113–120, April 1979.

[17] M. Bouchard and S. Quednau. Multi-channel recursive-least-squares algorithms and fast-transversal-filter algorithms for active noise control and sound reproduction systems. *IEEE Trans. Speech, Audio Processing*, 8(5):606–618, September 2000.

[18] H. Buchner, J. Benesty, and W. Kellermann. *Multichannel frequency-domain adaptive filtering with application to multichannel acoustic echo cancellation*, chapter 4 in "Adaptive Signal Processing: Applications to Real-World Problems"(Benesty, J. and Huang, Y., Eds.), pages 95–128. Springer-Verlag, 2003.

[19] M. Buck. Aspects of first-order differential microphone arrays in the presence of sensor imperfections. *European Transactions on Telecommunications*, 13(2c):115–122, Mar.-Apr. 2002.

[20] K. M. Buckley. Broad-band beamforming and the Generalized Sidelobe Canceller. *IEEE Trans. Acoust., Speech, and Signal Processing*, 34(5):1322–1323, October 1986.

[21] D. K. Bustamante, T. L. Worrall, and M. J. Williamson. Measurement and adaptive suppression of acoustic feedback in hearing aids. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 2017–2020, Glasgow, Scotland, May 1989.

[22] O. Cappé. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Trans. Speech, Audio Processing*, 2(2):345–349, April 1994.

[23] J. Cezanne and E. W. Gary. Adaptive microphone array. U.S. Patent US 5303307, December 1995.

[24] H.-F. Chi, S. X. Goa, S. D. Soli, and A. Alwan. Band-limited feedback cancellation with a modified filtered-X LMS algorithm for hearing aids. *Speech Communication*, 39(1-2):147–161, January 2003.

[25] I. Claesson and S. Nordholm. A spatial filtering approach to robust adaptive beamforming. *IEEE Trans. Antennas Propag.*, 40(9):1093–1096, September 1992.

[26] G. A. Clark, S. K. Mitra, and S. R. Parker. Block implementation of adaptive digital filters. *IEEE Trans. Acoust., Speech, and Signal Processing*, 29(3):744–752, 1981.

[27] H. Cox, R. M. Zeskind, and T. Kooij. Resolving power and sensitivity to mismatch of optimum array processors. *Journal of the Acoustical Society of America*, 54(3):771–785, 1974.

[28] H. Cox, R. M. Zeskind, and T. Kooij. Practical Supergain. *IEEE Trans. Acoust., Speech, and Signal Processing*, 34(3):393–398, June 1986.

[29] H. Cox, R. M. Zeskind, and M. M. Owen. Robust Adaptive Beamforming. *IEEE Trans. Acoust., Speech, and Signal Processing*, 35(10):1365–1376, 1987.

[30] R. de Vries and B. de Vries. Towards SNR-loss restoration in digital hearing aids. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume IV, pages 4003–4007, Orlando, Florida, USA, May 2002.

[31] J. R. Deller, J. G. Proakis, and J. H. L. Hansen. *Discrete-Time Processing of Speech Signals*. Macmillan Publishing Company, Englewood Cliffs, New Jersey, 1993.

[32] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek. Microphone-array hearing aids with binaural output–Part I: Fixed-processing systems. *IEEE Trans. Speech, Audio Processing*, 5(6):529–542, November 1997.

[33] S. Doclo. *Multi-microphone noise reduction and dereverberation techniques for speech applications*. PhD thesis, Faculty of Engineering, K.U. Leuven, Belgium, May 2003. available as Technical Report ESAT-SISTA/TR 03-77.

[34] S. Doclo, I. Dologlou, and M. Moonen. A novel iterative signal enhancement algorithm for noise reduction in speech. In *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, pages 1435–1438, Sydney, Australia, December 1998.

[35] S. Doclo and M. Moonen. *GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement*, chapter 6 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 111–132. Springer-Verlag, May 2001.

[36] S. Doclo and M. Moonen. GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement. *IEEE Trans. Signal Processing*, 50(9):2230–2244, September 2002.

[37] S. Doclo and M. Moonen. Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. *IEEE Trans. Signal Processing*, 51(10):2511–2526, October 2003.

[38] S. Doclo and M. Moonen. Design of broadband beamformers robust against microphone position errors. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 267–270, Kyoto, Japan, September 2003.

[39] S. Doclo and M. Moonen. Design of far-field and near-field broadband beamformers using eigenfilters. *Signal Processing*, 83(12):2641–2673, December 2003.

[40] S. Doclo and M. Moonen. Multi-microphone noise reduction using recursive GSVD-based optimal filtering with ANC postprocessing stage. Accepted for publication in *IEEE Trans. Speech, Audio Processing*, 2004.

[41] S. Doclo, A. Spriet, and M. Moonen. Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction. In *Proc. European Signal Processing Conf. (EUSIPCO)*, Vienna, Austria, September 2004.

[42] S. Doclo, A. Spriet, and M. Moonen. Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering

for noise reduction. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS)*, pages 195–198, Hilvarenbeek, The Netherlands, April 2004.

[43] S. Doclo, A. Spriet, M. Moonen, and J. Wouters. Design of a robust multi-microphone noise reduction algorithm for hearing instruments. In *Proc. of the International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, July 2004.

[44] A. J. Duquesnoy. *Speech intelligibility of the hearing impaired*. PhD thesis, Free University Amsterdam, The Netherlands, 1982.

[45] A. J. Duquesnoy and R. Plomp. The effect of a hearing aid on the speech-reception threshold of hearing-impaired listeners in quiet and noise. *Journal of the Acoustical Society of America*, 73(6):2166–2173, June 1983.

[46] O. Dyrlund and N. Bisgaard. Acoustic feedback margin improvements in hearing instruments using a prototype DFS (digital feedback suppression) system. *Scandinavian Audiology*, 20(1):49–53, 1991.

[47] O. Dyrlund, L. B. Henningsen, and J. H. Jensen. Digital feedback suppression (DFS). Characterization of feedback-margin improvements in a DFS hearing instrument. *Scandinavian Audiology*, 23(2):135–138, 1994.

[48] G. W. Elko. Microphone array systems for hands-free telecommunication. *Speech Communication*, 20(3-4):229–2240, 1996.

[49] G. W. 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, 2000.

[50] K. Eneman and M. Moonen. DFT modulated filter bank design for over-sampled suband systems. *Signal Processing*, 81(9):1947–1973, September 2001.

[51] A. M. Engebretson and M. F. George. Properties of an adaptive feedback equalization algorithm. *Journal of Rehabilitation Research and Development*, 30(1):8–16, 1993.

[52] Y. Ephraim and D. Malah. Speech enhancement using a minimun mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech, and Signal Processing*, 32(6):1109–1121, December 1984.

[53] Y. Ephraim and D. Malah. Speech enhancement using a minimun mean-square error log-spectral amplitude estimator. *IEEE Trans. Acoust., Speech, and Signal Processing*, 33(2):443–445, April 1985.

[54] Y. Ephraim and H. L. Van Trees. A signal subspace approach for speech enhancement. *IEEE Trans. Speech, Audio Processing*, 3(4):251–266, July 1995.

[55] J. M. Festen and R. Plomp. Relations between auditory functions in hearing impaired. *Journal of the Acoustical Society of America*, 73(2):652–662, February 1983.

[56] S. Fischer and K. U. Simmer. Beamforming microphone arrays for speech acquisition in noisy environments. *Speech Communication*, 20(3-4):215–227, December 1996.

[57] D. A. Florêncio and H. S. Malvar. Multichannel filtering for optimum noise reduction in microphone arrays. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, Utah, May 2001.

[58] M. C. Flynn. Opening ear fittings: nine questions and answers. *The Hearing Review*, 11(3), March 2004.

[59] U. Forssell. *Closed-loop identification - methods, theory and applications*. PhD thesis, Linköping Universitet, Linköping, Sweden, 1999.

[60] U. Forssell and L. Ljung. Closed-loop identification revisited. *Automatica*, 35:1215–1241, 1999.

[61] B. A. M. Franck, C. Sidonne, G. M. van Kreveld-Bos, W. A. Dreschler, and H. Verschuure. Evaluation of spectral enhancement in hearing aids, combined with phonemic compression. *Journal of the Acoustical Society of America*, 106(3):1452–1464, September 1999.

[62] O. L. Frost III. An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60:926–935, August 1972.

[63] M. Gabrea. A single microphone Kalman filter-based noise canceller. *IEEE Signal Processing Letters*, 6(3):55–57, March 1999.

[64] S. Gannot, D. Burshtein, and E. Weinstein. Iterative and sequential Kalman filter-based speech enhancement algorithms. *IEEE Trans. Speech, Audio Processing*, 6(4):373–385, July 1998.

[65] S. Gannot, D. Burshtein, and E. Weinstein. Signal enhancement using beamforming and non-stationarity with applications to speech. *IEEE Trans. Signal Processing*, 49(8):1614–1626, August 2001.

[66] S. Gannot, Burshtein D., and Weinstein E. Theoretical analysis of the general transfer function GSC. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 103–106, Darmstadt, Germany, September 2001.

[67] S. Gao, S. Soli, and H.-F. Chi. Band-limited adaptive feedback canceller for hearing aids. European patent, EP 1118247, 1999.

[68] G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Institute of Physics, Inverse Problems*, 19(2):R1–R26, April 2003. available at http://www.iop.org/EJ/article/0266-5611/19/2/201/ip32r1.pdf.

[69] G. H. Golub and C. F. Van Loan. *Matrix Computations*. MD : John Hopkins University Press, Baltimore, 2nd edition, 1989.

[70] N. Grbíc and S. Nordholm. Soft contrained subband beamforming for hands-free speech enhancement. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 885–888, May 2002.

[71] J. E. Greenberg. Modified LMS algorithms for speech processing with an adaptive noise canceller. *IEEE Trans. Speech, Audio Processing*, 6(4):338–350, July 1998.

[72] J. E. Greenberg, J. G. Desloge, and P. M. Zurek. Evaluation of array-processing algorithms for a headband hearing aid. *Journal of the Acoustical Society of America*, 113(3):1646–1657, March 2003.

[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] J. E. Greenberg and P. M. Zurek. Evaluation of an adaptive beamforming method for hearing aids. *Journal of the Acoustical Society of America*, 91(3):1662–1676, March 1992.

[75] J. E. Greenberg and P. M. Zurek. *Microphone-array hearing aids*, chapter 11 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 229–255. Springer-Verlag, May 2001.

[76] J. E. Greenberg, P. M. Zurek, and M. Brantley. Evaluation of feedback-reduction algorithms for hearing aids. *Journal of the Acoustical Society of America*, 108(5):2366–2376, November 2000.

[77] L. J. Griffiths and C. W. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propag.*, 30(1):27–34, January 1982.

[78] T. Gülzow, A. Engelsberg, and U. Heute. Comparison of a discrete wavelet transformation and a nonuniform polyphase filterbank applied to spectral-subtraction speech enhancement. *Signal Processing*, 64(1):5–19, January 1998.

[79] S. Gustafsson, P. Jax, and P. Vary. A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 397–400, Seattle WA, USA, May 1998.

[80] S. Gustafsson, R. Martin, P. Jax, and P. Vary. A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. *IEEE Trans. Speech, Audio Processing*, 10(5):245–256, July 2002.

[81] V. Hamacher. Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4008–4011, Orlando, Florida, USA, May 2002.

[82] J. H. L. Hansen and M. A. Clements. Constrained iterative speech enhancement with application to speech recognition. *IEEE Trans. Signal Processing*, 39(4):795–805, April 1991.

[83] P. S. K. Hansen. *Signal subspace methods for speech enhancement*. PhD thesis, Technical University of Denmark, Lyngby, Denmark, September 1997.

[84] R. Harjani and R. Wang. Apparatus for eliminating acoustic oscillation in a hearing aid by using phase equalization. U.S. patent, US 5621802, 1997.

[85] S. Haykin. *Adaptive filter theory*. Information and system sciences series. Prentice Hall, Englewood Cliffs, New Jersey, 4th edition edition, 2001.

[86] K. S. Helfer and L. A. Wilber. Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech and Hearing Research*, 33(1):149–155, 1990.

[87] J. Hellgren. *Compensation for hearing loss and cancellation of acoustic feedback in digital hearing aids*. PhD thesis, Linköpings Universitet, Linköping, Sweden, April 2000.

[88] J. Hellgren. Analysis of feedback cancellation in hearing aids with filtered-X LMS and the direct method of closed loop identification. *IEEE Trans. Speech, Audio Processing*, 10(2):119–131, February 2002.

[89] J. Hellgren and U. Forssell. Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification. *IEEE Trans. Speech, Audio Processing*, 9(7):906–913, November 2001.

[90] J. Hellgren, T. Lunner, and S. Arlinger. System identification of feedback in hearing aids. *Journal of the Acoustical Society of America*, 105(6):3481–3496, June 1999.

[91] 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.

[92] W. Herbordt, H. Buchner, and W. Kellermann. An acoustic human-machine front-end for multimedia applications. *EURASIP journal on Applied Signal Processing*, 1:21–31, January 2003.

[93] W. Herbordt and W. Kellermann. Frequency-domain integration of acoustic echo cancellation and a Generalized Sidelobe Canceller with improved robustness. *European Transactions on Telecommunications*, 13(2):123–132, Mar.-Apr. 2002.

[94] W. Herbordt and W. Kellermann. *Adaptive Beamforming for Audio Signal Acquisition*, chapter 6 in "Adaptive Signal Processing: Applications to Real-World Problems"(Benesty, J. and Huang, Y., Eds.), pages 155–188. Springer-Verlag, 2003.

[95] I. Hochberg, A. Boothroyd, M. Weiss, and S. Hellman. Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear and Hearing*, 13(4):263–271, August 1992.

[96] M. W. Hoffman and K. M. Buckley. Robust time-domain processing of broadband acoustic data. *IEEE Trans. Speech, Audio Processing*, 3:193–203, May 1995.

[97] M. W. Hoffman, X. F. Lu Pinkelman, and Z. Li. Real-time and off-line comparisons of standard array configurations containing three and four microphones. *Journal of the Acoustical Society of America*, 2000.

[98] M. W. Hoffman, T. D. Trine, K. M. Buckley, and D. J. Van Tasell. Robust adaptive microphone array processing for hearing aids: realistic speech enhancement predictions. *Journal of the Acoustical Society of America*, 96:759–770, 1994.

[99] V. Hohmann, V. Hamacher, I. Holube, B. Kollmeier, and T. Wittkop. Method for the operation of a hearing aid device or hearing device system as well as hearing aid device or hearing aid device system. U.S. patent, US 2002/0176595 A1, 2002.

[100] V. Hohmann, Nix J., G. Grimm, and T. Wittkop. Binaural noise reduction for hearing aids. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4000–4003, Orlando, Florida, USA, may 2002.

[101] V. Hohmann and B. Kollmeier. Recruitment compensation based on simplified loudness estimation model. *Journal of the Acoustical Society of America*, 95(5):2991–2992, May 1994.

[102] V. Hohmann and B. Kollmeier. The effect of multichannel dynamic compression on speech intelligibility. *Journal of the Acoustical Society of America*, 97(2):1191–1195, February 1995.

[103] O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Trans. Signal Processing*, 47:2677–2683, 1999.

[104] Y. Hu and P. C. Loizou. A perceptually motivated approach for speech enhancement. *IEEE Trans. Speech, Audio Processing*, 11(5):457–465, September 2003.

[105] J. Huang and Y. Zhao. Energy-constrained signal subspace method for speech enhancement and recognition. *IEEE Signal Processing Letters*, 4(10):283–285, October 1997.

[106] International Collegium of Rehabilitative Audiology. Noise Signals ICRA, Compact Disc, Ver 0.3 1997.

[107] Hear it AISBL. www.hear-it.org.

[108] N. K. Jablon. Adaptive beamforming with the Generalized Sidelobe Canceller in the presence of array imperfections. *IEEE Trans. Antennas Propag.*, 34:996–1012, August 1986.

[109] F. Jabloun and B. Champagne. A multi-microphone signal subspace approach for speech enhancement. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 205–208, Salt Lake City UT, USA, May 2001.

[110] F. Jabloun and B. Champagne. A perceptual signal subspace approach for speech enhancement in colored noise. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 569–572, Orlando, Florida, USA, May 2002.

[111] M. J. Jacobson. Space-time correlation in spherical and circular noise fields. *Journal of the Acoustical Society of America*, 34(7):971–978, July 1962.

[112] C. P. Janse and P. A. A. Timmermans. Signal amplifier system with improved echo cancellation. U.S. patent, US 5748751, 1998.

[113] L. B. Jensen. A method for controlling the directionality of the sound receiving characteristic of a hearing aid and a hearing aid for carrying out the method. WO patent, WO 01/001731, 2001.

[114] L. B. Jensen. Hearing aid with adaptive matching of input transducers. U.S. patent, 6741714, May 2004.

[115] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen. Reduction of broad-band noise in speech by truncated QSVD. *IEEE Trans. Speech, Audio Processing*, 3(6):439–448, November 1995.

[116] C. W. Jim. A comparison of two LMS constrained optimal array structures. *Proceedings of IEEE*, 65(12):1730–1731, December 1977.

[117] D.H. Johnson and D.E. Dudgeon. *Array Signal Processing*. Prentice Hall, Englewood Cliffs, 1993.

[118] H. A. L. Joson, F. Asano, Y. Suzuki, and S. Toshio. Adaptive feedback cancellation with frequency compression for hearing aids. *Journal of the Acoustical Society of America*, 94(6):3248–3254, December 1993.

[119] A. Kaelin, A. Lindgren, and S. Wyrsch. A digital frequency-domain implementation of a very high gain hearing aid with compensation for recruitment of loudness and acoustic echo cancellation. *Signal Processing*, 64(1):71–85, 1998.

[120] M. Kajala and M. Hämäläinen. Broadband beamforming optimization for speech enhancement in noisy environements. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 19–22, New Paltz, NY, USA, October 1999.

[121] S. D. Kampe and M. K. Wynne. The influence of venting on the occlusion effect. *The Hearing Journal*, 49(4):59–66, 1996.

[122] J. M. Kates. Feedback cancellation in hearing aids: Results from a computer simulation. *Signal Processing*, 39(3):553–562, March 1991.

[123] J. M. Kates. Superdirective arrays for hearing aids. *Journal of the Acoustical Society of America*, 94(4):1930–1933, October 1993.

[124] J. M. Kates. Classification of background noises for hearing-aid applications. *Journal of the Acoustical Society of America*, 97(1):461–470, January 1995.

[125] J. M. Kates. Constrained adaptation for feedback cancellation in hearing aids. *Journal of the Acoustical Society of America*, 106(2):1010–1019, August 1999.

[126] J. M. Kates. Feedback cancellation apparatus and methods. U.S. patent, US 6072884, 2000.

[127] J. M. Kates. Feedback cancellation improvements. U.S. patent, US 6219427, 2001.

[128] J. M. Kates. Room reverberation effects in hearing aid feedback cancellation. *Journal of the Acoustical Society of America*, 109(1):367–378, January 2001.

[129] J. M. Kates. *Adaptive feedback cancellation in hearing aids*, chapter 2 in "Adaptive signal processing: Applications to real-world problems"(Benesty, J. and Huang, Y.), pages 23–55. Springer-Verlag, 2003.

[130] J. M. Kates and J. L. Melanson. Feedback cancellation apparatus and methods utilizing adaptive reference filter mechanisms. U.S. patent, US 6434247, 2002.

[131] J. M. Kates and M. R. Weiss. A comparison of hearing-aid array-processing techniques. *Journal of the Acoustical Society of America*, 99(5):3138–3148, May 1996.

[132] T. Kaulberg. A hearing aid with an adaptive filter for suppression of acoustic feedback. European patent, EP 1191814, 2002.

[133] G. Keidser. The relationship between listening conditions and alternative amplification schemes for multiple memory hearing aids. *Ear and Hearing*, 16(6):575–586, 1995.

[134] M. C. Killion. Guest Editorial: Hearing aids: Past, present, future: Moving toward normal conversations in noise. *British Journal of Audiology*, 31:141–148, 1997.

[135] 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.

[136] S. B. Knebel and R. A. Bentler. Comparison of two digital hearing aids. *Ear and Hearing*, 19(4):280–289, August 1998.

[137] S. Köbler and A. Leijon. Noise analysis of real-life listening situations for maximal speech audiblility in hearing aid fitting. *Scandinavian Audiology*, 28(3):179–189, 1999.

[138] M. Kompis. Method for producing an electric signal or method for boosting acoustic signals from a preferred direction, transmitter and associated device. WO patent, WO 00/41436, 2000.

[139] M. Kompis and N. Dillier. Noise reduction for hearing aids: Combining directional microphones with an adaptive beamformer. *Journal of the Acoustical Society of America*, 96(3):1910–1913, September 1994.

[140] M. Kompis and N. Dillier. Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise ratio improvement. *Journal of the Acoustical Society of America*, 109(3):1123–1133, March 2001.

[141] M. Kompis and N. Dillier. Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. II. Experimental verification of the predictions. *Journal of the Acoustical Society of America*, 109(3):1134–1143, March 2001.

[142] D. Korompis, K. Yao, and F. Lorenzelli. Broadband maximum energy array with user imposed spatial and frequency constraints. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 529–532, Adelaide, Australia, April 1994.

[143] F. Kuk, C. Ludvigsen, and T. Kaulberg. Understanding feedback and digital feedback cancellation. *The Hearing Review*, 9(2), February 2002.

[144] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine. Splitting the unit delay. *IEEE Signal Processing Magazine*, 13(1):30–60, January 1996.

[145] S. Laugesen, K. V. Hansen, and J. Hellgren. Acceptable delays in hearing aids and implications for feedback cancellation. *Journal of the Acoustical Society of America*, 105(2):1211–1212, 1999.

[146] R. Le Bouquin-Jeannès, A. A. Azirani, and G. Faucon. Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator. 5(5):484–487, September 1997.

[147] R. Leber and W. Schaub. Circuit and method for the adaptive suppression of an acoustic feedback. U.S. patent, US 6611600, 2003.

[148] H. Lev-Ari and Y. Ephraim. Extension of the signal subspace speech enhancement approach to colored noise. *IEEE Signal Processing Letters*, 10(4):104–106, April 2003.

[149] H. Levitt. Noise reduction in hearing aids: An overview. *Journal of Rehabilitation Research and Development*, 38(1):111–123, Jan/Feb 2001.

[150] M. J. Link and K. M. Buckley. Prewhitening for intelligibility gain in hearing aid arrays. *Journal of the Acoustical Society of America*, 93(4):2139–2140, April 1993.

[151] R. P. Lippmann, L. D. Braida, and N. I. Durlach. Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss. *Journal of the Acoustical Society of America*, 69(2):524–534, February 1981.

[152] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, MA, 1983.

[153] P. C. Loizou. Mimicking the human ear. *IEEE Signal Processing Magazine*, 15(5):101–130, September 1998.

[154] G. Long, F. Ling, and J. G. Proakis. The LMS algorithm with delayed coefficient adaptation. *IEEE Trans. Speech, Audio Processing*, 37(9):1397–1405, September 1989.

[155] G. Long, F. Ling, and J. G. Proakis. Corrections to 'The LMS algorithm with delayed coefficient adaptation'. *IEEE Trans. Signal Processing*, 40(9):230–232, September 1992.

[156] T. Lotter and P. Vary. Noise reduction by maximum a posteriori spectral amplitude estimation with supergaussian speech modeling. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 83–86, Kyoto, Japan, September 2003.

[157] F. Luo, J. Yang, C. Pavlovic, and A. Nehorai. Adaptive null-forming scheme in digital hearing aids. *IEEE Trans. Signal Processing*, 50(7):1583–1590, July 2002.

[158] H. Luts, J.-B. Maj, W. Soede, and J. Wouters. Better speech perception in noise with an assistive multi-microphone array for hearing aids. Accepted for publication in *Ear and Hearing*, 2004.

[159] R. J. Mailloux. *Phased Array Antenna Handbook*. Artech House, Boston, 1994.

[160] J.-B. Maj, J. Wouters, and M. Moonen. Noise reduction results of an adaptive filtering technique for behind-the-ear hearing aids with two microphones. *Ear and Hearing*, 25(3):215–229, June 2004.

[161] J. Makhoul. Stable and efficient lattice methods for linear prediction. *IEEE Trans. Acoust., Speech, and Signal Processing*, 25(5):423–428, October 1977.

[162] C. Marro, Y. Mahieux, and K. U. Simmer. Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering. *IEEE Trans. Speech, Audio Processing*, 6(3):240–259, May 1998.

[163] R. Martin. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Trans. Speech, Audio Processing*, 9(5):504–512, July 2001.

[164] R. Martin. Statistical methods for the enhancement of noisy speech. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 1–6, Kyoto, Japan, September 2003.

[165] R. Martin and C. Breithaupt. Speech enhancement in the DFT domain using Laplacian speech priors. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 83–86, Kyoto, Japan, September 2003.

[166] J. A. Maxwell and P. M. Zurek. Reducing acoustic feedback in hearing aids. *IEEE Trans. Speech, Audio Processing*, 3(4):304–313, July 1995.

[167] M. Mboup, M. Bonnet, and N. Bershad. LMS coupled adaptive prediction and system identification: a statistical model and transient mean analysis. *IEEE Trans. Signal Processing*, 42(10):2607–2615, October 1994.

[168] R. J. McAulay and M. L. Malpass. Speech enhancement using a soft-decision noise suppression filter. *IEEE Trans. Acoust., Speech, and Signal Processing*, 28(2):137–145, April 1980.

[169] I.A. McCowan and H. Bourlard. Microphone array post-filter for diffuse noise field. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 905–908, Orlando, Florida, USA, May 2002.

[170] I. Merks. *Binaural application of microphone arrays for improved speech intelligibility in a noisy environment*. PhD thesis, Technische Universiteit Delft, January 2000.

[171] J. Meyer. Microphone array for hearing aids taking into account the scattering of the head. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 27–30, 2001.

[172] U. Mittal and N. Phamdo. Signal/noise KLT based approach for enhancing speech degraded by colored noise. *IEEE Trans. Speech, Audio Processing*, 8(2):159–167, March 2000.

[173] M. Moonen and I. Proudler. An introduction to adaptive signal processing. Course Notes, Katholieke Universiteit Leuven, 2001.

[174] B. C. J. Moore. *Perceptual consequences of cochlear damage*. Oxford University Press, Oxford, 1995.

[175] B. C. J. Moore. *Cochlear hearing loss*. Whurr Publishers Ltd, London, 1998.

[176] B. C. J. Moore. Psychoacoustics of cochlear hearing impairment and the design of hearing aids. *Journal of the Acoustical Society of America*, 103(5):2991, May 1998.

[177] B. C. J. Moore. 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.

[178] 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.

[179] M. N. Murthi and B. D. Rao. All-pole modeling of speech based on minimum variance distortionless response spectrum. 8(3):221–239, May 2000.

[180] J. Neumann and S. Laugesen. Method for noise reduction and microphone array for performing noise reduction. WO patent, WO 03/003349, 2003.

[181] L. Ngia. *System modeling using basis functions and applications to echo cancellation*. PhD thesis, Chalmers Institute of Technology, Sweden, 2000.

[182] J. Nielsen and M. Ekelid. Feedback cancellation using bandwidth detection. WO Patent, WO 01/06746 A2, 2001.

[183] J. Nielsen and M. Ekelid. Feedback cancellation with low frequency input. WO Patent, WO 01/06812 A1, 2001.

[184] 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, July 1999.

[185] 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.

[186] R. Nishimura, Y. Suzuki, and F. Asano. A new adaptive binaural microphone array system using a weighted least squares algorithm. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1925–1928, Orlando, Florida, USA, may 2002.

[187] S. Nordebo, I. Claesson, and S. Nordholm. Adaptive beamforming: spatial filter designed blocking matrix. *IEEE Journal of Oceanic Engineering*, 19(4):583–590, October 1994.

[188] 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, July 1994.

[189] S. Nordholm, I. Claesson, and M. Dahl. Adaptive microphone array employing calibration signals: an analytical evaluation. *IEEE Trans. Speech, Audio Processing*, 7(3):241–22, May 1999.

[190] S. Nordholm, I. Claesson, and N. Grbíc. *Optimal and adaptive microphone arrays for speech input in automobiles*, chapter 14 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 307–326. Springer-Verlag, May 2001.

[191] S. Nordholm, K. Rehbock, and S. Nordebo. Chebyshev optimization for the design of broadband beamformers in the near field. *IEEE Trans. Circuits and Systems–II: Analog and Digital Signal Processing*, 45(1):141–143, January 1998.

[192] S. E. Nordholm and Y. H. Leung. Performance limits of the broadband Generalized Sidelobe Cancelling structure in an isotropic noise field. *Journal of the Acoustical Society of America*, 107(2):1057–1060, February 2000.

[193] P. Nordqvist and A. Leijon. An efficient robust sound classification algorithm for hearing aids. *Journal of the Acoustical Society of America*, 115(6):3033–3041, June 2004.

[194] L. Olsen, H. Müsch, and C. Struck. Digital solutions for feedback control. *The Hearing Review*, May 2001.

[195] J. Ouellette. The incredible shrinking microphone. *The Industrial Physicist*, 5(4):7–9, August 1999.

[196] Y. Park, D. Kim, and I. Kim. An efficient feedback cancellation for multiband compression hearing aids. In *Proc. of 20th Annual Int. Conf. Eng. in Medicine and Biology Soc.*, volume 5, pages 2706–2709, 1998.

[197] A. J. Parkinson, W. S. Parkinson, R. S. Tyler, M.W. Lowder, and B. J. Gantz. Speech perception performance in experienced cochlear implant patients receiving the SPEAK processing strategy in the Nucleus Spectra-22 cochlear implant. *Journal of Speech, Language, and Hearing Research*, 41(5):1073–1087, October 1998.

[198] P. M. Peterson. *Adaptive array processing for multiple microphone hearing aids*. PhD thesis, Dept. Elect. Eng. and Comp. Sci., M.I.T, Cambridge, MA, 1989. available as Res. Lab. Elect. Techn. Rept. 541.

[199] P. M. Peterson, A.-M. Wie, W. M. Rabinowitz, and P. M. Zurek. Robustness of an adaptive beamforming method for hearing aids. *Acta Otolaryngol. Suppl.*, 469:85–90, 1990.

[200] R. Plomp. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America*, 63(2):533–549, February 1978.

[201] R. Plomp and A. M. Mimpen. Speech-reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America*, 66(5):1333–1342, 1979.

[202] R. Porayath and J. Daniel. Acoustic feedback elimination using adaptive notch filter algorithm. U.S. patent, US 5999631, 1999.

[203] T. A. Powers and V. Hamacher. Three-microphone instrument is designed to extend benefits of directionality. *The Hearing Journal*, 55(10):38–45, October 2002.

[204] B. Rafaely and M. Roccasalva-Firenze. Control of feedback in hearing aids – a robust filter design approach. *IEEE Trans. Speech, Audio Processing*, 8(6):754–756, November 2000.

[205] 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.

[206] B. Rafaely, N. A. Shusina, and J. L. Hayes. Robust compensation with adaptive feedback cancellation in hearing aids. *Speech Communication*, 39(1-2):163–170, January 2003.

[207] L. J. Revitt. Two techniques for dealing with the occlusion effect. *Hearing Instruments*, 43(12):16–18, 1992.

[208] A. Rezayee and S. Gazor. An adaptive KLT approach for speech enhancement. *IEEE Trans. Speech, Audio Processing*, 9(2):87–95, February 2001.

[209] T. Ricketts and P. Henry. Evaluation of an adaptive directional-microphone hearing aid. *International Journal of Audiology*, 41(2), 2002.

[210] G. Rombouts. *Adaptive filtering algorithms for acoustic echo and noise cancellation*. PhD thesis, Faculty of Engineering, K.U. Leuven, Belgium, May 2003. available as Technical Report ESAT-SISTA/TR 03-89.

[211] G. Rombouts and M. Moonen. An integrated approach to acoustic echo and noise suppression. Submitted to *Signal Processing*, 2002.

[212] G. Rombouts and M. Moonen. QRD-based unconstrained optimal filtering for acoustic noise reduction. *Signal Processing*, 83(9):1889–1904, September 2003.

[213] G. Rombouts and M. Moonen. Fast-QRD-based optimal filtering for acoustic noise reduction. Accepted for publication in *IEEE Trans. Speech, Audio Processing*, 2004.

[214] G. H. Saunders and J. M. Kates. Speech intelligibility enhancement using hearing-aid array processing. *Journal of the Acoustical Society of America*, 102(3):1827–1837, September 1997.

[215] T. Schneider, R. Brennan, P. Balsiger, and A. Heubi. An ultra low-power programmable DSP system for hearing aids and other audio applications. In *Proc. IEEE Int. Conf. Signal Processing Applications and Technology (ICSPAT)*, Orlando, Florida, USA, November 1999.

[216] M. R. Schroeder. Improvement of acoustic-feedback stability by frequency shifting. *Journal of the Acoustical Society of America*, 36(9):1718–1724, September 1976.

[217] N. A. Shusina and B. Rafaely. Feedback cancellation in hearing aids based on indirect closed-loop identification. In *Proc. IEEE Benelux Signal Processing Symposium (SPS)*, pages 177–180, Leuven, Belgium, March 2002.

[218] J. J. Shynk. Adaptive IIR filtering. *IEEE ASSP Magazine*, 6(2):4–21, April 1989.

[219] J. J. Shynk. Frequency-domain and multirate adaptive filtering. *IEEE Signal Processing Magazine*, pages 14–37, January 1992.

[220] K. U. Simmer, J. Bitzer, and C. Marro. *Post-filtering techniques*, chapter 3 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 39–57. Springer-Verlag, May 2001.

[221] M. G. Siqueira and A. Alwan. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. *IEEE Trans. Speech, Audio Processing*, 8(4):443–453, July 2000.

[222] H. R. Skovgaard. Hearing aid compensating for acoustic feedback. US Patent, US 5680467, 1997.

[223] S. D. Snyder and C. H. Hansen. The effect of transfer function estimation errors on the filtered-X LMS algorithm. *IEEE Trans. Signal Processing*, 42(4):950–953, April 1994.

[224] W. Soede, A. J. Berkhout, and F. A. Bilsen. Assesment of a directional microphone array for hearing-impaired listeners. *Journal of the Acoustical Society of America*, 94(2):799–808, August 1993.

[225] W. Soede, A. J. Berkhout, and F. A. Bilsen. Development of a directional hearing instrument based on array technology. *Journal of the Acoustical Society of America*, 94(2):785–798, August 1993.

[226] S. D. Soli, P. Widin, and K. M. Buckley. Auditory prosthesis, noise suppression apparatus and feedback suppression apparatus having focused adaptive filtering. U.S. patent, US 5402496, 1993.

[227] A. Spriet, M. Moonen, and I. Proudler. An unbiased modelling approach to feedback cancellation in hearing aids. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS)*, pages 5–8, Belgium, March 2002.

[228] A. Spriet, M. Moonen, and I. Proudler. Feedback cancellation in hearing aids: an unbiased modelling approach. In *Proc. European Signal Processing Conf. (EUSIPCO)*, volume 1, pages 531–534, Toulouse, France, September 2002.

[229] A. Spriet, M. Moonen, and I. Proudler. Feedback cancellation in hearing aids: an unbiased modelling approach. In *Proc. of the IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA)*, pages 29–32, Belgium, November 2002.

[230] A. Spriet, M. Moonen, and J. Wouters. More than 2 microphones in digital hearing aids: a clear benefit for speech intelligibility in noise! Technical Report ESAT-SISTA/TR 2000-143, Katholieke Universiteit Leuven, Belgium, December 2000.

[231] A. Spriet, M. Moonen, and J. Wouters. Multi-channel subband GSVD for speech enhancement in hearing aids. Technical Report ESAT-SISTA/TR 2000-142, Katholieke Universiteit Leuven, Belgium, December 2000.

[232] A. Spriet, M. Moonen, and J. Wouters. A multichannel subband GSVD based approach for speech enhancement in hearing aids. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 187–191, Darmstadt, Germany, September 2001.

[233] A. Spriet, M. Moonen, and J. Wouters. Multi-microphone noise reduction algorithms for hearing aids. In *Proc. of Nederlands Akoestisch Genootschap (NAG)*, number 157, pages 51–60, Utrecht, The Netherlands, September 2001.

[234] A. Spriet, M. Moonen, and J. Wouters. Robustness analysis of GSVD based optimal filtering and Generalized Sidelobe Canceller for hearing aid applications. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 31–34, New Paltz, NY, USA, October 2001.

[235] A. Spriet, M. Moonen, and J. Wouters. A multi-channel subband generalized singular value decomposition approach to speech enhancement. *European Transactions on Telecommunications*, 13(2):149–158, March-April 2002.

[236] A. Spriet, M. Moonen, and J. Wouters. The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 501–504, Hong Kong, China, April 2003.

[237] A. Spriet, M. Moonen, and J. Wouters. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction in hearing aids. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 147–150, Kyoto, Japan, September 2003.

[238] A. Spriet, M. Moonen, and J. Wouters. Robustness analysis of multichannel Wiener filtering and Generalized Sidelobe Cancellation for multimicrophone noise reduction in hearing aid applications. Accepted for publication in *IEEE Trans. Speech, Audio Processing*, 2004.

[239] A. Spriet, M. Moonen, and J. Wouters. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. Accepted for publication in *Signal Processing* (extended version: Technical Report ESAT-SISTA/TR 2003-46a, K.U. Leuven), 2004.

[240] 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. Accepted for publication in *IEEE Trans. on Signal Processing* (extended version: Technical Report ESAT-SISTA/TR 2003-47a, K.U. Leuven), 2004.

[241] A. Spriet, M. Moonen, and J. Wouters. Stochastic gradient implementation of spatially pre-processed multi-channel Wiener filtering for noise reduction in hearing aids. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 57–60, Montreal, Quebec, Canada, May 2004.

[242] A. Spriet, M. Moonen, and J. Wouters. The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and GSC. Accepted for publication in *Signal Processing*, 2004.

[243] A. Spriet, I. Proudler, M. Moonen, and J. Wouters. Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal. Accepted for publication in *IEEE Transactions on Signal Processing*, 2004.

[244] A. Spriet and K. Vanbleu. Ruisonderdrukking met behulp van iteratieve Wiener filters. Master's thesis, Katholieke Universiteit Leuven, Belgium, 1999.

[245] 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, September 1993.

[246] 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 2004.

[247] M. A. Stone and B. C. J. Moore. Tolerable hearing aid delays. I. estimation of limits imposed by the auditory path alone using simulated hearing losses. *Ear and Hearing*, 20(3):182–191, 1999.

[248] M. A. Stone, B. C. J. Moore, J. I. Alcantara, 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.

[249] B. Strope and A. Alwan. A novel structure to compensate for frequency-dependent loudness recruitment of sensorineural hearing loss. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3539–3542, Detroit, USA, 1995.

[250] C. Sydow. Broadband beamforming for a microphone array. *Journal of the Acoustical Society of America*, 96(2):845–849, August 1994.

[251] J. Tchorz. Effects of microphone mismatch in directional hearing instruments. *The Hearing Review*, 8(10), October 2001.

[252] H. Teutsch and G. W. Elko. First- and second-order adaptive differential microphone arrays. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 35–38, Darmstadt, Germany, September 2001.

[253] S. Thipphayathetthana and C. Chinrungrueng. Variable step-size of the least-mean-square algorithm for reducing acoustic feedback in hearing aids. In *IEEE Asi-Pacific Conf. on Circuits and Systems*, pages 407–410, 2000.

[254] S. C. Thompson. Directional patterns obtained from two or three microphones. Technical report, Knowles Electronics, Illinois, USA, September 2000.

[255] Z. Tian, K.L. Bell, and H.L. Van Trees. A recursive least squares implementation for LCMP beamforming under quadratic constraint. *IEEE Trans. Signal Processing*, 49(6):1138–1145, June 2001.

[256] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis. Speech enhancement based on audible noise suppression. *IEEE Trans. Speech, Audio Processing*, 5(6):497–513, November 1997.

[257] M. Valente, D. A. Fabry, L. G. Potts, and R. E. Sandlin. Comparing the performance of the Widex SENSO digital hearing aid with analog hearing aids. *Journal of the American Academy of Audiology*, 9(5):342–360, October 1998.

[258] D. Van Compernolle. Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, pages 833–836, Albuquerque, April 1990.

[259] D. Van Compernolle and S. Van Gerven. Beamforming with microphone arrays. In *COST 229: Applications of Digital Signal Processing to Telecommunications (V. Cappelini and A. R. Figueiras-Vidal, Eds.)*, pages 107–131. 1995.

[260] H. L. Van Trees. *Optimum array processing: Part IV of detection, estimation and modulation theory.* John Wiley, 2002.

[261] B. D. Van Veen and K. M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, 5(2):4–24, April 1988.

[262] T. van Waterschoot, G. Rombouts, and M. Moonen. On the performance of decorrelation by prefiltering for adaptive feedback cancellation in public address systems. In *Proc. of IEEE Benelux Signal Processing Symposium (SPS)*, pages 167–170, Hilvarenbeek, The Netherlands, 2004.

[263] J. Vanden Berghe and J. Wouters. An adaptive noise canceller for hearing aids using two nearby microphones. *Journal of the Acoustical Society of America*, 103:3621–3626, June 1998.

[264] I. Veit and H. Sander. Production of spatially limited diffuse sound field in an anechoic room. *Journal of the Audio Engineering Society*, 35(3):128–142, 1987.

[265] N. Virag. Single channel speech enhancement based on masking properties on the human auditory system. *IEEE Trans. Speech, Audio Processing*, 7(2):126–137, March 1999.

[266] A. M. Vural. Effects of perturbations on the performance of optimum/adaptive arrays. *IEEE Trans. Aerospace and Electronic Systems*, 15(1):76–87, January 1979.

[267] B. E. Walden, R. K. Surr, M. T. Cord, B. Edwards, and L. Olson. Comparison of benefits provided by different hearing aid technologies. *Journal of the American Academy of Audiology*, 11(10):540–560, Nov-Dec 2000.

[268] R. Wang and R. Harjani. Acoustic feedback cancellation in hearing aids. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 137–140, Minneapolis, USA, April 1993.

[269] T. Weidner. Method for feedback recognition in a hearing aid and a hearing aid operating according to the method. U.S. patent, US 6404895, 2002.

[270] S. Weiss and R. W. Stewart. Performance limitations of subband adaptive filters. In *Proc. European Signal Processing Conf. (EUSIPCO)*, pages 1245–1248, Rhodos, Greece, September 1998.

[271] D. P. Welker, J. E. Greenberg, J. G. Desloge, and P. M. Zurek. Microphone-array hearing aids with binaural output–part II: A two-microphone adaptive system. *IEEE Trans. Speech, Audio Processing*, 5(6):543–551, November 1997.

[272] S. Westermann. The occlusion effect. *Hearing Instruments*, 38(6):443–453, 1987.

[273] B. Widrow and S. Stearns. *Adaptive Signal Processing*. Prentice Hall, Englewood Cliffs, 1985.

[274] T. Wittkop. *Two-channel noise reduction algorithms motivated by models of binaural interaction*. PhD thesis, Universität Oldenburg, Oldenburg, Germany, 2001.

[275] B. D. Woodruff and D. A. Preves. Fixed filter implementation of feedback cancellation for in-the-ear hearing aids. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 22–23, New Paltz, NY, 1995.

[276] J. Wouters, L Geurts, J. Laneau, H. Luts, J.-B. Maj, M. Moonen, L. Royackers, A. Spriet, T. Van den Bogaert, and A. van Wieringen. Digital hearing aids and future directions for hearing aids. *Acta Oto-Rhino-Laryngologica Belgica*, 56(4):357–361, June 2002.

[277] J. Wouters, L. Litière, and A. Van Wieringen. Speech intelligibility in noisy environments with one- and two-microphone haring aids. *Audiology*, 38(2):91–98, Mar.-Apr. 1999.

[278] J. Wouters and J. Vanden Berghe. Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system. *Ear and Hearing*, 22(5):420–430, October 2001.

[279] J. Yang, F.-L. Luo, and A. Nehorai. Spectral contrast enhancement: Algorithms and comparisons. *Speech Communication*, 39(1-2):33–46, January 2003.

[280] K.F.C. Yiu, N. Grbíc, K.-L. Teo, 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.

[281] P. Young. *Recursive estimation and time-series analysis*. Springer-Verlag, 1984.

[282] R. Zelinski. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2578–2581, New York, April 1988.

[283] R. Zelinski. Noise reduction based on microphone array with LMS adaptive post-filtering. *Electronics Letters*, 26(24):2036–2037, November 1990.

List of publications

International journal papers

- A. Spriet, M. Moonen and J. Wouters. A multichannel subband GSVD approach to speech enhancement. *European Transactions on Telecommunications (special issue on Acoustic Echo and Noise Control)*, 13(2):149–158, March/April 2002.

- J. Wouters, L. Geurts, J. Laneau, H. Luts, J.-B. Maj, M. Moonen, L. Royackers, A. Spriet, T. Van den Bogaert and A. van Wieringen. Digital hearing aids and future directions for hearing aids. *Acta Oto-Rhino-Laryngologica Belgica,* 56(4):357–361, June 2002.

- A. Spriet, M. Moonen and J. Wouters. Robustness analysis of multichannel Wiener filtering and Generalized Sidelobe Cancellation for multimicrophone noise reduction in hearing aid applications. *Accepted for publication in IEEE transactions on Speech and Audio processing*, 2004.

- A. Spriet, M. Moonen and J. Wouters. Stochastic gradient based implementation of spatially pre-processed multi-channel Wiener filter for noise reduction in hearing aids. *Accepted for publication in IEEE transactions on Signal Processing* (extended version: Tech. Rep. ESAT-SISTA/TR 2003-47a, K.U. Leuven), 2004.

- A. Spriet, M. Moonen and J. Wouters. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. *Accepted for publication in Signal Processing* (extended version: Tech. Rep. ESAT-SISTA/TR 2003-46a, K.U. Leuven), 2004.

- A. Spriet, M. Moonen and J. Wouters. The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation. *Accepted for publication (minor revision) in Signal Processing*, 2004.

- A. Spriet, I. Proudler, M. Moonen and J. Wouters. Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal. *Accepted for publication (minor revision) in IEEE transactions on Signal Processing,* 2004.

International conference papers

- A. Spriet, M. Moonen and J. Wouters. A multichannel subband GSVD based approach for speech enhancement in hearing aids. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 187–191, Darmstadt, Germany, September 2001.

- A. Spriet, M. Moonen and J. Wouters. Multi-microphone noise reduction algorithms for hearing aids. In *Proc. of Nederlands Akoestisch Genootschap (NAG)*, number 157, pages 51–60, Utrecht, The Netherlands, September 2001.

- A. Spriet, M. Moonen and J. Wouters. Robustness analysis of GSVD based optimal Filtering and Generalized Sidelobe Canceller for hearing aid applications. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 31–34, New Paltz, NY, USA, October 2001.

- A. Spriet, M. Moonen and I. Proudler. An unbiased modelling approach to feedback cancellation in hearing aids. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS)*, pages 5–8, Belgium, March 2002.

- A. Spriet, M. Moonen and I. Proudler. Feedback cancellation in hearing aids: an unbiased modelling approach. In *Proc. European Signal Processing Conf. (EUSIPCO)*, volume 1, pages 531–534, Toulouse, France, September 2002.

- A. Spriet, M. Moonen and I. Proudler. Feedback cancellation in hearing aids: an unbiased modelling approach. In *Proc. of the IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA)*, pages 29–32, Belgium, November 2002.

- A. Spriet, M. Moonen and J. Wouters. The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages 501–504, Hong Kong, China, April 2003.

- A. Spriet, M. Moonen and J. Wouters. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction in hearing aids. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 147–150, Kyoto, Japan, September 2003.

- A. Spriet, M. Moonen and J. Wouters. Stochastic gradient implementation of spatially pre-processed multi-channel Wiener filtering for noise reduction in hearing aids. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 57–60, Montreal, Quebec, Canada, May 2004.
- S. Doclo, A. Spriet and M. Moonen. Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS)*, pages 195–198, Hilvarenbeek, The Netherlands, April 2004.
- S. Doclo, A. Spriet, M. Moonen and J. Wouters. Design of a robust multi-microphone noise reduction algorithm for hearing instruments. In *Proc. of the International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Leuven, Belgium, July 2004.
- S. Doclo, A. Spriet and M. Moonen. Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction. In *Proc. European Signal Processing (EUSIPCO)*, Vienna, Austria, September 2004.

Abstracts

- A. Spriet, M. Moonen and J. Wouters. Multichannel subband GSVD-based filtering for noise reduction in hearing aids. In *20th Benelux Meeting on Systems and Control*, Houffalize, Belgium, March 2001.
- J. Wouters, L. Geurts, J. Laneau, H. Luts, J.-B. Maj, M. Moonen, A. Spriet and A. van Wieringen. More signal processing for better speech perception? In *Conference on Implantable Auditory Prostheses*, Asilomar, California, US, August 2001.
- J. Wouters, J.-B. Maj, M. Moonen and A. Spriet. Zooming in on speech: multimicrophone noise reduction strategies in hearing aids. In *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe, California, US, August 2002.
- J. Wouters, J.-B. Maj, M. Moonen, A. Spriet and J. Vanden Berghe. Speech recognition in noise for cochlear implantees with a two-microphone adaptive noise reduction system. In *7th International Cochlear Implant Conference*, Manchester, UK, September 2002.
- A. Spriet, J.-B. Maj, J. Wouters, M. Moonen, B. Van Dijk and J. Janssen. Next generation, multi-microphone noise reduction system for cochlear implants. In *DSP Valley's Annual Research and Technology Symposium (DARTS)*, DSP valley, Leuven, Belgium, October 2002.

- J. Wouters, H. Luts, J.-B. Maj, M. Moonen, W. Soede and A. Spriet. What have we learned about multi-microphone noise reduction strategies for a BTE hearing aid? In *Workshop of the International Collegium for Rehabilitative Audiology (ICRA)*, Elsinore, Denmark, June 2003.

- A. Spriet, J. Wouters and M. Moonen. The impact of signal model errors and speech detection errors on the noise reduction performance of GSC and multi-channel Wiener filtering. In 5*th International Workshop on Microphone Array Systems - Theory and Practice,* Erlangen-Nuremberg, Germany, May 2003.

- S. Doclo, A. Spriet, J.-B. Maj, M. Moonen, J. Wouters, B. Van Dijk and J. Janssen. Design and low-cost implementation of a robust multichannel noise reduction scheme for cochlear implants. In *DSP Valley's Annual Research and Technology Symposium (DARTS)*, DSP valley, Leuven, Belgium, October 2003.

- S. Doclo, A. Spriet, J. Wouters and M. Moonen. Design, implementation and evaluation of a robust multi-microphone noise reduction algorithm for hearing instruments. In *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe CA, USA, August 2004.

- J. Wouters, S. Doclo, T. Klasen, J.-B. Maj, M. Moonen, L. Royackers, A. Spriet and T. Van den Bogaert. Noise reduction approaches for improved speech perception. In *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe CA, USA, August 2004.

Internal reports

- A. Spriet, M. Moonen, J. Wouters. Multichannel subband GSVD for speech enhancement in hearing aids. *Tech. Rep. ESAT-SISTA/TR 2000-142*, Katholieke Universiteit Leuven, Belgium, December 2000.

- A. Spriet, M. Moonen, J. Wouters. More than 2 microphones in digital hearing aids: a clear benefit for speech intelligibility in noise! *Tech. Rep. ESAT-SISTA/TR 2000-143*, Katholieke Universiteit Leuven, Belgium, December 2000.

Curriculum vitae

Ann Spriet was born in Izegem, Belgium, in 1976. In 1999, she received the electrical engineering degree in applied sciences (summa cum laude) from the Katholieke Universiteit Leuven, Belgium. For her Master's thesis, entitled 'Speech enhancement using iterative Wiener filtering techniques', she visited Aalborg University, Aalborg, Denmark (Prof. Søren Holdt Jensen). In October 1999, she started her Ph.D. at the SCD-SISTA laboratories of the Department of Electrical Engineering (Faculty of Applied Sciences) and the laboratory for experimental ORL (Faculty of Medicine) of the K.U. Leuven. Her supervisors are Prof. Marc Moonen and Prof. Jan Wouters. Her research interests are in the area of digital signal processing techniques for speech enhancement in hearing aids and cochlear implants.

From 1999 until 2003, she was funded by the Fonds voor Wetenschappelijk Onderzoek - Vlaanderen. She has been involved in the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry) project 'Innovative speech processing algorithms for improved performance of cochlear implants' in cooperation with Cochlear Technology Centre Europe and in a research project with GN ReSound.

In 2001, she received the Lapperre Award. In 2003, she received a Best Student Paper Award at the IEEE International Workshop on Acoustic Echo and Noise Control.